Current Practices in Ophthalmology
Series Editor: Parul Ichhpujani

Parul Ichhpujani
Sahil Thakur *Editors*

Artificial Intelligence and Ophthalmology

Perks, Perils and Pitfalls

Springer

Current Practices in Ophthalmology

Series Editor

Parul Ichhpujani
Department of Ophthalmology
Government Medical College and Hospital
Chandigarh, India

This series of highly organized and uniform handbooks aims to cover the latest clinically relevant developments in ophthalmology. In the wake of rapidly evolving innovations in the field of basic research, pharmacology, surgical techniques and imaging devices for the management of ophthalmic disorders, it is extremely important to invest in books that help you stay updated. These handbooks are designed to bridge the gap between journals and standard texts providing reviews on advances that are now part of mainstream clinical practice. Meant for residents, fellows-in-training, generalist ophthalmologists and specialists alike, each volume under this series covers current perspectives on relevant topics and meets the CME requirements as a go-to reference guide. Supervised and reviewed by a subject expert, chapters in each volume provide leading-edge information most relevant and useful for clinical ophthalmologists. This series is also useful for residents and fellows training in various subspecialties of ophthalmology, who can read these books while at work or during emergency duties. Additionally, these handbooks can aid in preparing for clinical case discussions at various forums and examinations.

More information about this series at http://www.springer.com/series/15743

Parul Ichhpujani • Sahil Thakur
Editors

Artificial Intelligence and Ophthalmology

Perks, Perils and Pitfalls

Springer

Editors
Parul Ichhpujani
Department of Ophthalmology
Government Medical College and Hospital
Chandigarh
India

Sahil Thakur
Ocular Epidemiology Research Group
Singapore Eye Research Institute
Singapore

ISSN 2523-3807 ISSN 2523-3815 (electronic)
Current Practices in Ophthalmology
ISBN 978-981-16-0633-5 ISBN 978-981-16-0634-2 (eBook)
https://doi.org/10.1007/978-981-16-0634-2

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

About the Authors

Parul Ichhpujani is currently a professor in the Department of Ophthalmology at Government Medical College and Hospital, Chandigarh, India, where she is chiefly responsible for glaucoma and neuro-ophthalmology services. She completed her glaucoma training at the Advanced Eye Centre, Postgraduate Institute of Medical Education and Research, Chandigarh, India, and in a subsequent clinical research fellowship, under Dr. George L Spaeth, at Wills Eye Institute, Philadelphia, USA. She currently serves on the education committee of the World Glaucoma Association and is the associate managing editor of the *Journal of Current Glaucoma Practice*. She was ranked among the Powerlist 2015 for the "Best 40 ophthalmologists under 40."

An avid researcher, Dr. Ichhpujani has coauthored three books: *Pearls in Glaucoma Therapy*, *Living with Glaucoma*, and *Smart Resources in Ophthalmology*; and has edited another eight: *Expert Techniques in Ophthalmology*, *Glaucoma: Basic and Clinical Perspectives*, *Manual of Glaucoma*, *Clinical Cases in Glaucoma: An Evidence-Based Approach*, *Glaucoma: Intraocular Pressure and Aqueous Dynamics*, *Current Advances in Ophthalmic Technology*, *Glaucoma*, and *Ophthalmic Instruments and Surgical Tools*. She is also the Springer series editor for the "Current Practices in Ophthalmology" series. She has also contributed several research articles and book chapters in national and international books and serves as a reviewer for many ophthalmology journals. She is heavily involved in resident and fellow teaching.

Sahil Thakur, MBBS, MS graduated from Government Medical College and Hospital, Chandigarh, India. He is currently a clinical research fellow at the Singapore Eye Research Institute. Apart from pursuing clinical ophthalmology, he has a keen interest in developing affordable medical technology and clinical photography. He is also involved in the research and development of digital solutions for faster and efficient diagnosis of common ophthalmic disorders. Previously he has coauthored a book, *Smart Resources in Ophthalmology: Applications and Social Networking*, and has more than 50 peer-reviewed publications. Dr. Thakur also has an interest in developing tools for ophthalmic education and has published two Android applications (Ophthalmaster and Ophthalminion) for use by residents and medical school students.

A Clinician's Introduction to Artificial Intelligence

1

Sahil Thakur and Ching-Yu Cheng

1.1 Artificial Intelligence

To understand the concept of artificial intelligence (AI) and how it is being used in applications today, we first need to understand the concept of intelligence. The term intelligence is derived from the Latin noun '*intellēctus*' or verb '*intelligere*', which means to comprehend or perceive. This concept is however abstract and is better understood with examples of different types of intelligence and how humans display them.

1. Visual-spatial: physical environment characteristics (architects when designing a building according to terrain and surroundings, navigating a boat in water).
2. Kinaesthetic: body movements (technical skill and precision of a ballerina, surgeons or athletes).
3. Creative: novel thought, typically expressed in art, music and writing (imagination-driven authors, painters and musicians).
4. Interpersonal: interaction with others (interviewers, shopkeepers, businessmen).
5. Intrapersonal: self-realisation (meditation, goal planning, self-preservation).

S. Thakur (✉)
Ocular Epidemiology Research Group, Singapore Eye Research Institute, Singapore

C.-Y. Cheng
Ocular Epidemiology Research Group, Singapore Eye Research Institute, Singapore, Singapore

Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore

Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_1

6. Linguistic: manipulation of words for communication (day-to-day communication).
7. Logical-mathematical: calculations, identifying patterns, analysing relationships (logic, puzzles, computing numbers).

From this classification, it is easy to understand what AI today is capable of and where we may be heading in the future. The simulation potential of logical-mathematical intelligence is the maximum and early development in AI almost exclusively focused on this domain. Robotics aims to mimic the kinaesthetic intelligence while sensor-driven (LiDAR scanners in self driven cars) applications leverage on visual-spatial intelligence. Chatbots are trying to mimic linguistic and interpersonal intelligence while the creative and intrapersonal intelligence are domains with limited to no simulation potential. Utility of algorithms has been explored to create music and draw art, but this is mainly driven by logical-mathematical intelligence.

When we understand what an algorithm can and cannot do, that is when we can maximise the utility of the algorithm. Thus, it is imperative to stay away from over-optimistic predictions and avoid false promises to increase the acceptability of algorithms and their potential widespread use. Some algorithms that have achieved this level of acceptance are the 'search engine' algorithms that offer personalised search results, spam filters in email clients, recommendations in applications like Netflix or Amazon and computational photography algorithms on mobile devices. Algorithms that have been developed for use in hospitals however are yet to see such levels of acceptance. This has been due to the inherent nature of patient–physician relationship, potential regulatory hurdles and multiple types of bias that confound these algorithms. However, with FDA approvals being given to 64 algorithms (SaMD: software as medical device) over the last 3 years, we can expect widespread availability of these options for clinicians in the future [1, 2]. Currently for ophthalmology only the IDx-DR has been approved as an autonomous AI diagnostic system for diabetic retinopathy [3–5].

1.2 The Past and What We Can Learn from It

The earliest examples of humans trying to build intelligent devices were the abacus like devices namely the *nepohualtzintzin* (Aztecs), *suanpan* (Chinese) or the *soroban* (Japan) [6]. These devices though based on simple concepts, reduced the time required for mathematical computations. This concept of reducing time and effort for repetitive computational tasks remains one of the driving concepts behind algorithm development.

The *Antikythera* mechanism was another ancient computing device that was probably used to track dates of important events, predict eclipses and even planetary motions [7]. Ramon Llull's *Ars Magna* was another device that used simple paper-based rotating concentric circle to generate combinations of new words and ideas. It was a rudimentary step towards generating a logical system to produce knowledge

[8]. Examples of such systems also exist in fictional literature like the book-writing engine in the city of Lagado in Gulliver's Travels. Attempts to create a similar algorithm include the RACTER program which generated text for the first computer authored book titled 'The Policeman's Beard is Half Constructed' in 1983 [9].

Perhaps one of the most significant inventions in primitive computing was the Difference Engine, proposed by Charles Babbage in 1822 [10]. In addition to this engine, Babbage also wanted to create the Analytical Engine which could be programmed using punch cards and had separate areas for number storage and computation. Ada Lovelace, the daughter of English poet, Lord Byron, gave the specifications for designing a program for this Engine. She is now considered by many as the first computer programmer [11].

Currently we know that data is stored in computers as a series of binary 1 s and 0 s called bits. Eight bits make up one byte. The fundamentals of this concept were published in a book, titled, 'An Investigation into the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities', by George Boole in 1854 [12]. He wanted to reduce logic to simple algebra involving only 0 and 1, with three simple operations: and, or and not. Boolean algebra, which is named after him, is one of the foundations of this digital age.

Over the next few decades, there were incremental improvements in algorithms for applications like optical character recognition (OCR), handwriting recognition (HWR) and speech synthesis. The next breakthrough was the 1943 paper 'A Logical Calculus of the Ideas Immanent in Nervous Activity' by Warren McCulloch and Walter Pitts [13]. In this paper, they described the basic mathematical model of the biological neuron. This formed the basis for the development of artificial neural networks (ANN) and deep learning (DL).

ENIAC, short for Electronic Numerical Integrator and Computer, was unveiled in 1946 and represented the pinnacle of specialised electronic, reprogrammable, digital computers built to solve a range of computing problems [14]. This started the race for development of powerful computer hardware for specialised operations by different countries. However, by today's standards even the Apollo Space Mission Guidance Computer (AGC) only had 64 KB memory and operated at 0.043 MHz, when compared to today's smartphones running with GHz speed processors (A14 chips in iPhones and iPads run at 3.0GHz and thus clock 70,000 times faster) shows how far we have come in terms of computing power due to the development of semiconductor technology [15, 16].

The term, 'artificial intelligence' (AI) was coined by John McCarthy at the Dartmouth conference for experts in this field in 1956 [17]. The expectations from this conference were extremely high despite limited computing power and hardware at that time. Inability to meet the hype generated by this conference, thus led to the AI winters of 1974–1980 and 1987–1993 [18].

Meanwhile during this time interesting developments were taking place in the backdrop, like:

- Rosenblatt concept of the perceptron [19] (1957).
- Arthur Lee Samuel's concept of machine learning [20] (1959).

- ELIZA: The program that could respond to text input simulating a conversation [21] (1964).
- Early deep learning using supervised multilayer perceptrons (1965).
- MYCIN: Rule-based expert system to identify sepsis and to recommend antibiotics [22] (1970).
- Fuzzy logic and its applications in automation [23] (1965–1974).
- Lighthill Report (criticised the utter failure of artificial intelligence in achieving its 'grandiose objectives') that triggered the first AI winter [24] (1973).
- Joseph Weizenbaum's early idea of ethics in AI, suggestion that AI should not be used as substitutes for humans in jobs requiring compassion, interpersonal respect, love, empathy and care [25] (1976).
- Expert system boom driven by LISP machines; however, LISP was soon overtaken by IBM/Apple with more powerful and cheaper consumer desktop computers, this led to collapse of the demand for expert systems [26] (1980–1987).
- Alex Waibel's Time Delay Neural Network (TDNN) which was the first convolutional network [27] (1987).
- Moravec's paradox: Tasks simple for humans like walking, talking, face/voice recognition are difficult for AI while humanly complex computational tasks involving mathematics and logic are simple [28] (1988).
- Yan LeCun developed system to recognise handwritten ZIP codes [29] (1989).
- Chinook (checkers playing algorithm) vs Marion Tinsley [30] (1994).
- IBM Deep Blue (chess playing algorithm) vs Garry Kasparov [31] (1997).
- Logistello (othello playing algorithm) vs Takeshi Murakami [32] (1997).
- Oh and Jung demonstrated power of graphical processing units (GPUs) for network training [33] (2004).
- ImageNet database [34] (2009).
- IBM DeepQA-based Watson winning the quiz show Jeopardy [35] (2011).
- Google DeepMind AlphaGo (based on ANN and Monte Carlo tree search algorithm defecting Lee Sedol) and AlphaGo Zero (trained by self-play without using previous data) which subsequently defected AlphaGo [36] (2017).
- Adversarial patches and perturbations [37, 38] (2018).
- Stanford death predictor [39] (2019).

Perhaps the most important developments that renewed interest in the field of AI and allowed widespread access over the last decade are the availability of large amounts of data and increased computational power at cheaper costs using modalities like graphical processing units (GPUs). ImageNet has especially been used to train popular models like the AlexNet [40], VGG16 [41], Inception modules [42] and the currently used ResNet [43].

Other datasets are also available for applications like music, facial recognition, text and speech processing [44]. As AI is a rapidly evolving field, today new innovations also happen with the same pace. However, understanding the history of AI is vital in predicting how it may affect the future. In further sections, we discuss why AI has become so popular today and how it may help in optimising patient care by evolving into an effective decision support system.

1.3 Why Should a Clinician Bother About AI?

A quick PubMed search shows how the number of articles published in the field of AI has grown to 112,594 results with 35,140 (31.2%) being published since 2018 [45]. Another insight comes from the Gartner Hype Index that monitors and predicts how a technology will evolve over time [46]. Machine learning (ML) was at the peak of inflated expectation indicating impact of publicity and expectations in 2016, DL at the same peak in 2018. These peaks also translate to the increase in applications that were developed using these technologies in this time. PubMed search shows a total of 49, 721 results till 2020 for ML, with 3885 results in 2016, 5217 in 2017 and 8169 in 2018. The last 2 years have seen 24,230 results which is 48.73% of total results [47]. Similarly, for DL, PubMed search shows 18,082 results till 2020 with 3020 results in 2018, 5401 in 2019 and 7383 in 2020 [48]. The last 2 years represent 70.7% of the total results. These numbers show how these technologies are being increasingly tried and tested for use in medicine.

Due to the lack of special training for understanding or evaluating these applications or their underlying concepts, a lot of effort has been recently initiated to make the clinicians more aware and sensitised about the use of AI in providing patient care [49–51]. In the next section, we describe a checklist approach to reading an AI paper with emphasis on evidence assessment and evaluation of future potential for translation to clinical use. We believe that this approach can help in better understanding of the scientific merit of the publication and its potential impact on care delivery practice patterns.

1.4 How to Read an Artificial Intelligence Paper?

Jaeschke et al. provided a framework to evaluate diagnostic tests in clinical medicine [52]. We have expanded the same framework to include relevant information about AI-based algorithms. We will initially describe the framework and then provide example of using the framework [53, 54]. The framework is as follows:

- Step 1: Evaluate if the study results are valid.
  Primary Guide
  - Was there an independent, blind comparison with a reference standard?
  - Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?
    For AI-based algorithms these can be adapted as:
      Are the datasets appropriate and described in sufficient detail?
      Was the gold standard for algorithm training appropriate and reliable?

Secondary Guide

- Did the results of the test being evaluated influence the decision to perform the reference standard?
- Were the methods for performing the test described in sufficient detail to permit replication?

  For AI-based algorithms these can be adapted as:

  Is the methodology of algorithm development described in sufficient detail to allow replication?

  Are the algorithm/datasets used available for external validation?

- Step 2: Evaluate the presented results.
  - Are likelihood ratios for the test results presented or data necessary for their calculation provided?

    For AI-based algorithms these can be adapted as:

    Are adequate and appropriate performance metrics reported? [50].

- Step 3: Evaluate the utility of results in providing care for your patients.
  - Will the reproducibility of the test result and its interpretation be satisfactory in my setting?
  - Are the results applicable to my patient?
  - Will the results change my management?
  - Will the patients be better off because of the test?

    For AI-based algorithms these can be adapted as:

    Are the findings of the algorithm explainable? Does the algorithm exhibit generalisability (can it be easily adapted for a different machine input or population)? Was the original algorithm performance too optimistic?

    Has the algorithm been validated in my local population?

    Is there any independent comparison of the algorithm with existing standard of care? Is there a cost-effectiveness analysis for rationale of algorithm use?

    Will there be a significant impact on patient well-being after algorithm deployment? Is there an attempt to measure this impact?

Table 1.1 shows how this framework can be used to evaluate an artificial intelligence paper.

Table 1.1 Framework for evaluation of artificial intelligence papers in medicine (adopted from Jaeschke et al.) [52]

Paper Title: Clinically applicable deep learning for diagnosis and referral in retinal disease [53]. *Purpose: Develop an artificial intelligence-based patient triage system using 3D OCT data*		
Step 1: Evaluate if the study results are valid		
• Was there an independent, blind comparison with a reference standard?	Are the datasets appropriate and described in sufficient detail?	The authors describe in detail the training set for OCT (Topcon) segmentation (877 scans), validation set for segmentation (224 scans), training set for classification (14,884 scans), validation set for classification (993 scans) and the testing set for comparison of algorithm (997 random scans) with standard of care
• Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?	Was the reference standard for algorithm training/testing appropriate and reliable?	The data for training the segmentation algorithm was manually segmented by trained ophthalmologists, reviewed and edited by senior ophthalmologists. The training set for classification used labels from automatic note search and trained ophthalmologists/optometrists reviewed the scans. The validation set for classification was graded by three junior graders, while for the test set, the referral gold standard was from full patient clinical records to determine the diagnosis and referral path considering subsequently obtained information. The algorithm performance was compared to four medical retina consultant ophthalmologists and four specialist optometrists
• Did the results of the test being evaluated influence the decision to perform the reference standard?	Did the results of the algorithm influence the decision to perform the reference standard?	No, the referral gold standard was retrospective data based on full clinical records of patients undergoing current standard of care
• Were the methods for performing the test described in sufficient detail to permit replication?	Is the methodology of algorithm development described in sufficient detail to allow replication? Are the algorithm/datasets used available for external validation?	The authors describe the algorithm (U-net architecture) in detail but mention that the data is not available in the public domain and may be available on request subject to local and national ethical approvals. In a subsequent paper, the authors mention about releasing the segmentation algorithm and dataset in the public domain for validation [54]

(continued)

Table 1.1 (continued)

Step 2: Evaluate the presented results		
• Are likelihood ratios for the test results presented or data necessary for their calculation provided?	Are adequate and appropriate performance metrics reported? [50]	The authors report ROC curves, confusion matrices, total error rates and impact of additional information (OCT alone, OCT + fundus + full case summary) on expert referral decisions. The algorithm had an AUC of 99.21 and error rate of 5.5% (55/997)
Step 3: Evaluate the utility of results in providing care for your patients		
• Will the reproducibility of the test result and its interpretation be satisfactory in my setting?	Are the findings of the algorithm explainable? Does the algorithm exhibit generalisability? Was the original algorithm performance too optimistic?	The authors report data for generalising using of the algorithm using another OCT device. (Spectralis), though initially the algorithm performs poorly and has error rate of 46.6% for referral decisions, retraining of the segmentation algorithm improves the AUC to 99.93 and reduces error rate to 3.4% (4/116). This shows that the algorithm is flexible and adaptable to a different machine. The authors also report results in a third OCT machine (Cirrus 5000) where initial error rate of 16.4% was reduced to 9.8% after retraining the segmentation algorithm The developers of the algorithm also tried to incorporate elements of explainable artificial intelligence by providing a segmentation maps with highlighted retinal structure, pathology, artefacts and predicted diagnostic probabilities and referral suggestions. However, in the videos provided as supplementary material, the automatic segmentation is not always accurate
• Are the results applicable to my patient?	Has the algorithm been validated in my local population?	No, the results are from the patient population at Moorfields eye hospital, London, United Kingdom. It will need further validation in different ethnic populations and research settings before it can be applicable to your patients
• Will the results change my management?	Is there any independent comparison of the algorithm with existing standard of care? Is there a cost effectiveness analysis for rationale of algorithm use?	The algorithm was compared to 4 medical retina consultant ophthalmologists and 4 specialist optometrists. The algorithm performed as well or outperformed the experts. There was no attempt however to assess the cost effectiveness of the algorithm as compared to standard of care
• Will the patients be better off because of the test?	Will there be a significant impact on patient well-being after algorithm deployment? Is there an attempt to measure this impact?	The algorithm has potential to be deployed as a clinician decision support tool but immediate impact on patient well-being cannot be assessed. No attempt was made by the authors to measure this impact in the real world

1.5 Conclusion

We have exciting times ahead of us, due to the immense potential of AI as a clinical decision support tool. However potential ethical and legal issues of liability management, reduction in clinical skills due to excessive algorithm use, inappropriate data representation especially for minorities, lack of personal privacy, 'biomarkup' due to excessive testing and inadequate understanding of algorithm results (AI black box) can hamper the deployment and acceptance of these AI algorithms [55–59]. Humans are intelligent, flexible and tenacious but are also liable to make mistakes. The embarrassing inability of Apple HealthKit to track menstrual cycles while tracking innocuous parameters for health monitoring like weight, height, inhaler use, alcohol content, blood sugar, sodium intake is just one example of this oversight [60]. Inherently the algorithms are unbiased but the bias from data used for training and the developers inherent bias can ultimately create complex ethical problems. An attitude of critical evaluation by all stakeholders before adoption of any new technology will thus help to separate the real from the hype. IBM Watson is an excellent example of how AI struggles with real-world medicine, messy hospital records and the expectations of industry, hospitals, physicians and patients [61]. We must keep in mind that our primary goal is always providing our patients the 'best' standard of care available. The affordability, availability and widespread social impact of the 'model of care' should also be considered while making this critical decision.

References

1. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med. 2020;3(1):118.
2. Software as a Medical Device (SAMD): clinical evaluation—guidance 2018 [updated 2018/08/31/]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/software-medical-device-samd-clinical-evaluation.
3. Abramoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci. 2016;57(13):5200–6.
4. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018;1(1):39.
5. Grzybowski A, Brona P. A pilot autonomous AI-based DR screening in Poland. Acta Ophthalmol. 2019;97(S263)
6. Projects CtW. Abacus—Wikipedia 2020 [updated 2020/11/02/]. Available from: https://en.wikipedia.org/w/index.php?title=Abacus&oldid=986730574.
7. Projects CtW. Antikythera mechanism—Wikipedia 2020 [updated 2020/11/03/]. Available from: https://en.wikipedia.org/w/index.php?title=Antikythera_mechanism&oldid=986880521.
8. Jensen T, editor. Ramon Llull's Ars magna. Cham: Springer; 2018.
9. Racter—visual melt 2020 [updated 2020/11/05/]. Available from: https://visualmelt.com/Racter.

10. Park E. What a difference the difference engine made: from Charles Babbage's calculator emerged today's computer. Smithsonian Magazine; 1996.
11. Ada Lovelace: founder of scientific computing 1998 [updated 1998/03/24/]. Available from: https://www.sdsc.edu/ScienceWomen/lovelace.html.
12. Boole G. An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities. Walton and Maberly; 1854.
13. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. 1943. Bull Math Biol. 1990;52(1–2):99–115. discussion 73–97
14. Levy S. The brief history of the ENIAC computer. Smithsonian Magazine; 2013.
15. Computers on board the Apollo spacecraft 2014 [updated 2014/07/04/]. Available from: https://history.nasa.gov/computers/Ch2-5.html.
16. Your smartphone is millions of times more powerful that all of NASA's combined computing in 1969 2020 [updated 2020/02/11/]. Available from: https://www.zmescience.com/science/news-science/smartphone-power-compared-to-apollo-432.
17. Rajaraman V. JohnMcCarthy—father of artificial intelligence. Resonance. 2014;19(3):198–207.
18. History of AI winters. Actuaries Digital 2020 [updated 2020/11/06/]. Available from: https://www.actuaries.digital/2018/09/05/history-of-ai-winters/#_ednref2.
19. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65(6):386–408.
20. Arthur Samuel 2007 [updated 2007/12/08/]. Available from: http://infolab.stanford.edu/pub/voy/museum/samuel.html.
21. Eliza, a chatbot therapist 2018 [updated 2018/07/18/]. Available from: https://web.njit.edu/~ronkowit/eliza.html.
22. van Melle W. MYCIN: a knowledge-based consultation program for infectious disease diagnosis. Int J Man-Mach Stud. 1978;10(3):313–22.
23. Cintula P, Fermüller CG, Noguera C. Fuzzy logic 2016 [updated 2016/11/15/]. Available from: https://plato.stanford.edu/entries/logic-fuzzy.
24. Lighthill report 2017 [updated 2017/12/20/]. Available from: http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm.
25. Weizenbaum J. Computer power and human reason: W.H. Freeman & Company; 1976.
26. Pollack A. Setbacks for artificial intelligence. NY Times; 1988.
27. Waibel A, Hanazawa T, Hinton G, Shikano K, Lang KJ. Phoneme recognition using time-delay neural networks. Readings in speech recognition. Morgan Kaufmann Publishers Inc.; 1990. p. 393–404.
28. Moravec H. Mind children. Cambridge, MA: Harvard University Press; 1990.
29. LeCun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1989;1(4):541–51.
30. Play Chinook—World man- machine checkers champion 2012 [updated 2012/03/08/]. Available from: https://webdocs.cs.ualberta.ca/~chinook/play.
31. IBM100—Deep blue. IBM corporation; 2012 [updated 2012/03/07/]. Available from: https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue.
32. LOGISTELLO' homepage 2011 [updated 2011/02/03/]. Available from: https://skatgame.net/mburo/log.html.
33. Oh K-S, Jung K. GPU implementation of neural networks. Pattern Recogn. 2004;37(6):1311–4.
34. Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. IEEE.
35. The DeepQA research team—IBM 2020 [updated 2020/11/06/]. Available from: https://researcher.watson.ibm.com/researcher/view_group_subpage.php?id=2160.
36. AlphaGo Zero: starting from scratch 2020 [updated 2020/11/06/]. Available from: https://deepmind.com/blog/article/alphago-zero-starting-scratch.
37. Brown T, Mane D, Roy A, Abadi M, Gilmer J. Adversarial patch. Google Research; 2017.
38. Su J, Vargas DV, Kouichi S. One pixel attack for fooling deep neural networks. arXiv. 2017.
39. Avati A, Jung K, Harman S, et al. Improving palliative care with deep learning. arXiv. 2017.

40. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.
41. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. 2014.
42. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. arXiv. 2014.
43. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv. 2015.
44. Datasets. Deep learning 2020 [updated 2020/11/06/]. Available from: http://deeplearning.net/datasets.
45. Artificial intelligence—Search results—PubMed 2020 [updated 2020/11/06/]. Available from: https://pubmed.ncbi.nlm.nih.gov/?term=artificial+intelligence&filter=years.1951-2020&timeline=expanded.
46. Hype cycle research methodology 2020 [updated 2020/11/06/]. Available from: https://www.gartner.com/en/research/methodologies/gartner-hype-cycle.
47. Machine learning—Search results—PubMed 2020 [updated 2020/11/06/]. Available from: https://pubmed.ncbi.nlm.nih.gov/?term=machine+learning&filter=years.1957-2020&timeline=expanded.
48. Deep learning—Search Results—PubMed 2020 [updated 2020/11/06/]. Available from: https://pubmed.ncbi.nlm.nih.gov/?term=deep+learning&filter=years.1954-2020&timeline=expanded.
49. Carin L, Pencina MJ. On deep learning for medical image analysis. JAMA. 2018;320(11):1192–3.
50. Yu M, Tham Y-C, Rim TH, et al. Reporting on deep learning algorithms in health care. Lancet Digit Health. 2019;1(7):e328–e9.
51. Liu Y, Chen P-HC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. JAMA. 2019;322(18):1806–16.
52. Jaeschke R, Guyatt GH, Sackett DL, et al. Users' guides to the medical literature: III. How to use an article about a diagnostic test B. what are the results and will they help me in caring for my patients? JAMA. 1994;271(9):703–7.
53. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342–50.
54. Moraes G, Fu DJ, Wilson M, et al. Quantitative analysis of OCT for neovascular age-related macular degeneration using deep learning. Ophthalmology. https://doi.org/10.1016/j.ophtha.2020.09.025.
55. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517–8.
56. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. JAMA. 2018;319(1):19–20.
57. Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. JAMA Intern Med. 2019;179(3):293–4.
58. Price WN II, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. JAMA. 2019;322(18):1765–6.
59. Mandl KD, Manrai AK. Potential excessive testing at scale: biomarkers, genomics, and machine learning. JAMA. 2019;321(8):739–40.
60. Eveleth R. The Atlantic 2018 [updated 2018/04/05/]. Available from: https://www.theatlantic.com/technology/archive/2014/12/how-self-tracking-apps-exclude-women/383673.
61. How IBM Watson overpromised and underdelivered on AI health care—IEEE Spectrum 2020 [updated 2020/11/06/]. Available from: https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care.

2 What You Need to Know About Artificial Intelligence: Technical Introduction

Oscar J. Perdomo, Santiago Toledo, Alvaro Orjuela, and Fabio A. González

2.1 Artificial Intelligence

The popularity of artificial intelligence (AI) has increased in different fields of application. Initial terminology pertaining to AI was proposed in the middle of the last century. However, in the nineties decade, expressions like expert systems, artificial neural networks, fuzzy systems, and others were based on statistical tools from the 60 and 70 decades, creating the field known as Machine Learning (ML). Currently, these topics are associated with automation applications and an emergent area of data science. Different processes have applied these technological tools thereby improving the solutions to problems in distinct fields and at different levels. In spite of the increment in the use of the terms AI and ML, it is difficult for many to determine what exactly AI is.

The Institute of Electrical and Electronics Engineers (IEEE) subdivides the AI field into three subfields:

1. artificial neural networks (ANN), which are based on connectionist models, trying to emulate the biological brain;
2. evolutionary algorithms that employ bioinspired methods of optimization as, for example, the mechanism of natural selection; and,
3. fuzzy logic, which use the natural language in human being, modifying the classical logic.

O. J. Perdomo (✉) · A. Orjuela
Universidad del Rosario, Bogotá, Colombia
e-mail: oscarj.perdomo@urosario.edu.co

S. Toledo · F. A. González
Universidad Nacional de Colombia, MindLab Research Group, Bogotá, Colombia

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_2

These paradigms have been included inside the concept of computational intelligence (CI). However, the similarities between AI and CI are notable, CI emerged from a community virtually different [1]. Since there is no unification, the terminology is wide and, currently, employ concepts related to the explanation as to how the machines learn [2]. Therefore, the main aspect of CI is the numeric representation of the knowledge compared to the symbolic representation of the AI.

Simultaneous to the development of the CI community, on the other side, Vapnik and Chervonenkis proposed more models that learn from data, which were coined as machine learning [3]. Mainly they used different statistical tools for different strategies to solve problems in classification and regression. The popularization of these models gave birth to the new field of ML.

The ML area is composed of models such as support vector machines (SVM), trees for regression and classification (RT), and ANNs, just to mention a few. More recently, Rosenblatt's work based on perceptron evinced improvements in the area, and later with the multilayer perceptron (MLP) and the backpropagation (backprop, BP) algorithm from McLellan [4] has exhibited its splendor.

Neural networks were reborn with the addition of layers to the ANN models. This field has been named deep learning (DL), and currently, it is one of the most popular methods to solve challenges in image processing and computer vision [5]. The number of parameters of the neural network has been increased with additional problems in the training model, where more synaptic weights have to be tuned, demanding more computation capacity and time processing. Figure 2.1 shows the association of the models in terms of AI as a big technological and study area and ML as a subfield of the AI and the DL as a particular scenario of the ML.

AI in health is a subspecialty, which includes different methods and techniques to address different challenges or problems in areas associated to health sciences. These new tools can be employed to strengthen task in assistance in clinical decisions [6], data mining in medicine [7–9], proposing extra-help that allows to provide an additional insight to professional staff in healthcare.

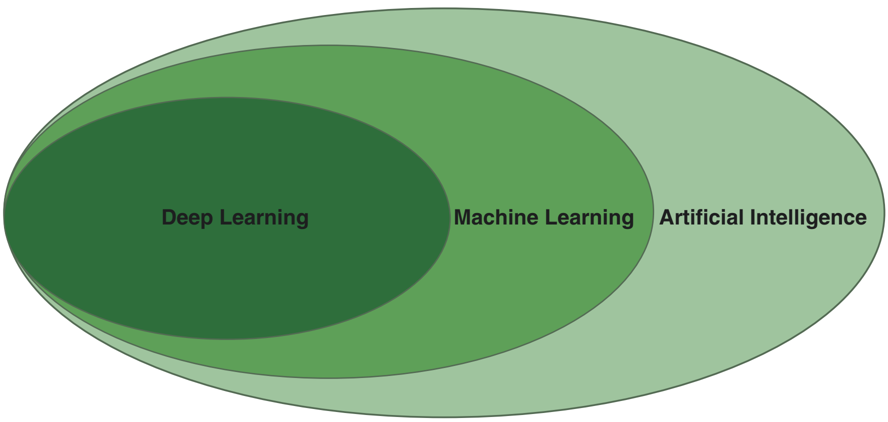


Fig. 2.1 Representation of the AI, ML, and DL

In medical specialties where medical staff spends lot of time analyzing images, AI offers alternatives that automatize these processes, for example, in radiology [10, 11], ophthalmology [12, 13], and, in general, diagnostic support [14–16].

2.2 Difference Between Machine Learning Versus Deep Learning

As mentioned earlier, the DL is a specific case of the ML when ANN are employed. This depends on the number of layers and, consequently, more parameters interconnecting these layers (Fig. 2.2). However, this section aims to unveil details about the differences between ML and DL.

It is important to describe some aspects of the data for training of the network. Thus, a previous stage with a focus in preprocessing must be included, which aims to reduce the number of variables or to choose the more relevant ones. This process is called feature extraction, and it is important for any ML classification or regression application.

In general, there are many tools designed to contribute in the feature extraction step, where the experience, limitations, and necessities of the problem are addressed. As the performance of the ANN model depends on the treatment of data, this stage takes the most time for a robust project development.

In contrast, DL attempts to automatize the feature extraction processes, wherein, the first layers of the model work for obtaining the parameters that be employed for last layers in the classification or regression issue. It is essential to point out that for these kinds of applications more data must be available to enhance all process and results. However, there are sophisticated techniques that allow to treat this problem, as bootstrapping methods or data augmentation for image applications.

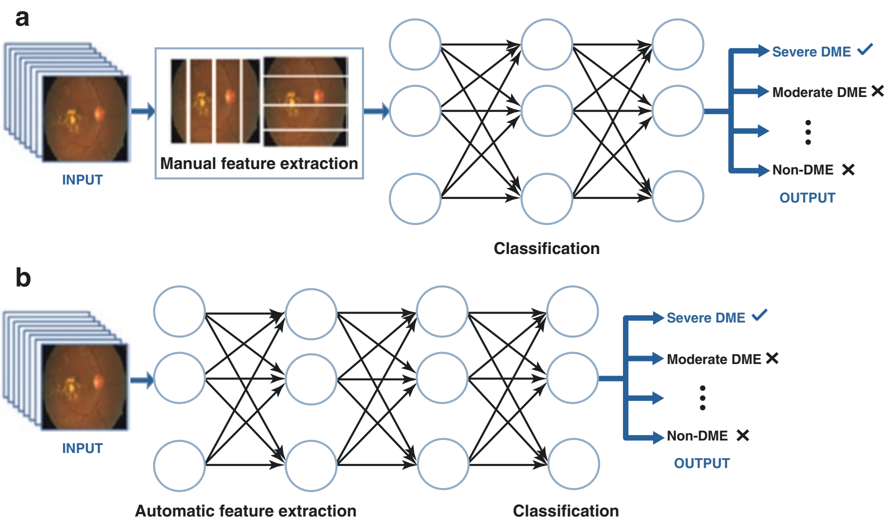


Fig. 2.2 ML and DL comparison: (**a**) ML and (**b**) DL

2.3 Machine Learning

Machine learning is defined as the area of study where the computers have the ability to extract relation and features of the data as a learning process, but without being programmed for this task. The automatic analysis of ocular images as a tool to support medical diagnosis has been a research challenge in terms of achieving the best model with the lowest computational cost [17–21]. In ophthalmology, the choice of the best method to represent, analyze, and make a diagnosis using eye fundus images is a complex computational problem [22–27].

2.3.1 Types of Learning Models for Machine Learning Methods

Machine learning algorithms are mainly organized according to the type of learning used by them. Supervised and unsupervised learning are widely used with ocular imaging and videos [17] while reinforcement learning is studied to automatically perform ocular surgery.

The differences between the three learning techniques are presented in the Fig. 2.3 and summarized as follows:

- *Supervised learning:* The algorithm is fed with inputs and outputs and has a function that maps the relationship between inputs and desired outputs.
- *Unsupervised learning:* The algorithm is fed with inputs and has to cluster in n-groups according to the similarity among the input data.
- *Reinforcement learning:* The algorithm is fed with actions and states, and it learns a policy of how to act, given a state.

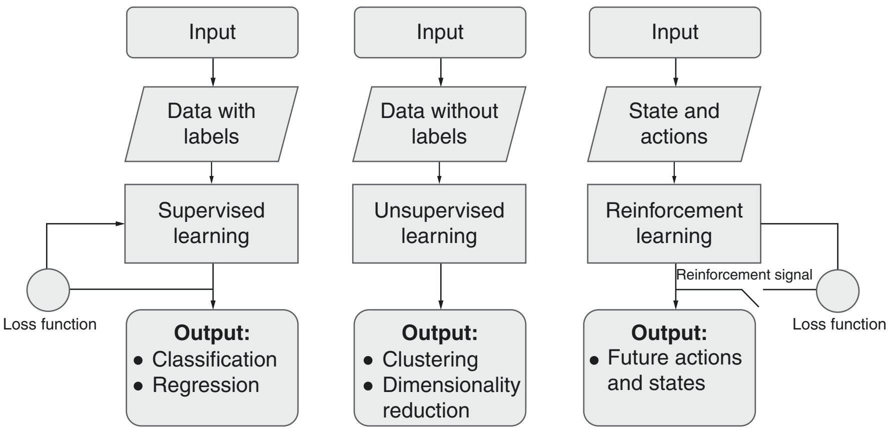


Fig. 2.3 Learning approaches of machine learning methods

In addition, ML techniques have been applied with some success to several eye conditions using as evidence individual sources of information [27–30]. Some researchers have studied how to support the diagnosis with different methodologies. Vandarkuhali and Ravichandran [27] detected the retinal blood vessels with extreme learning machine (ELM) approach and probabilistic neural network, Gurudath et al. [17] worked with ML identification from fundus images with a three-layered ANN and SVM to classify the retinal images, and Priyadarshini et al. studied clustering and classifications with data mining to give some useful predictions applied to diabetic retinopathy (DR) diagnosis [18]. Despite good results, the main problems with these results are that the datasets are small and the need for labels is expensive and cumbersome work.

2.4 Deep Learning

DL is a branch of ML that includes a whole family of algorithms with a common characteristic: an architecture organized by hierarchical levels. Its history dates back in the 1940s, with simple algorithms, that were variations of linear regression methods [31]. In the early 1960s, thanks to the inspiration produced by the nature of the cerebral cortex, Frank Rosenblatt introduced the fundamental pillar of the construction of neural networks: the perceptron. A perceptron is a basic processing unit, which receives information directly from the data to be analyzed, or from the output of other perceptrons. Each input has an associated connection weight and, in the simplest case, the output is a weighted linear combination of the inputs (Fig. 2.4).

ANN consist of a set of perceptrons (neurons) organized by layers. The whole structure can be divided into three sections: an input layer, a set of hidden layers, and an output layer. All the neurons are connected from one layer to the next in a way that the information moves through the network from the input to the output layer. The input information is transformed depending on the weights of the connections and the activation function of each neuron. An activation function acts as a

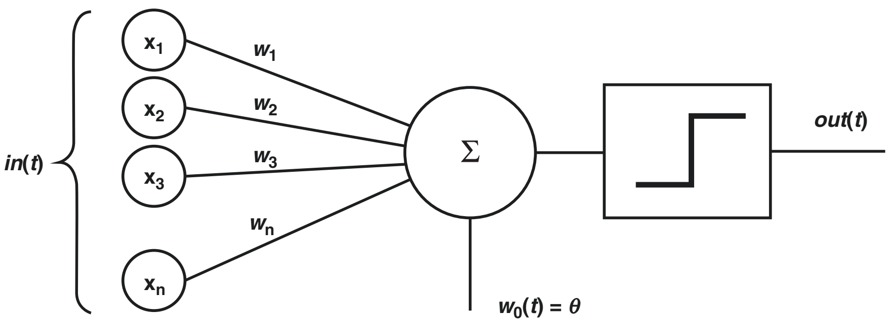


Fig. 2.4 Architecture of a simple perceptron

filter for the output of the neuron. As in most of the ML methods, the goal is to learn the parameters that better describe the data patterns and to reach this objective, a cost function should be minimized. The cost function could measure the error rate, or the precision of the predictions reached by the network, and it can vary depending on the model and the learning task. In any case, the cost function depends on the parameters of the network connections. In 1970, backpropagation was introduced as a gradient computing technique for the cost function minimization, based on Stochastic Gradient Descent method. Backpropagation tracks the error measured by the cost function in the opposite direction of the flow of information into the net, it means, from the output to the input layer, which allows the -necessary adjustment of the parameters to be calculated iteratively, as the model analyzes the training samples. Nowadays, backpropagation is part of the state of the art in the training process of neural networks [31].

DL is based on deep neural networks, using hundreds of hidden layers. Deep learning methods represent the state of the art in many fields of research [32], and thanks to the amount of efficient and specialized open-source software as Tensorflow, Keras, or Pytorch, they are the focus of an immense academic and industrial effort. Among the multiple options of DL models, one particular type stands out among medical imaging applications: the convolutional neural networks (CNNs).

2.4.1 Convolutional Neural Networks

CNNs were introduced in 1979 [33]. They are a particular type of neural network designed to recognize visual patterns (Fig. 2.5). CNNs include several hidden layers based on convolutional, subsampling, and/or normalization operators. This allows to exploit the structural information present in an image without the need to include an excessive number of trainable parameters.

The input of a CNN is a tensor (a multidimensional matrix) with shape defined by the image resolution. Convolutional layers are the core structure behind these models. They are composed of several convolutional kernels, which are rectangular matrices, whose size is a hyperparameter to be tuned by the user, and whose elements are parameters that the model must learn.

These kernels scan the image and output a new tensor from the linear combination of the kernel elements with the scanned input tensor elements. The new tensor is basically an image filtered by certain specific geometrical patterns, learned during training. These patterns increase in complexity as the depth of the network increases. Therefore, the first convolutional layers learn to recognize very local patterns, such as edges, and the following layers can learn to recognize structures created from combinations of the previous local patterns.

To date, CNNs are the most effective tool available for image analysis and have been successfully applied to medical imaging since 1995 [34], although they have only recently begun to be used in the study of eye diseases [35].

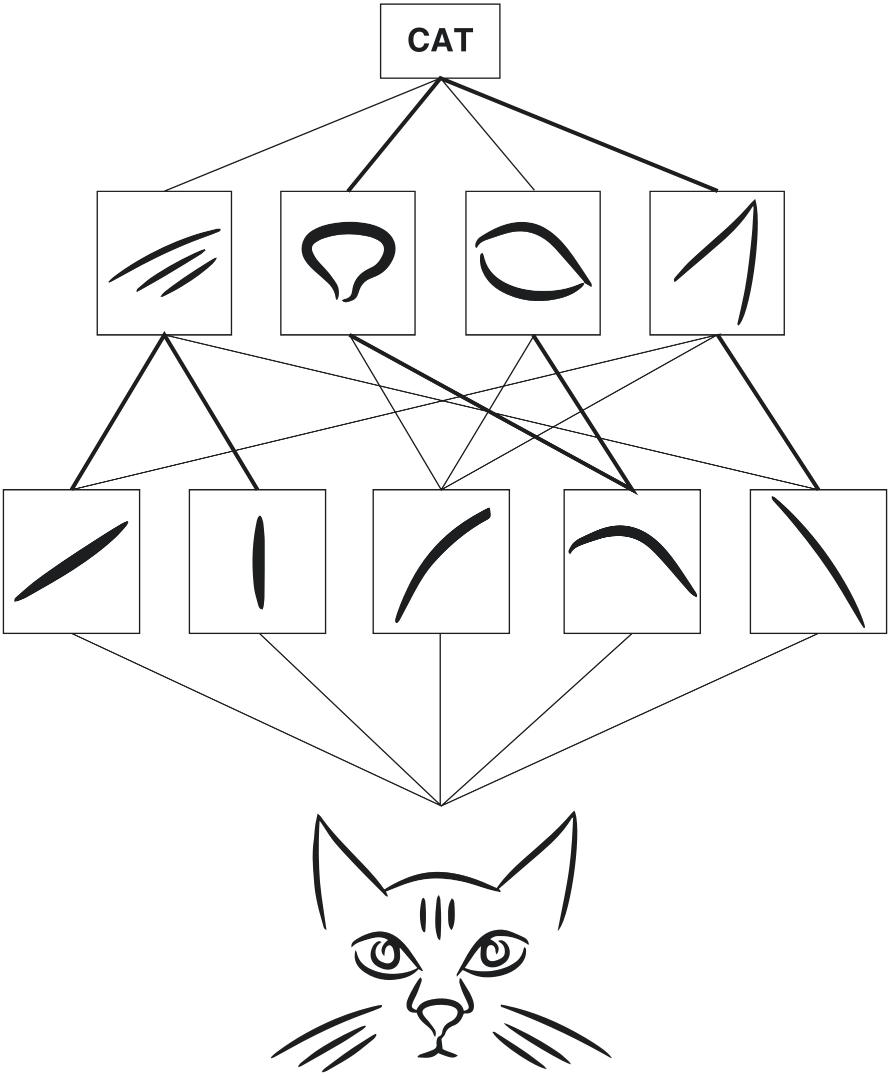


Fig. 2.5 Feature extraction performed by a CNN

2.4.2 Transfer Learning

In medial image analysis, most of the DL models make use of transfer learning, a technique that allows the researchers to take advantage of previously trained deep neural network models. Some of the most used DL models in the analysis of eye fundus images are Inception V1 and V3 [36]. Those models have been originally trained with ImageNet [37], one of the largest available datasets of natural images. While these models have shown good results in these image datasets, their direct application in medical image analysis is not feasible, due to several reasons. Natural and medical images have different and particular statistical characteristics. Therefore, the parameters of the models trained in ImageNet are not optimal for recognizing the specific patterns present in medical images. On the other hand, training these models with medical images datasets from scratch is not feasible, as they are usually very small sets. For eye fundus imaging, for example, one of the largest publicly available datasets is the EyePACS, which has only 35,126 images for training. Given the high number of parameters that deep models must optimize, the success of these DL models depends on the availability of very large volumes of data [38]. The transfer learning technique raises as an alternative to these problems. It consists of taking the models with the parameters learned with ImageNet (or with some other set of natural images as CIFAR) and then performing a fine-tuning process. It is, an adjustment of the parameters, continuing the learning process, using the specific data that we wish to analyze. This process has proven useful both for feature extraction, which could later be used in another ML model [38], and for the classification task. Although it is possible to create your own architectures, some studies have shown that under certain circumstances (such as poor availability of training data) the fine-tuning strategy gives better results than doing training from scratch [39].

Based on a classification of the most applied CNN, models to medical imaging can be made as follows: classification models, segmentation models, and multimodal architectures [40, 41].

2.4.3 Classification

The task of classifying healthy and sick patients, or classifying a disease among its different stages, is obviously important in supporting medical diagnosis. As mentioned earlier, the classification of ocular diseases based on eye fundus images focuses on fine-tuned CNNs. Inception V1 and Inception V3 are two of the most successful models used for this task. Inception V1 is a CNN that uses different sizes of convolutions for the same input. Furthermore, it includes a global average pooling layer at the end of its architecture. Inception V3 is an improved version that adds several layers of batch normalization and label smoothing strategies in order to prevent overfitting. The fine-tuned Inception V3 has reported impressive performance in the task of DR diagnosis [42]. However, it can also be used for representation learning. Among the ML methods, DL emerges as an unbeatable competitor for the tasks of representation learning. This is a fundamental factor in the success of any ML method. Models based on DL learn a representation space explicitly through the different processing levels, without much preprocessing [43]. By defining a hidden layer as the output of the Inception V3 model, it is possible to obtain a vector representation of the images, and use this representation as input for classical ML methods. In this way, the representational power of DL can be combined with the rigorous theoretical foundation of more classic and robust methods such as the SVM or probabilistic methods. This can allow to obtain in total more interpretable models, which not only offer a prediction but also, for example, indicators of uncertainty [38].

2.4.4 Segmentation

Segmentation models seek to recognize and isolate anatomical structures as organs or injuries within images. In many cases, segmentation is one of the first steps in the analysis tasks, prior to classification, since it precisely serves as a support to the diagnosis process. CNNs are used to make optic disc segmentation, useful for the diagnosis of glaucoma [44], segmentation of blood vessels and exudates for the monitoring of DR [45], or segmentation for the identification of druses useful in the diagnosis of Age-Related Macular Degeneration (AMD) [46]. Among the most common models used for segmentation tasks, there is the U-Net [47], a model that can be trained with few images, which consists of two parts, one of down sampling and another of upper sampling, interconnected in a symmetrical scheme, which allows the training to consider the full context of the image, and thus achieves a direct mapping between the original image and the segmented image.

Another common model is that of recurrent neural networks (RNN) [48], which involves recurrent connections and allows to temporarily store information from recent inputs. This is useful in the analysis of volumetric images, such as those obtained from an Optical Coherence Tomography (OCT) machine, because it allows finding patterns between sets of successive images related to each other.

2.4.5 Multimodal Learning

Multimodal learning attempts to find the best way to combine information from different sources, so that they complement each other and allow for better results than would be obtained by analyzing the sources of information separately. CNNs can be modified to receive not only an image but also more information in the form of extra channels [49].

Although the application of multimodal strategies with DL in the diagnosis of eye-related diseases has just begun, recent studies have shown that the combination of different sources of information generates models with greater predictive capacity and robustness [49]. This is the case with models designed to process volumetric OCTs, or those that combine eye fundus images with OCT volumes for the diagnosis of AMD [50]. More recently, most work has focused on the process of segmentation and classification of glaucoma and AMD. Golabbakhsh proposed the registration of the retina vessels combining information from eye fundus images and OCT. [51] Segmentation of retina vessels is useful in the diagnosis of AMD and DR. This combination of fundus images and OCT has been explored for the AMD diagnosis as well as for the segmentation of the optic cup and optic disc for glaucoma diagnosis.

However, the data from multiple modalities may result in the problem of appropriate representation and consequent fusion. How to learn the right representation for each information source? How and where to fuse the representation features into the model? Data from different modalities may have different statistical properties, so a simple concatenation of representative features is not necessarily a good strategy [52]. One must look for latent spaces, and around these various strategies can be designed.

Regarding the fusion, there are three types: early, late, and hybrid fusion. In early fusion, a selection or combination of the representations of each modality is made before solving the problem. In late fusion, the results of independent models are combined in order to draw conclusions. Deciding on the optimal type of fusion is part of the exploratory process in the application of DL methods. An example of this can be found in, where the authors showed that the extraction of several features, obtained separately from the same source, can be combined within the same model to provide complementary information. The authors experimented with merging information at different stages of the process and found the best results by intermediate fusion, it means, within the classifier. A similar concept was developed in [42], where structural and nonstructural features are extracted from the eye fundus images, and then correlated in a late fusion module.

But the combination of information can go beyond visual data, for instance, combining images and text. The first work on this subject was developed by Schlegl [53], combining CNN to analyze OCT images, with semantic information extracted from medical reports. The fusion there consisted of a concatenation of parameters into a fully connected layer, and they reported an improvement in the performance of retinal tissue classification tasks. It is also possible to combine visual information with morphological data, as proposed by Perdomo in [54], combining deep neural networks with morphological features in the detection and classification of glaucoma. Image analysis and extraction of morphological features can be done separately and mixed within the network in a layer, and then go through a fully connected layer and the output layer for the final decision.

References

1. Marks RJ II. Intelligence: computational versus artificial. IEEE Trans Neural Netw. 1993;4(5):737–9.
2. Bezdek JC. What is computational intelligence? In: Zurada J, Marks R, Robinson C, editors. Computational intelligence: imitating life. Piscataway: IEEE Press; 1994. p. 1–12.
3. Vapnik VN. Statistical learning theory: Wiley-Interscience; 1998. p. 521–6.
4. McLellan T, Endler JA. The relative success of some methods for measuring and describing the shape of complex objects. Syst Biol. 1998;47(2):264–81.
5. Raví D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ. Deep learning for health informatics. IEEE J Biomed Health Inform. 2017;21(1):4–21.
6. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med. 2012;157(1):29–43.
7. Houlton S. How artificial intelligence is transforming healthcare. Prescriber. 2018;29(10):13–7.
8. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y. Artificial intelligence in healthcare: past, present and future. Stroke Vascul Neurol. 2017;2(4):230–43.
9. Jorge Fernández Garcí A, Solveigh H, Angela S, Beck JP, Jenkins J. Transforming healthcare with AI. EIT and McKinsey and Company; 2020. https://eithealth.eu/wp-content/uploads/2020/03/EIT-Health-and-McKinsey_Transforming-Healthcare-with-AI.pdf.
10. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500–10.

11. Rulaningtyas R, Suksmono AB, Mengko TLR, Saptawati P. Colour segmentation of multi variants tuberculosis sputum images using self-organizing map. J Phys Conf Ser. 2017;853(1):012012.
12. Perdomo O, Andrearczyk V, Meriaudeau F, Müller H, González FA. Glaucoma diagnosis from eye fundus images based on deep morphometric feature estimation. In: Computational pathology and ophthalmic medical image analysis; 2018. p. 319–27.
13. Perdomo O, Rios H, Rodriguez FJ, Otálora S, Meriaudeau F, Müller H, et al. Classification of diabetes-related retinal diseases using a deep learning approach in optical coherence tomography. Comput Methods Prog Biomed. 2019;178:181–9.
14. Das N, Topalovic M, Janssens W. Artificial intelligence in diagnosis of obstructive lung disease: current status and future potential. Curr Opin Pulm Med. 2018;24(2):117–23.
15. de Oliveira E Souza Filho JB, Sanchez M, de Seixas JM, Maidantchik C, Galliez R, da Silva Rezende Moreira A, da Costa PA, Oliveira MM, Harries AD, Kritski AL, et al. Screening for active pulmonary tuberculosis: development and applicability of artificial neural network models. Tuberculosis. 2018;111:94–101.
16. Orjuela-Cañón AD, Mendoza JEC, García CEA, Vela EPV. Tuberculosis diagnosis support analysis for precarious health information systems. Comput Methods Prog Biomed. 2018;157:11–7.
17. Gurudath N, Celenk M, Riley HB. Machine learning identification of diabetic retinopathy from fundus images. In: IEEE signal processing in medicine and biology symposium (SPMB); 2014. p. 1–7.
18. Priyadarshini R, Dash N, Mishra R. A novel approach to predict diabetes mellitus using modified extreme learning machine. In: Electronics and communication systems (ICECS); 2014. p. 1–5.
19. Quellec G, Lamard M, Cazuguel G, Bekri L, Daccache W, Roux C, Cochener B. Automated assessment of diabetic retinopathy severity using content-based image retrieval in multimodal fundus photographs. Invest Ophthalmol Vis Sci. 2011;52(11):8342–8.
20. Welikala RA, Dehmeshki J, Hoppe A, Tah V, Mann S, Williamson TH, Barman SA. Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification. Comput Methods Prog Biomed. 2014;114(3):247–61.
21. Roychowdhury S, Koozekanani DD, Parhi KK. DREAM: diabetic retinopathy analysis using machine learning. IEEE J Biomed Health Inform. 2014;18(5):1717–28.
22. Usher D, Dumskyj M, Himaga M, WilliamsonT H, Nussey S, Boyce J. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. Diabet Med. 2004;21(1):84–90.
23. Philip S, Fleming AD, Goatman KA, Fonseca S, Mcnamee P, Scotland GS, Prescott GJ, Sharp PF, Olson JA. The efficacy of automated “disease/no disease” grading for diabetic retinopathy in a systematic screening program. Br J Ophthalmol. 2007;91(11):1512–7.
24. Chen SC, Huang YM. A novel approach to diagnose diabetes based on the fractal characteristics of retinal images. IEEE Trans Inf Technol Biomed. 2003;7(3):163–70.
25. Schaefer G, Leung E. An investigation into neural network for detection of exudates in retinal images. Appl Soft Comput. 2009;52:169–77.
26. Decencière E, Cazuguel G, Zhang X, Thibault G, Klein JC, Meyer F, et al. TeleOphta: machine learning and image processing methods for teleophthalmology. Ingènierie et Recherche Biomèdicale. 2013;34:196–203.
27. Vandarkuhali T. Ravichandran DRCS. ELM based detection of abnormality in retinal image of eye due to diabetic retinopathy. J Theor Appl Inf Technol. 2014;66(2):423–8.
28. Antal B, Hadju A. An ensemble-based system for automatic screening of diabetic retinopathy. Knowl-Based Syst. 2014;60:20–7.
29. Oh E, Keun Yoo T, Park EC. Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study. BMC Med Inform Decis Mak. 2013;13:106.
30. Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. Surv Ophthalmol. 2018;64(2):233–40.
31. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117.

32. Bietti A, Mairal J. Group invariance stability to deformations, and complexity of deep convolutional representations. J Mach Learn Res. 2019;20:1–49.
33. Fukushima K. Neocognition: a self-organized neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybernetics. 1980;202:193–202.
34. Lo S-CB, Lin J-S, Freedman MT, Chien MV, Lou S-LA, Mun SK. Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Trans Med Imaging. 1995;14(4):711–8.
35. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal. 42, 2017:60–88.
36. Perdomo O, González F. A systematic review of deep learning methods applied to ocular images. Ciencia e Ingenieria Neogranadina. 2019;30(1)
37. Deng J, Dong W, Socher R, Li LJ, Li K, FeiFei L. ImageNet: a large-scale hierarchical image database. In:CVPR09; 2009.
38. Toledo-Cortés S, de la Pava M, Perdomo O, González FA. Hybrid deep learning gaussian process for diabetic retinopathy diagnosis and uncertainty quantification. 2020. http://arxiv.org/abs/2007.14994.
39. Menegola A, Fornaciali M, Pires R, Avila S, Valle E. Towards automated melanoma screening: exploring transfer learning schemes. 2016. p. 1–4.
40. Zhou SK, Greenspan H, Shen D. Deep learning for medical image analysis. 2017. p. 1–433.
41. Müller H, Unay D. Medical decision support using increasingly large multimodal data sets. In: Big data analytics for large-scale multimedia search. 2019. p. 317–36.
42. Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. PLoS One. 2019;14(6):1–11.
43. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell. 2013;35(8):1798–828.
44. Claro MDL, Santos L, De M, Lima e Silva W, FHD DA, De Moura NH, Santana AM. Automatic glaucoma detection based on optic disc segmentation and texture feature extraction. CLEI Electr J. 2016;19(2):1–10.
45. Kumar Silpa A, Satheesh Kumar J. A review on recent developments for the retinal vessel segmentation methodologies and exudate detection in fundus images using deep learning algorithms. In: Smys S (Hrsg.), Tavares J (Hrsg.), Balas VE (Hrsg.), Iliyasu AM (Hrsg.). Computational vision and bio-inspired computing. Cham: Springer; 2020. p. 1363–70.
46. Checco P, Corinto F. CNN-based algorithm for drusen identification. In: Proceedings IEEE international symposium on circuits and systems. 2006. p. 2181–4.
47. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics. 2015. p. 234–41.
48. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. J Digit Imaging. 2019;32(4):582–96.
49. Andrearczyk V, Müller H. Deep multimodal classification of image types in biomedical journal figures. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). 2018. p. 3–14.
50. Yoo Tae K, Choi Joon Y, Seo Jeong G, Ramasubramanian B, Selvaperumal S, Kim DW. The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: a preliminary experiment. Med Biol Eng Comput. 2019;57(3):677–87.
51. Golabbakhsh M, Rabbani H. Vessel-based registration of fundus and optical coherence tomography projection images of retina using a quadratic registration model. IET Image Proc. 2013;7(8):768–76.
52. Vanegas JA. Large-scale non-linear multimodal semantic embedding largescale non-linear multimodal semantic embedding. Dissertation. 2017;3:14–20.

53. Schlegl T, Waldstein S, Vogl WD, Schmidt Erfurth U, Langs G. Predicting semantic descriptions from medical images with convolutional neural networks. Inform Proc Med Imag. 2015;9123(58):733–45.
54. Perdomo OJ, Arevalo J, González FA. Combining morphometric features and convolutional networks fusion for glaucoma diagnosis. In: Proceedings of the SPIE 10572, 13th international conference on medical information processing and analysis, 105721G. 2017. https://doi.org/10.1117/12.2285964.

3 Artificial Intelligence and Ophthalmology: An Overview

Parul Ichhpujani and Gagan Kalra

3.1 Introduction

As population aging has become a major demographic trend around the world, a steep increase is expected in patients suffering from ocular diseases. Early detection and appropriate treatment of ocular diseases are of great significance to prevent vision loss and improve quality of life. Conventional methods for diagnosing eye diseases primarily depend on the professional experience and knowledge of the ophthalmologist, which may result in high misdiagnosis rate and underutilization of medical data. An integration of ophthalmology and artificial intelligence (AI) has the potential to objectively revolutionize screening, diagnosing and management patterns of various ocular diseases. Applications of AI can make great contributions to provide support to patients in remote areas by sharing expert knowledge and limited resources.

Traditionally, an ophthalmic examination involves description of findings using words, drawings and images followed by establishing a diagnosis. This method of diagnosing is highly subjective, qualitative and inconsistent.

Since, majority of ophthalmic diagnosis is image-based, so a lot depends on the use of computers for analyzing and quantifying various parameters in the images. Research in medical image processing typically aims to extract features that might be difficult to assess with the naked eye. There are two types of features. The first is the well-known semantic feature defined by human experts, and the other is the agonistic feature defined by mathematical equations.

With increased access to big data and analytics and advancements in the neural network approach, the computers have helped in learning the combinations and

P. Ichhpujani (✉) · G. Kalra
Department of Ophthalmology, Government Medical College and Hospital, Chandigarh, India

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_3

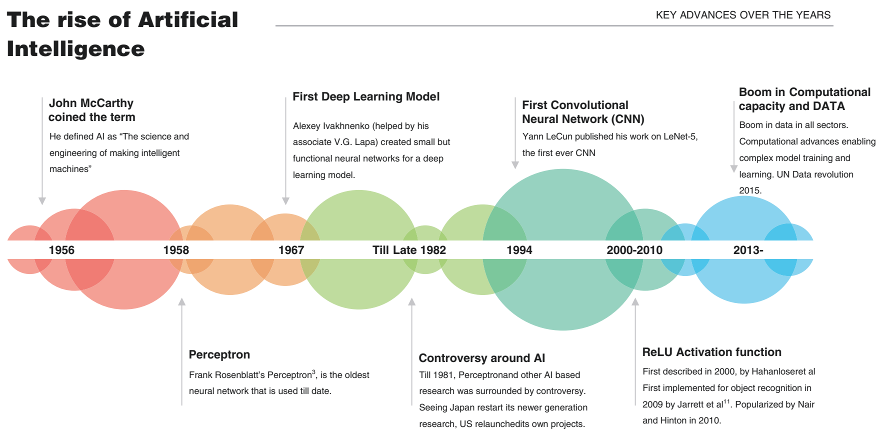


Fig. 3.1 The rise of artificial intelligence

permutations of important features [1]. This chapter gives an overview of AI (using both semantic and agonistic features) for early detection and treatment of anterior as well as posterior segment disorders. Figure 3.1 shows the timeline of key advances of AI.

3.2 Artificial Intelligence and Anterior Segment Disorders

3.2.1 Artificial Intelligence and Ectasia

With the boom in refractive surgeries, iatrogenic keratectasia has been noted in cases without anterior surface alterations. This highlights the need to screen for susceptible cases likely to have a post procedure biomechanical failure. For developing a screening technique/protocol, several methods such as risk scores, linear models and artificial intelligence models have been explored.

Lopes et al. collected Pentacam HR (Oculus, Wetzlar, Germany) data of three groups of patients viz., those with stable LASIK, post LASIK ectasia and clinical keratoconus from multiple centres of three countries to develop a 'Pentacam Random Forest Index' (PRFI) which had a sensitivity of 94.2% and specificity of 98.8% to detect corneal ectasia with an AUC of 0.992 [2].

Yoo et al. predicted candidacy for corneal refractive surgery by analyzing data from patient demographics, corneal tomography and ophthalmic examination using five different machine learning algorithms. They used an ensemble classifier algorithm which was validated using internal and external datasets with an AUC of 0.983 and 0.972, respectively [3].

AI-derived models bridge the gap between lab and clinic, by providing a simple output parameter which serves as a risk profiling tool. Different corneal tomographers and biomechanical analyzers have added these AI-based indices in the software as objective screening parameters.

3.2.2 Artificial Intelligence and Keratoconus

Artificial intelligence for keratoconus detection is an active area of research with many different methods implemented over the years to detect presence of the disease, differentiate normal from *form fruste* keratoconus and even classify severity grade of the disease in some studies [4].

Kamiya et al. used deep learning (DL) to classify eyes into keratoconus present or absent based on arithmetic calculation of output data from colour-coded corneal maps on swept-source anterior segment optical coherence tomography (AS-OCT). Six different colour-coded maps were used to generate AUC of 0.991 for detection of keratoconus and classification of grade of the disease in keratoconic eyes [5].

Valdés-Mas et al. studied corneal curvature and astigmatism in keratoconus patients post intra-corneal ring implantation to predict visual outcome using machine learning [6].

Yousefi et al. studied corneal parameters from both the anterior and posterior surface, grouped them into principal components by conducting linear transformation before conducting non-linear tSNE (t-Distributed Stochastic Neighbour Embedding) transformation, identified unique non-overlapping clusters using unsupervised machine learning and conducted a post-hoc analysis for these clusters to predict likelihood of requiring a future keratoplasty intervention [7].

3.2.3 Artificial Intelligence and Lens

3.2.3.1 Cataract Grading

Cataracts are the leading cause of visual impairment worldwide, accounting for over 50% of cases of blindness in middle- and low-income countries. As mentioned earlier in the chapter, with the global trend of increased longevity, the prevalence of cataracts is also going to increase. Unfortunately, the distribution of medical resources in the developing countries is not satisfactory for cataract diagnosis and appropriate surgical management. Therefore, a universal AI platform is needed for the management of cataracts involving multiple clinical scenarios and improving medical resource coverage.

Kumar M and Gunasundari developed a computer-aided diagnosis system to detect corneal arcus and cataract from the photographs of eyes taken with a standard digital camera [8]. They developed a multiclass computer-aided diagnosis (CAD) system using visible wavelength (VW) eye images to diagnose anterior segment ocular conditions. They pre-processed the input VW eye images for specular reflection removal and then segmented the iris circle region using a Circular Hough Transform (CHT)-based approach.

From the segmented iris circle, first-order statistical features and wavelet-based features were extracted and used for classification. They achieved a predictive accuracy of 96.96% with 97% sensitivity and 99% specificity.

Gao and Wong proposed an automated system to ascertain features for grading the severity of nuclear cataracts from slit-lamp images [9]. Image patches from lenses within the same grading class were clustered to develop local filters. These filters were fed into a convolutional neural network (CNN), followed by a set of recursive neural networks, in order to extract higher order features. Support vector regression was applied to higher order features to grade the cataract.

Currently, no universal AI tool or platform is available that can recognize different capture modes, aetiologies and stages of treatment for cataract.

3.2.3.2 Intraocular Lens Power Calculation

The erstwhile simple cataract extraction has now evolved to ‘refractive cataract surgery’ as now the aim is to neutralize the patient’s pre-existing refractive error at the time of addressing the lenticular opacity. By using the axial length and keratometry, only about 80% of patients can achieve within 0.5 Dioptres (D) of the intended refractive target using third-generation static formulae such as SRK/T, Holladay 1 and Hoffer Q. These formulae consist of a single equation for all eyes. The more recent formulae, such as the Holladay 2, the Barrett Universal, the Hill-RBF and the Ladas Super Formula 1.0, are probably better classified as methodologies because they are more than just single equations and involve use of additional inputs such as anterior chamber depth and white-to-white diameter.

Recently, DL has been applied to an amalgamation of multiple IOL calculation formulae with Ladas Super Formula 1.0 as framework, to develop a new improved formula, Ladas 2.0. Using data sets to learn from multiple surgeons and then applying it to another surgeon, Ladas 2.0 AI gives an accuracy of 87% and if the same surgeon’s data is used to improve his/her own calculations, then its accuracy improves to 94% [10].

González et al. deployed 11 DL models for IOL power calculation of which SVM with Gaussian kernel RBF and multivariate analysis regression spline (MARS) had the best results and an ensembled model was generated by combining the two models, Karmona. This was better than other models as it incorporated the ratio between the curvatures of the posterior and anterior corneal surfaces. Best results were obtained with Karmona stacked regression model, where 90.38% and 100% of eyes were within ±0.50 and ± 1.00 D, respectively [11].

AI-based self-calibrating biometers are going to revolutionize the future of ‘refractive cataract surgery’.

3.3 Artificial Intelligence and Posterior Segment Disorders

3.3.1 Artificial Intelligence and Diabetic Retinopathy

About 600 million people are likely to have diabetes by 2040, and nearly one-third will have diabetic retinopathy (DR) [12].

Screening for DR is performed by different eyecare and technical professionals, including ophthalmologists, optometrists, clinical photographers and screening technicians. Any of the screening methods such as direct ophthalmoscopy, dilated slit-lamp biomicroscopy with a handheld lens, mydriatic or non-mydriatic fundus photography, teleretinal screening and retinal video recording may be used for clinic or community. A major hurdle in robust implementation of DR screening programmes is related to availability of trained human assessors and long-term financial sustainability.

Deep learning has revolutionized the diagnostic performance for detecting DR. The goal is to have such a DL system that can be generalized to populations of different ethnicities, and works well with retinal images captured using different cameras.

El Tanboly et al. developed a DL-based computer-aided system to detect DR through 52 OCT images, achieving an AUC of 0.98 [13]. Despite the good outcomes in the cross-validation process, the system needs to be further validated in larger patient cohorts. A computer-aided diagnostic (CAD) system based on continuous machine learning (CML) algorithms using optical coherence tomography angiography (OCTA) images to automatically diagnose non-proliferative DR (NPDR) also achieved high accuracy and AUC [14, 15].

The results of highly biased and heterogeneous studies assessing the diagnostic performance of OCTA highlight the need for further analyses of methodologically sound and sufficiently sized clinical evaluations.

Algorithms for the diagnosis of diabetic retinopathy have been amongst the first to receive regulatory approval for routine clinical use [16].

3.3.2 Artificial Intelligence and Retinal Vein Occlusion

Automated detection of branch retinal vein occlusion (BRVO) was attempted by Zhang et al. with hierarchical local binary pattern (HLBP) recognition and maximal pooling as a feature extraction method. AUC was noted to be 0.961 [17].

Nagasato et al. used ultra-wide field fundus photographs from central retinal vein occlusion (CRVO) patients and normal patients to train a DL convolutional neural network and a support vector machine (SVM) to detect presence of CRVO. They found that the DL model had better performance than SVM model with AUC of 0.989 with sensitivity and specificity of 98.4% and 97.9%, respectively. Same group performed another study using ultra-wide field fundus photographs from BRVO and normal patients to train DL and SVM models to compare performance. The AUC for DL model that outperformed the SVM model was 0.976 with sensitivity and specificity of 94.0% and 97.0%, respectively [18, 19].

A random forest model was implemented to classify images based on presence of vitreomacular adhesion (VMA) in BRVO patients as it is an important biomarker to predict response to anti-VEGF therapy. It was noted that eyes with VMA had more visual acuity gains with therapy as compared to eyes that did not have VMA [20].

3.3.3 Artificial Intelligence and Retinopathy of Prematurity

Retinopathy of prematurity (ROP) is one of the leading causes of preventable childhood blindness globally. Blindness from ROP is largely preventable with early case detection and timely treatment [21].

Although improvement in neonatal care in middle-income 'developing' countries has resulted in improved survival of premature infants but unfortunately infrastructure needed for ROP case detection and treatment has not improved proportionately. ROP screening is a time-consuming practice that requires trained personnel. Interclinician subjectivity and regional variation in the diagnosis of different stages of ROP, especially the plus disease has led to delayed treatment and poorer visual outcomes [22].

Attempts have been made to improve ROP screening by digital retinal imaging and remote diagnosis via telemedicine by trained ophthalmologists. In recent times, automated techniques using AI deep learning technologies have been validated for diagnosis of ROP using retrospective data, but are yet to be fully validated in the real-world setting [23, 24].

In order to improve the generalizability of the DL algorithm, technical variations in the fundus imaging process such as different camera models, lenses and optical aberrations must be accounted for.

3.3.4 Artificial Intelligence and Age-Related Macular Degeneration

Age-related macular degeneration (ARMD) is a chronic and irreversible macular disease characterized by drusen, retinal pigment changes, choroidal neovascularization, haemorrhage and geographic atrophy. It is one of the leading causes of central vision loss in people over 50 years. With the population aging and the severity of ARMD, it is necessary to perform regular screening. Automatic ARMD diagnosis may obviously reduce the workload of clinicians and improve efficiency.

Lee et al. used a VGG16 CNN to analyze central 11 images from the OCT volumes of normal and ARMD patients to automatically detect ARMD. The data was analyzed for individual images, averaged probabilities from comprising images from OCT volumes and averaged probabilities from all scans of individual patients showing AUC of 0.928, 0.938 and 0.975, respectively [25].

Burlina et al. used deep learning to detect and grade ARMD. In a follow-up study, they deployed deep learning to classify ARMD according to the Age Related Eye Disease Study (AREDS) Severity Scale and estimate five-year risk of progression to advanced disease by soft prediction, hard prediction and DL-based regression mapping. They found that mean estimation error in overall five-year risk ranged between 3.4% and 5.8% with higher AREDS classes having a higher mean error and vice versa [26, 27].

DL and convolutional neural networks were deployed by Grassmann et al. to detect and classify ARMD into 13 classes using colour fundus photographs [28]. They deployed an ensemble random forest model of six different convolutional neural network architectures to achieve an accuracy of 0.943 and quadratic weighted k of 92% for classification into 13 classes. Recently, Peng et al. used a newer DeepSeeNet model to classify colour fundus photographs into ARMD severity classes [29].

DeepSeeNet closely resembles human grading process by initially taking into account the risk factors for ARMD and then classifying according to the AREDS Simplified Severity Scale. This model was shown to have better accuracy compared to retina specialists for patient-based ARMD classification (0.671 vs 0.599) with AUC of 0.94, 0.93 and 0.97 for large drusen, pigmentary abnormalities and detection of late ARMD, respectively.

3.3.5 Artificial Intelligence and Glaucoma

AI can help revolutionize the screening, diagnosis and classification of glaucoma. AI allows automated processing of large data sets, and detection of new disease patterns. Initially, fundus photographs were processed via machine learning to identify glaucomatous optic nerve damage [30, 31]. Later, larger databases were processed with DL technology [32]. Ting et al. analyzed a database of 125,189 fundus photographs and reported a sensitivity of 96.4% and specificity of 87.2% [33].

Additionally, AI applications are being developed using computerized visual field and OCT data, studies have also been published describing programs that are able to evaluate patients based on data from both of these examination devices.

3.3.6 Artificial Intelligence and Retinal Detachment

Researchers in Japan used a DL algorithm to detect rhegmatogenous retinal detachment from Optos ultra-widefield fundus images. It demonstrated a sensitivity of 97.6% and specificity of 96.5% [34].

3.3.7 Artificial Intelligence and Geographic Atrophy

Advanced AI methods in high-resolution retinal imaging allow to identify, localize and quantify biomarkers such as hyper reflective foci (HRF). Increased HRF concentrations in the junctional zone and future macular atrophy may represent progressive migration and loss of retinal pigment epithelium [35]. AI-based biomarker monitoring may pave the way into the era of individualized risk assessment and objective decision-making processes.

3.4 Artificial Intelligence and Miscellaneous Ocular Disorders

3.4.1 Refractive Error Calculation

Varadarajan et al. deployed a combination of ResNet and Soft attention DL architecture to predict refractive error from features extracted retinal fundus images with a mean absolute error (MAE) of 0.56D (95% confidence interval) and 0.91D (95% confidence interval) on the two tested data sets [36].

Chun et al. conducted deep learning-based refractive error estimation using smartphone captured eccentric photorefractive images in paediatric age group with an overall estimation accuracy of 81.6% [37].

Subjective refraction data was split into power vectors (M, J0 and J45) which were then predicted using three boosted gradient trees (XGBoost) algorithms by training using patient data. It was noted that this model yielded more accurate results compared to paraxial matching method for spectacle correction. The mean absolute error with this model was noted to be 0.301 ± 0.252 D for the M vector, 0.120 ± 0.094 D for the J0 vector and 0.094 ± 0.084 D for the J45 vector [38].

3.4.2 Ocular Oncology

Damato et al. deployed conditional hazard estimating neural network (CHENN) to estimate survival in patients with choroidal melanoma and compared results with standard Kaplan-Meier analysis [39]. They found that the all-cause survival curves matched between the two methods ($p < 0.05$), except in older patients where CHENN estimated lower mortality than Kaplan-Meier analysis.

Nguyen et al. described a two-stepped approach to achieve automated uveal melanoma segmentation. The first step comprising of a class activation map, conditional random field and active shape model was deployed to localize the tumour in the MRI scan. The second step deployed a 2D-Unet CNN for tumour segmentation [40].

Sun et al. classified uveal melanomas for BAP1 expression with AUC of 0.99 by using histopathological slides, creating 8176 patches of 256 × 256 pixels each and feeding them through a dense CNN. BAP1 expression is associated with poor prognosis in these tumours as it confers metastatic potential [41].

3.4.3 Paediatric Ophthalmology

As described above, automated refraction using eccentric photorefractive images has a huge potential in the paediatric age group. In a recent review, retinopathy of prematurity (also described above) was noted to have high yield from AI-based detection and grading with accuracy approaching that of experts [42]. Other areas in paediatric ophthalmology that have had recent expansion in AI-based applications include paediatric cataract detection, cataract classification, prediction of complications post-cataract surgery, strabismus detection, prediction of potential high myopia development, fundus vessel segmentation and visual development analysis to name a few.

To conclude, by building systematic and interpretable AI platforms using advanced techniques with sufficient multimodal and high-quality data, the applicability of AI in clinical scenarios can be enhanced.

References

1. Abràmoff MD. Image processing. In: Schachat AP, Wilkinson CP, Hinton DR, Sadda SR, Wiedemann P, editors. Ryan's Retina. 6th ed. Amsterdam: Elsevier; 2017. p. 196–223.
2. Lopes BT, Ramos IC, Salomão MQ, Guerra FP, Schallhorn SC, Schallhorn JM, et al. Enhanced tomographic assessment to detect corneal ectasia based on artificial intelligence. Am J Ophthalmol. 2018;195:223–32.
3. Yoo TK, Ryu IH, Lee G, Kim Y, Kim JK, Lee IS, Kim JS, Rim TH. Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. NPJ Digit Med. 2019;2(1):1–9.
4. Lin SR, Ladas JG, Bahadur GG, Al-Hashimi S, Pineda R. A review of machine learning techniques for keratoconus detection and refractive surgery screening. Semin Ophthalmol. 2019;34(4):317–26.
5. Kamiya K, Ayatsuka Y, Kato Y, Fujimura F, Takahashi M, Shoji N, Mori Y, Miyata K. Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: a diagnostic accuracy study. BMJ Open. 2019;9(9):e031313.
6. Valdés-Mas MA, Martín-Guerrero JD, Rupérez MJ, Pastor F, Dualde C, Monserrat C, Peris-Martínez C. A new approach based on machine learning for predicting corneal curvature (K1) and astigmatism in patients with keratoconus after intracorneal ring implantation. Comput Methods Prog Biomed. 2014;116(1):39–47.
7. Yousefi S, Takahashi H, Hayashi T, Tampo H, Inoda S, Arai Y, et al. Predicting the likelihood of need for future keratoplasty intervention using artificial intelligence. Ocul Surf. 2020;18(2):320–5.
8. Mahesh K, Gunasundari R. Computer-aided diagnosis of anterior segment eye abnormalities using visible wavelength image analysis based machine learning. J Med Syst. 2018;42:128.
9. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. IEEE Trans Biomed Eng. 2015;62:2693–701.
10. Kane JX, Van Heerden A, Atik A, Petsoglou C. Accuracy of 3 new methods for intraocular lens power selection. J Cataract Refract Surg. 2017;43(3):333–9.
11. González DC, Bautista CP. Accuracy of a new intraocular lens power calculation method based on artificial intelligence. Eye. 2020;28:1–6.
12. Yau JW, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35:556–64.
13. ElTanboly A, Ismail M, Shalaby A, Switala A, El-Baz A, Schaal S, Gimel'farb G, El-Azab M. A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images. Med Phys. 2017;44(3):914–23.
14. Sandhu HS, Eladawi N, Elmogy M, Keynton R, Helmy O, Schaal S, El-Baz A. Automated diabetic retinopathy detection using optical coherence tomography angiography: a pilot study. Br J Ophthalmol. 2018;102(11):1564–9.
15. Faes L, Bodmer NS, Locher S, Keane PA, Balaskas K, Bachmann LM, Schlingemann RO, Schmid MK. Test performance of optical coherence tomography angiography in detecting retinal diseases: a systematic review and meta-analysis. Eye (Lond). 2019;33(8):1327–38.
16. Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. 2018. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye.
17. Zhang H, Chen Z, Chi Z, Fu H. Hierarchical local binary pattern for branch retinal vein occlusion recognition with fluorescein angiography images. Electron Lett. 2014;50(25):1902–4.
18. Nagasato D, Tabuchi H, Ohsugi H, Masumoto H, Enno H, Ishitobi N, Sonobe T, Kameoka M, Niki M, Hayashi K, Mitamura Y. Deep Neural Network-Based Method for Detecting Central Retinal Vein Occlusion Using Ultrawide-Field Fundus Ophthalmoscopy. J Ophthalmol. 2018:1875431.

19. Nagasato D, Tabuchi H, Ohsugi H, Masumoto H, Enno H, Ishitobi N, Sonobe T, Kameoka M, Niki M, Mitamura Y. Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion. Int J Ophthalmol. 2019;12(1):94.
20. Waldstein SM, Montuoro A, Podkowinski D, Philip AM, Gerendas BS, Bogunovic H, Schmidt-Erfurth U. Evaluating the impact of vitreomacular adhesion on anti-VEGF therapy for retinal vein occlusion using machine learning. Sci Rep. 2017;7(1):1–1.
21. Early Treatment for Retinopathy Of Prematurity Cooperative Group. Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. Arch Ophthalmol. 2003;121:1684–94.
22. Fleck BW, Williams C, Juszczak E, et al. An international comparison of retinopathy of prematurity grading performance within the Benefits of Oxygen Saturation Targeting II trials. Eye. 2018;32:74–80.
23. Brown JM, Campbell J, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol. 2018;136:803–10.
24. Wang J, Ju R, Chen Y, et al. Automated retinopathy of prematurity screening using deep neural networks. Ebio Med. 2018;35:361–8.
25. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. Ophthalmol Retina. 2017;1(4):322–7.
26. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA Ophthalmol. 2017;135(11):1170–6.
27. Burlina PM, Joshi N, Pacheco KD, Freund DE, Kong J, Bressler NM. Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration. JAMA Ophthalmol. 2018;136(12):1359–66.
28. Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, Peters A, Heid IM, Palm C, Weber BH. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology. 2018;125(9):1410–20.
29. Peng Y, Dharssi S, Chen Q, Keenan TD, Agrón E, Wong WT, Chew EY, Lu Z. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. Ophthalmology. 2019;126(4):565–75.
30. Singh A, Dutta MK, ParthaSarathi M, Uher V, Burget R. Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image. Comput Methods Prog Biomed. 2016;124:108–20.
31. Salam AA, Khalil T, Akram MU, Jameel A, Basit I. Automated detection of glaucoma using structural and non structural features. Springerplus. 2016;5:1519.
32. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology. 2018;125:1199–206.
33. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 2017;318:2211–23.
34. Ohsugi H, Tabuchi H, Enno H, Ishitobi N. Accuracy of deep learning, a machine-learning technology, using ultra-wide- field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. Sci Rep. 2017;7:9425.
35. Schmidt-Erfurth U, Bogunovic H, Grechenig C, Bui P, Fabianska M, Waldstein S, Reiter GS. Role of deep learning quantified hyperreflective foci for the prediction of geographic atrophy progression. Am J Ophthalmol. 2020; pii: S0002-9394(20)30154-9.
36. Varadarajan AV, Poplin R, Blumer K, Angermueller C, Ledsam J, Chopra R, Keane PA, Corrado GS, Peng L, Webster DR. Deep learning for predicting refractive error from retinal fundus images. Invest Ophthalmol Vis Sci. 2018;59(7):2861–8.
37. Chun J, Kim Y, Shin KY, Han SH, Oh SY, Chung TY, Park KA, Lim DH. Deep learning–based prediction of refractive error using photorefraction images captured by a smartphone: model development and validation study. JMIR Med Inform. 2020;8(5):e16225.

38. Rampat R, Debellemanière G, Malet J, Gatinel D. Using artificial intelligence and novel polynomials to predict subjective refraction. Sci Rep. 2020;10(1):1–9.
39. Damato B, Eleuteri A, Fisher AC, Coupland SE, Taktak AF. Artificial neural networks estimating survival probability after treatment of choroidal melanoma. Ophthalmology. 2008;115(9):1598–607.
40. Nguyen HG, Pica A, Rosa FL, Hrbacek J, Weber DC, Schalenbourg A, Sznitman R, Cuadra MB. A novel segmentation framework for uveal melanoma based on magnetic resonance imaging and class activation maps. In: Proceedings of the 2nd international conference on medical imaging with deep learning, PMLR. 2019; 102:370–379.
41. Sun M, Zhou W, Qi X, Zhang G, Girnita L, Seregard S, Grossniklaus HE, Yao Z, Zhou X, Stålhammar G. Prediction of BAP1 expression in uveal melanoma using densely-connected deep classification networks. Cancers. 2019;11(10):1579.
42. Reid JE, Eaton E. Artificial intelligence for pediatric ophthalmology. Curr Opin Ophthalmol. 2019;30(5):337–46.

4 Artificial Intelligence in Cornea and Refractive Surgery

Sartaj Singh Grewal and S. P. S. Grewal

4.1 Introduction

In the field of ophthalmology, the subspecialty of cornea has seen pioneering work in the evolution of technological advances aimed at aiding diagnosis and treatment. This dates back to the beginning of the fifteenth century, with Christoph Scheiner's experiments using reflections of images from the cornea [1]. Over time, advancements in technology have enabled the development of instruments such as the keratometer, keratoscope, pachymeters, tomographer, confocal microscopes, corneal hysteresis, meibography and wavefront analyzers [2–4]. The components of a comprehensive ocular surface evaluation today include a tomographic assessment of cornea, biomechanical measurements for ocular hysteresis, optical wavefront analysis and tear film evaluation including meibomian gland function.

Busy clinics with an increasing demand for refractive surgery have led to a paradigm shift in practice patterns. The clinician has to his disposal a wide range of tools to perform a comprehensive qualitative and quantitative evaluation of the ocular surface. However, the interpretation of the scans requires a thorough assessment and due to the vast amount of data generated by modern instruments, this may be time consuming. The approach has to be individualized given the huge amount of naturally occurring variations. In addition, detection of pathology at a subclinical stage can help in timely and more effective management while avoiding complications. The tools for evaluation remain the same but the diagnostic and therapeutic threshold criteria for different disorders vary—from selecting candidates for refractive surgery to diagnosing meibomian gland dysfunction or detecting early corneal ectasias.

S. S. Grewal (✉) · S. P. S. Grewal
Grewal Eye Institute, Chandigarh, India

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_4

The overwhelming amount of information provided by these modern instruments can at times be difficult to interpret, and prevent us from arriving at definitive conclusions and decisions. An algorithmic approach has shown to reduce chances of oversight and human errors by defining step-by-step protocols to interpret all the available data and detect red flags. With an ever-increasing workload, these human errors are unfortunately not completely avoidable.

We need improved strategies to extract, process and correlate meaningful information from the various complementary instruments, and artificial intelligence (AI) is the most efficient means to accomplish this. Convolutional neural networks (CNNs), deep learning (DL) and other machine learning (ML) techniques are becoming vital tools to help clinicians deliver the best quality of care to patients while streamlining data analytics. Once a comprehensive work up has been done, if we are able to feed the data and images from the scans onto an AI system, which allows us to check for subtle variations in measurements, recognize patterns prior to them being clinically visible as well as detect and screen out those patients having "red flag" measurements which require additional attention.

The long-term aim is to ease the interpretation of investigations in a time-efficient manner and minimize risk of errors and subsequent complications. Second aim is to provide a systematic approach to objectively monitor a disease process and its response to treatment. Within the corneal subspecialty, refractive surgery and progressive ectatic disorders have seen the most significant development in algorithms and implementation of ML for improving patient care. This chapter aims to discuss the existing role of AI in cornea and refractive surgery and scope for further development.

4.2 Artificial Intelligence for Ectasia Diagnosis

Keratoconus (KCN) is a progressive, usually bilateral corneal ectasia which varies in prevalence from 1 in 50 people in Central India, to 1 in 2000 people in the United States [5, 6]. The biggest challenge in KCN is to achieve an early diagnosis, when the patient is still asymptomatic. Very early diagnosis is essential due to two main reasons, firstly, the progressive nature of KCN and the fact that treatments such as crosslinking only allow us to halt progression, and secondly, to accurately screen the large number of patients undergoing refractive surgery who are at risk of iatrogenic ectasia. The first report of progressive ectasia after LASIK occurred in a case with subclinical *forme fruste* KCN [7]. Such patients are susceptible to biomechanical failure of the cornea following the removal of corneal tissue by laser vision correction (LVC). The challenge is to identify patients with KCN even before any anterior corneal surface alterations are evident.

To accomplish the goal of early identification of such cases, various methods including risk scorers, linear models and more recently artificial intelligence and machine learning models evaluating data from different tomographers have been proposed. One of the first published use of AI in the analysis of corneal abnormalities was done by Maeda et al. in 1995 [8]. They selected topographic maps obtained from videokeratography and subsequently corneal experts classified these into seven categories: normal, with-the-rule astigmatism, KCN (mild, moderate, advanced), post photorefractive keratectomy and post keratoplasty. The maps were divided into a training set (108 maps) and a test set (75 maps). For each map, 11 topography-characterizing indices that were calculated from the data provided by the videokeratoscope, along with the corresponding diagnosis category, were used to train a neural network. Correct classification was achieved by the trained neural network for all 108 maps in the training set. In the test set, the neural network correctly classified 60 of 75 maps (80%). For every category, accuracy and specificity were greater than 90%, whereas sensitivity ranged from 44% to 100%. They used the backpropagation method as a neural network model which consists of three or more layers of "neurons", that form the basic processing unit. The layers include an input layer, at least one hidden layer and an output layer. Each neuron in one layer is connected with each neuron in the following layer. The input layer consists of investigator defined neurons, as in the data that is provided, and the output layer consists of neurons corresponding to the different classification categories. In the training process, the neural network compares its output response with the correct response and produces an error value, which is then further used to adjust the weighting scheme and minimize output error. The aim of training a neural network with training data is to allow the system to memorize the relationship between input and output data as a matrix of adjusted weights. One limitation in such neural networks is that the accuracy of results depends on the size and accuracy of the data set.

In 1997, Smolek et al. further developed two networks to answer two questions by analyzing videokeratoscope examinations: "Is keratoconus present?" and "How severe is it?" [9] They trained it to classify the examinations into nine different categories: Normal, astigmatism, KC, KCS, contact lens induced warpage, pellucid marginal degeneration, photorefractive keratectomy, radial keratotomy and penetrating keratoplasty. One separate network was then trained to grade KC maps as mild (KC1), moderate (KC2) or advanced (KC3). The study used a single hidden layer and both networks were trained to an error tolerance of 0.1. Three hundred examinations were randomly divided into 150 examinations each for training and test set. On the test set, the classification network was 100% accurate, sensitive and specific. The network outperformed the diagnostic methods available on the TMS-2 machine (Tomey USA, Cambridge, MA) at the time. The network was as accurate as the Klyce/Maeda keratoconus Index (KCI) and the Rabinowitz test.

Smolek et al., in 2001, then trained a neural network to screen wavelet data from videokeratography examinations and determine whether the cases had undergone previous refractive surgery. The trained network correctly identified post-refractive surgery corneas with a 99.3% accuracy, 99.1% sensitivity and 100% specificity [10].

Carvalho et al. in 2007 compared the accuracy of neural networks and discriminant analysis (DA) techniques for the classification of corneal shapes, using Zernike coefficients (ZC) as inputs. The study used ZC data from 80 patient examinations with an input set composed of only the first 15 of 5760 available ZC data points from the videokeratograph due to limitations of computational power. The neural networks achieved better overall mean results with an accuracy of 94%, while the

DA techniques achieved a mean accuracy of 84.8%. The study showed that Zernike polynomials could be used for diagnosis automation, by both neural network and DA techniques and they also predicted better results as computational abilities and implementation costs reduced over time [11].

With newer tomography platforms using technologies such as Scheimpflug imaging and slit scanning, it became possible to measure and analyze more than just the anterior surface of the cornea [12, 13]. According to prediction theory, when more variables describing an event can be measured, the model can predict the outcome more precisely. Souza et al. used this theorem to hypothesize and eventually prove that they could use supervised learning methods to combine all the attributes from the Orbscan II (Bausch and Lomb) and improve its accuracy in detecting KCN [14]. While all the previous studies used only multilayer perceptron (MLP) neural networks, Souza et al. also incorporated support vector machine (SVM) and radial basis function neural networks (RBFNN) methods and a receiver operating characteristic (ROC) analysis. The RBFNN is characterized by a layer of input nodes, a layer of output nodes and one intermediate or hidden layer and is the main practical alternative to the MLP for non-linear modelling [15]. Each processing unit in the hidden layer implements a radial basis function, and for their study, Souza et al. chose the Gaussian function, as it is preferred in pattern classification applications. The SVM is a supervised learning method useful for classifying data that is not linearly separable. In SVM, when a two-dimensional separation is not possible, the algorithm searches for a plane in a higher dimension that is able to separate the group with the highest margin. In both, RBFNN and SVM, the hidden layer performs a non-linear transformation from the input space into a high dimensional space. Maximal separation of the data within the high dimensional space is applied in both methods, using a kernel function to find a hyperplane in SVM, and performing a subsequent linear transformation in RBFNN. The study used a standard MLP with a single hidden layer, and weights and biases were measured and averaged prior to the training data run in order to prevent any biases, and a scaled conjugate gradient was used along with a "cross-entropy error function" in the training algorithm.

To understand cross-entropy error function, we must first understand the Information Theory given by Claude Shannon which says that the occurrence of an unlikely event gives you more information than the occurrence of a likely event, so in ML, information (measured in bits) quantifies the uncertainty in one single event, but what if you are interested in a sequence of events, and not just one event? That is where entropy enters, as it gives us the amount of information required to transmit a randomly selected event from a probability distribution. A skewed distribution has a low entropy, whereas a distribution where events have equal probability has a larger entropy. Cross entropy is a measure of the difference between two probability distributions over the same set of events. This method prevents the problem of over fitting, where the classifier accurately models the training data, but performs poorly on test data. RBFNN offers advantages over MLP as it is more resilient to a poor training set, and the simple linear transformation in the output layer can be better optimized than methods seen in MLP techniques. SVM also offers advantages over

MLP, such as improved accuracy with a smaller training data set, and fewer errors such as local minima as the solution to SVM is unique. Compared with MLP, a disadvantage with RBFNN and SVM is that they give equal weight to every attribute, and consequently, they cannot deal effectively with irrelevant attributes. Despite all the unique advantages of RBFNN and SVM over MLP, the study results showed no statistical difference between the three different methods, although each of the three methods was significantly more accurate than any one single attribute provided by the Orbscan.

After establishing the validity of neural networks as a tool for automated diagnosis of KCN, researchers looked into methods to widen and refine its application. Accardo et al. published their work on neural network in KCN diagnosis wherein they compared six different neural networks with several combinations of the number of inputs, hidden layers, output nodes and learning rates [16]. They also compared whether KCN screening should be performed on both eyes of the same subject, or evaluate each eye separately. They found that screening on both eyes provided more accurate results, and utilizing nine input parameters and three outputs, with fewer hidden neurons improved the discriminant ability of the neural network, and produced a network that was easy to implement and quick in both training and test phases. Kovacs et al. used ML to enable even earlier diagnosis of KCN, by developing an algorithm to identify and compare characteristics of the subtle morphologic changes in the clinically normal fellow eyes of patients with keratoconus [17]. The study analyzed tomographic, topographic and keratoconus indices from the Pentacam HR (Oculus, Wetzlar, Germany) in patients with bilateral KCN, and normal eyes of patients with unilateral KCN, and a group of normal control eyes. An MLP classifier trained on bilateral index of height decentration values had the highest accuracy in discriminating fellow eyes of unilateral KCN patients, from control eyes (area under the ROC 0.96). The ability to accurately detect KCN during the preclinical stages is the goal in KCN and refractive surgery screening.

To further improve the understanding of AI applications in ectasia diagnosis, Lopes et al. compared five different machine learning techniques to analyze tomographic data and detect ectasia [18]. The models they compared were regularized discriminant analysis (RDA), SVM, naïve Bayes (NB), neural networks and random forest (RF). At the time of its publication, it was one of the largest most diverse studies on AI in KCN diagnosis, incorporating data from 3693 eyes and five centres across three different continents. The training data set was divided into three groups, stable or controls, ectasia cases and preoperative data of patients that developed post LASIK ectasia. To overcome a limitation in the study by Kovacs et al., they composed their test set with cases not included in the training model (external validation) [17]. To avoid over-fitting and assess the external validity, a hold-out validation method was performed along with independent tests. Once the five different AI models were trained, their accuracies were measured and the AUC was compared. The RF had the highest accuracy, 0.992. Compared with the RF model, the AUC was significantly lower in the other four models. The RF model was named as the Pentacam Random Forest Index (PRFI), and it had better diagnostic accuracy than the pre-existing tomographic indices, including the Belin/Ambrósio Display

(BAD-D), which correctly classified only 55.3% of the post LASIK ectasia while the PRFI correctly classified 80%. PRFI was the first model trained with the preoperative exam of patients that later developed ectasia. The PRFI had a sensitivity of 85.2%, and specificity of 96.6%, which is lower than some of the previous studies that trained models to identify very asymmetric ectasia with normal topography. This difference can be explained by the fact that their training and testing sets were exclusive, their sample population was more heterogenous and they had a much larger number of cases. These factors dramatically reduce the risk of over-fitting and provide a more accurate representation of the algorithm's performance in real-world scenarios.

Newer diagnostic platforms have allowed us to move beyond analyzing only the shape of the cornea, and also bring into focus the corneal biomechanical properties. The Corvis ST (Oculus Optikgeräte GmbH, Wetzlar, Germany) is a non-contact tonometer with a collimated air pulse of fixed pressure that uses an ultra-high-speed Scheimpflug camera to monitor corneal deformation. It measures parameters such as the inverse concave radius of curvature during the concave phase of the deformation response, the ratio between deformation amplitude at the apex and at 2 mm from the apex, the stiffness parameter at first applanation and the horizontal thickness profile [14, 18, 19].

Corneal Biomechanical Index Vinciguerra et al. combined these parameters by logistic regression analysis for the development of the Corneal Biomechanical Index, which provides high accuracy to detect KCN. To reduce over-fitting, they included only one eye per case from over 600 cases and used two databases, one for training and one for validation. In the validation dataset, the AUC was 0.999 with 98.4% specificity and 100% sensitivity and correctly classified 98.8% of the cases. The study introduced the Corneal Biomechanical Index for KCN diagnosis and demonstrated it to be highly sensitive and specific to separate healthy from ectatic eyes [20].

Ambrosia et al. developed an index combining Scheimpflug-based corneal tomography from the Pentacam and biomechanical assessments from the Corvis for improving ectasia detection [19]. They enrolled 850 eyes from 778 patients from two clinics in Brazil and Italy and divided the eyes into three groups, KCN, very asymmetric ectasia with clinical ectasia (VAE-E), and very asymmetric ectasia with normal topography (VAE-NT). The data from corneal deformation response and corneal tomography, including the Corneal Biomechanical Index and Belin/Ambrósio Deviation (BAD-D) was analyzed and combined into indices using three different artificial intelligence methods, including logistic regression analysis with forward stepwise inclusion, support vector machine and random forest. The leave-one-out cross validation (LOOCV) technique was adopted for validation. The LOOCV method increases computational time and complexity, but also significantly increases the reliability or robustness of the model in classifying new data and provides a more conservative and truthful representation of the generalized performance for the indices in a novel population. The most accurate method was the random forest, which is referred to as the Tomographic and Biomechanical Index

(TBI), and it provided 100% sensitivity for detecting clinical ectasia (KCN and VAE-E groups) with 100% specificity. The TBI has now been commercially deployed by Oculus.

Keratodetect Lavric et al. developed Keratodetect, an algorithm to determine whether an eye is affected by keratoconus or not [21]. Keratodetect used only images of corneal topography as the input parameter, rather than numerical data from topographers. Interestingly, input images were not captured from a topography device, but rather generated using the SyntEyes KTC model. The algorithm was developed using topographies from 1500 healthy eyes and 1500 eyes with KCN. 1350 topographies were used for training, 150 for validation and 200 for testing. The input images were pre-processed to have the same resolution before being applied to the CNN. The algorithm decomposes the image into pixels, and then each image passes through a series of kernel convolutional filters, pooling layers and fully connected layers. The convolutional layer generates an image matrix, by first performing feature extraction from the image, and then using filters to learn the characteristics of the image. The convolutional layer is followed by a pooling layer which has the purpose of removing redundant information from the layers, by increasing the number of filters. The last layer is a fully connected neural layer which combines all the features extracted and learned by the previous layers so as to identify patterns in the input data. This study classified data into only two classes, normal and KCN, and it was able to do so with 99.33% accuracy. The use of topography images as the sole input parameter is unique in this study.

Pentacam InceptionResNet V2 Screening System (PIRSS) Xie et al. also developed a classification system, the PIRSS based on an image learning technique [22]. Their study utilized tomographic images from the Pentacam to obtain the overall profile of the cornea, comprising the axial curvature, front elevation, back elevation and corneal thickness. They employed the InceptionResNetV2 architecture in a convolutional neural network on the TensorFlow platform with transfer learning technique using 6465 tomographic images from 1385 patients. The images were divided into two independent training data sets with 5130 images, and 1335 images were used for validation of the model, and 100 new images collected from different patients were used in the test data set to compare accuracy with human specialists. Images were divided in to five categories: normal corneas, suspected irregular corneas, early-stage KC, KC, and post myopic refractive surgery. The model was compared against human specialists divided according to the level of experience into the following groups: senior ophthalmologists who perform refractive surgery, fellows of refractive surgery, senior ophthalmologists who are not refractive surgeons and medical students who are not studying refractive surgery. PIRSS achieved an overall accuracy of 95.0%, comparable with 92.8% achieved by the senior ophthalmologists performing refractive surgery. InceptionResNetV2 algorithm and TensorFlow were more advanced than the models used in previous studies. The use of tomographic heat maps rather than limited parameters derived from the maps can pose an advantage as it is analyzing a relatively larger amount of data from each cornea. The team is also working on deploying PIRSS as a web service where other ophthalmologists and possibly even patients can upload their scans and receive advice on suitability for refractive surgery.

4.3 Artificial Intelligence in Refractive Surgery

Over a few decades, refractive surgery has evolved from a procedure with limited predictability in the form of radial keratotomy to minimally invasive laser-based procedures that can deliver excellent visual outcomes. Overall refractive surgery has become one of the most successful and commonly performed elective procedures in the world, and today there are multiple different procedures and technologies available to perform refractive surgery. With more and more people undergoing these procedures, it is becoming crucial to ensure proper candidate selection to minimize complications after surgery. Algorithms developed using data sourced from corneal tomography and biomechanics can be very accurate and sufficient for diagnosing ectasia; however, screening candidates for refractive surgery requires a more wholesome approach and consideration of various other factors. We are all well aware of the complicated relationships that exist between a patient's optical parameters and the final results achieved. Preoperative parameters that have been shown to influence refractive outcomes include the patient's age, gender, spherical equivalent, pupil size, corneal tomography, intraocular pressure and tear film characteristics.

Yoo et al. aimed to develop an algorithm that could better mimic a refractive surgeon's decision matrix to provide clinical decision support based on multiple different parameters such as age, gender, spherical equivalent, corrected distance visual acuity, intraocular pressure, central corneal thickness, non-invasive tear break up time and corneal tomography from the Pentacam [23]. The patients selected were classified into two groups: those who had already undergone a refractive surgery, labelled as "candidates for refractive surgery", and those who had been refused surgery based on the presence of a contraindication, labelled as "contraindication for corneal refractive surgery". They recruited >15,000 patients in the study, and 10,561 subjects were used in the training set, 2640 in the internal validation set and 5279 in the external validation set. For algorithm development, they did not rely on just one ML technique, rather they used five different techniques to develop five different algorithms. They used some techniques such as SVM, RF, ANN, which had already been used in ectasia classification studies, and also used AdaBoost and LASSO (least absolute shrinkage and selection operator) which at the time were unique to this application. AdaBoost, short for adaptive boosting, is an ensemble learning method that uses an iterative approach to learn from the mistakes of weak learners to build a strong classifier. Ensemble learning combines several base algorithms to form one optimized algorithm. Boosting algorithms, like humans, learn from their mistakes and try not to repeat them again. It starts with creating a model from the training data, then creating a second model from the previous data set by trying to reduce the errors from the previous model. Sequentially more models are added till the training data is accurately predicted. A weak classifier is one that performs better than simply guessing, but is poor at classifying accurately. AdaBoost is applied on top of any other model to learn from its shortcomings and improve its accuracy. LASSO is a form of penalized regression and can assist investigators interested in predicting an outcome by selecting the subset of the variables that minimizes prediction error. Finally, data from all five ML techniques was combined using a weighted majority vote ensemble classifier to further improve accuracy. On the training data set, the RF decision tree out performed SVM, ANN, AdaBoost and LASSO, and while the ensemble classifier further improved accuracy, the difference was not statistically significant. On the internal and external validation data set, the RF model was the single best performing algorithm, but the difference was not statistically significant. The weighted majority vote ensemble classifier predicted refractive surgery suitability with an accuracy of 93.4% and an AUC of 0.972 in external validation, matching the accuracy of experts with consistent performance in the high-risk subgroups. Yoo et al.'s work showed that ML can help automate and potentially standardize the comprehensive refractive surgery screening process.

In order to investigate the role of AI in helping automate the process of making decisions in refractive surgery candidates, Yoo et al. developed a multiclass ML model that determined which type of laser vision correction procedure would be most [24]. The study used XGBoost, a more recently developed decision tree-based ensemble ML algorithm, that uses a gradient boosting framework. While ANN work best for unstructured data such as images or text, decision tree-based algorithms are considered best for structured data. XGBoost includes many enhancements such as regularization to prevent overfitting, sparsity awareness and built-in cross validation. The model was trained to predict surgery class options between LASEK, LASIK, SMILE and contraindication to LVC. The study also used the Shapley Additive exPlanations (SHAP) explainable model that allows users to verify if the model operates properly by observing the rationale behind decisions. Most ML models are like black boxes that are unable to provide reasoning and explanations behind decisions. While earlier algorithms provide predictors with global feature importance, the SHAP technique determines the contribution of each input variable in each decision of a model. The study also adopted the one-versus-rest (OVR) and one-versus-one (OVO) strategies for explainable classification as they better facilitate representation through an intuitive diagram. During data preprocessing, synthetic minority oversampling technique (SMOTE) was used to overcome any imbalances in the data. Out of the total 18,480 subjects included in the study, 10,561 and 2640 subjects were assigned to the testing and internal validation datasets, respectively. The external validation set consisted of 5279 subjects who were considered as independent prospective cohorts to validate the ML model prospectively. The final model achieved an accuracy of 78.9% in predicting the correct surgery option in external validation. The highlight of this study was the implementation of explainable ML, allowing us to understand that although the XGBoost model managed to match the clinician's decision in 92.7% of instances, it relied on the anticipated choice of the patient, rather than other clinical data as the most influential factor in classification. Including more investigative modalities in these algorithms can further help improve their efficiency. However, implementation of such algorithms has many challenges due to the availability of numerous different laser platforms and customized treatments, variation in diagnostic platforms, surgeon-related factors and patient demographics.

Table 4.1 summarizes the landmark studies on the use of AI in cornea and refractive surgery.

4.4 Artificial Intelligence in Dry Eye Disease

Dry eye syndrome is one of the most common complaints in general ophthalmology clinics. Meibography to help diagnose and asses severity of evaporative dry eye is also gaining popularity, particularly with the advent of treatments such as the vectoring LipiFlow thermal pulsation system (Johnson & Johnson Vision, Jacksonville, FL, USA) and intense pulsed light therapy. Work has been done to help automate and standardize the grading of meibography images with the help of ML algorithms. In 2012, Koh et al. employed a SVM to classify meibography images as healthy and unhealthy [25]. They used a small data set of 55 images with experts classifying 26 images as healthy and 29 and unhealthy, from these 13 healthy and 15 unhealthy were used in the training set, and the remaining in the testing set. The images were first processed to obtain the average arc length, the average width of the glands and the uniformity of the glands, captured as entropy. These three features were used to train the SVM, and it achieved a specificity of 96% and sensitivity of 98% on the test data. The study had some limitations, such as only using images of the upper eyelids and analyzing only three features. At the time of their work, good investigative platforms for meibography were not available and standardizing the process of capturing images was a challenge. The entire process of measuring the three gland parameters from the images was done using algorithms, but it was still a ten-step process, making it very laborious and time consuming. Remeseiro et al. developed an SVM to classify the tear film lipid layer using the features extracted by different texture analysis methods [26]. There is a lot of potential in the development of algorithms to differentiate healthy and unhealthy glands, and grade severity of damage to glands and finally develop a decision matrix for treatment on the basis of severity.

4.5 The Road Ahead

Over the past 25 years, there has been significant progress made in the research and clinical applications of machine learning for diagnosis and treatment of corneal diseases. These advances have been made possible due to synchronous developments in the fields of medicine, computers and data science. Algorithms are built and improved with data, and faster computers and better machine learning techniques have enabled the consumption of more and more data to help train more accurate AI models. To adequately train models and produce results that can be generalized, training data sets need to be in the order of tens of thousands of cases.

Table 4.1 Summary of landmark studies on the use of Artificial Intelligence in Cornea and Refractive surgery

Authors	Examinations	Machine learning techniques used	Inputs	Outputs	Training set size	Validation/test set size	Results
Maeda et al. [8]	TMS-1 videokeratoscope	Neural network	11 topographic indices	Normal, WTR, KCN mild, KCN moderate, KCN advanced, post-PRK, post-PKP	108	75	Total accuracy 80%, 44–100% sensitivity in subgroups >90% specificity in subgroups
Smolek and Klyce [9]	TMS-1 videokeratoscope	Neural network plus binary decision tree	10 topographic indices	Other, KCS KC1 KC2 KC3	150	150	Total accuracy 100%, KC accuracy 100%
Smolek and Klyce [10]	TMS-1 videokeratoscope	Neural network	Wavelet data	Normal, post-refractive surgery	138	138	Total accuracy 99.3%
Accardo and Pensiero [16]	EyeSys 2000	Neural network	9 unilateral topographic indices, 10 bilateral indices	Normal, KCN, other	95	103	Sensitivity 94.1%, specificity 97.6%
Carvalho and Barbosa [11]	EyeSys 2000	Neural network plus discriminant analysis	Zernike coefficients	Normal, KCN, WTR, ATR, post-PRK	40	40	Total accuracy: 94% by NN, 85% by DA
Souza et al. [14]	Orbscan	Radial basis function, neural network, support vector machine, multilayer perceptron	11 topographic indices	KCN, astigmatism, post-PRK	318	Cross validation	AUC of KC: 0.99 by RBFNN, 0.99 by SVM, 0.98 by MLP

(continued)

Table 4.1 (continued)

Authors	Examinations	Machine learning techniques used	Inputs	Outputs	Training set size	Validation/ test set size	Results
Kovacs et al. [17]	Pentacam HR	Neural network	Unilateral and bilateral indices	Normal, FFKC, KCN	135	Cross validation	Sensitivity of KC and FFKC: KC vs normal cornea, 100%; FFKC vs normal cornea, 90%
Lopes et al. [18]	Pentacam HR	Random forest, neural network, Bayes network, support vector machine, discriminant analysis	Keratometric values, topometric and topographic indices	Stable LASIK, post LASIK ectasia, KCN, FFKC, normal	3233	486	Differentiating normal vs. normal eye of ectasia, very asymmetric ectasia—Normal topography with 0.97AUC, 85.2% sensitivity and 96.6% specificity
Vinciguerra et al. [20]	Corvis ST	Logistic regression	16 dynamic corneal response parameters	Healthy, KCN	329	329	AUC of 0.999, 98.4% specificity, 100% sensitivity
Ambrosio et al. [19]	Corvis ST, Pentacam HR	Logistic regression, support vector machine, random forest	16 parameters from the Pentacam and Corvis ST	Normal, KCN, very asymmetric ectasia with clinical ectasia, very asymmetric ectasia with normal topography	850	Leave one out cross validation	90.4% sensitivity, 96% specificity in very asymmetric ectasia with normal topography
Lavric and Valentin [21]	SyntEyes	Convolutional neural network	Pixel values of an image representing corneal tomography	Normal, KCN	1350	150	On the test data set, the algorithm had an accuracy of 99.33%

Xie et al. [22]	Pentacam HR	Convolutional neural network	Images of the corneal axial curvature, the front elevation, the back elevation and the corneal thickness	Normal, suspect, early KC, KC, post-operative	5130	1335	Total accuracy 95%, KC accuracy 98%
Yoo et al. [23]	Pentacam, clinical examination, demographics	Support vector machine, artificial neural networks, random forest, least absolute shrinkage and selection operator, AdaBoost, ensemble classifier	Age, sex, anticipated surgery option, occupation, corneal anterior and posterior k values, CCT, IOP, spherical equivalent, NIBUT, pupil size	Candidate for corneal refractory surgery, contraindication for corneal refractive surgery	13,201	5279	The ensemble classifier had AUC of 0.972 in the external validation set
Yoo et al. [24]	Pentacam, keratograph 5 M, clinical examination, demographics	Explainable machine learning, XGBoost, shapley additive explanations, one vs rest, one vs one	Age, sex, spherical equivalent, CDVA, IOP, pupil diameter, central corneal thickness, NIBUT, anticipated surgery option, occupation, anticipated recovery time, concern about budget	LASEK, LASIK, SMILE, contraindication	13,201	5279	XGBoost model exhibited an accuracy of 78.9% in external validation, and explainable machine learning showed that the model gave greatest importance to the patient's anticipated choice of procedure rather than any clinical parameter

(continued)

Table 4.1 (continued)

Authors	Examinations	Machine learning techniques used	Inputs	Outputs	Training set size	Validation/test set size	Results
Koh et al. [25]	Infrared video camera mounted on a slit lamp	Support vector machine	Meibography images were processed to provide gland length, average thickness and uniformity	Healthy, unhealthy	28	27	Model achieved specificity of 96% and sensitivity of 98% on the test data

WTR with the rule astigmatism, *KCN* keratoconus, *PRK* photorefractive keratectomy, *PKP* penetrating keratoplasty, *KCS* keratoconus suspect, *ATR* against the rule astigmatism, *NN* neural networks, *DA* discriminant analysis, *AUC* area under the curve, *RBFNN* radial basis function neural network, *SVM* support vector machine, *MLP* multilayer perceptron, *FFKC* forme fruste keratoconus, *CCT* central corneal thickness, *IOP* intraocular pressure, *CDVA* corrected distance visual acuity, *NIBUT* non-invasive tear film break up time, *LASEK* laser epithelial keratomileusis, *LASIK* laser-assisted in situ keratomileusis, *SMILE* small incision lenticule extraction

Large data sets are also required to effectively deal with the high noise levels derived in biological data. Neural networks developed to classify retinal fundoscopic images have been trained on datasets of more than 100,000 images [27]. In comparison, ML models developed for corneal studies have seen the size of training data sets increasing from only around 100 eyes in the first few studies to over 10,500 cases in the more recent studies. These numbers are still very small compared to those seen in retinal applications of AI because it is far more challenging to build very large datasets using corneal imaging. The high cost of tomographers means they are not as commonplace in ophthalmic practices as retinal fundus imaging devices. The challenge faced in gaining access to a tomographer is highlighted by the fact that one of the studies discussed earlier in the chapter had to resort to using data synthesized by another AI-based model [21]. Another limitation is the variation across different devices, even those that use the same technology. For example, with Scheimpflug imaging, the scans obtained from the same patient with different devices and even different software versions are not easily interchangeable, while retinal images acquired from different devices can be processed to be used together with relative ease. This restricts data sets to a single device type, and with various clinicians preferring different devices, collecting large datasets becomes even more difficult. In the absence of a standardized dataset, it is not possible to directly compare the various models developed by different researchers. In other fields, advances in machine learning have been accelerated by the creation of large public datasets, and the release of similar corneal imaging datasets under strict patient privacy rules can help ophthalmic and data science researchers collaborate and further advance ML applications for corneal disorders. Finally, these models need to be available commercially to clinicians to be able to actually derive any benefit from them. They can be made available by either implementing them directly in the user interface of the various diagnostic machines on which they are validated or as a web-based tool that can be accessed easily anywhere. Even though seamless integration into the diagnostic machines themselves, as with the CBI and TBI, is the most convenient for clinicians, the responsibility for creating ML models must not be left solely to companies producing various diagnostic machines, as they will limit the benefits of the models to their own platforms, limiting its applications. It is important for researchers to build models that are validated across various platforms so that they benefit a larger cohort of patients and clinicians.

In ophthalmology, another problem faced in the development of predictive models is in the selection of training data. Some of the initial studies trained models using individual eyes from patients rather than considering them as cases and using both eyes together. Subsequently as studies began to use data from both eyes together to screen for KCN, the accuracy of models also began to improve. Studies by Accardo et al. and Kovacs et al. demonstrated the advantages of analyzing patients bilaterally [16, 17]. Early models were plagued with biases such as overfitting and thus overestimated their accuracy. As studies adopted better data preprocessing and validation techniques, the accuracy reported by newer models was lower than those of earlier studies. The employment of external validation by more recent studies has helped ensure that the algorithm works not only on the training data set but also on data the model has not previously seen, providing a better estimate of real-world performance of the model. Data preprocessing is a crucial step in achieving accurate predictions from a model and has improved, becoming far more sophisticated over time in the newer studies with a better understanding of biases such as over-fitting and techniques to improve them.

Artificial neural networks, DL and other ML techniques have demonstrated their potential to become useful tools in helping clinicians deliver the best quality of care to patients in a more efficient manner. It has the amazing ability to reveal characteristics and relationships between parameters that are initially imperceptible to the human brain. However, it is not easy to accept and act on the advice of a computer system without knowing the system's reasoning for the decision. Explainable ML can help overcome this hurdle as it can expose the shortcomings of a particular model. Explainable ML can help us avoid such pitfalls in the application of AI and make sure the algorithms are giving preferential weightage to the appropriate clinical factors. This could help improve acceptance and adoption of more AI models and increase our confidence in them. AI has tremendous potential to help improve diagnosis and refine outcomes from current corneal procedures and develop ML models that can match the diagnostic accuracy of experienced clinicians which in turn can greatly benefit younger doctors by reducing mistakes made due to relative inexperience. This democratization of clinical skills eventually helps improve the standard of healthcare for patients.

References

1. Daxecker F. Christoph Scheiner's eye studies. In: Henkes HE, editor. History of ophthalmology 5: sub auspiciis academiae ophthalmologicae internationalis. Dordrecht: Springer; 1993. p. 27–35.
2. Alarcón A, Rubiño M, Péééréz-Ocón F, Jiménez JR. Theoretical analysis of the effect of pupil size, initial myopic level, and optical zone on quality of vision after corneal refractive surgery. J Refract Surg. 2012;28(12):901–6.
3. Randleman JB, Trattler WB, Stulting RD. Validation of the ectasia risk score system for preoperative laser in situ keratomileusis screening. Am J Ophthalmol. 2008;145(5):813–8.
4. Ambrósio R Jr, Lopes BT, Faria-Correia F, Salomão MQ, Bühren J, Roberts CJ, Elsheikh A, Vinciguerra R, Vinciguerra P. Integration of scheimpflug-based corneal tomography and biomechanical assessments for enhancing ectasia detection. J Refract Surg. 2017;33(7):434–43.
5. Jonas JB, Nangia V, Matin A, Kulkarni M, Bhojwani K. Prevalence and associations of keratoconus in rural Maharashtra in central India: the central India eye and medical study. Am J Ophthalmol. 2009;148(5):760–5.
6. Kennedy RH, Bourne WM, Dyer JA. A 48-year clinical and epidemiologic study of keratoconus. Am J Ophthalmol. 1986;101:267–73.
7. Seiler T, Quurke AW. Iatrogenic keratectasia after LASIK in a case of forme fruste keratoconus. J Cataract Refract Surg. 1998;24(7):1007–9.
8. Maeda N, Klyce SD, Smolek MK. Neural network classification of corneal topography. Preliminary demonstration. Invest Ophthalmol Vis Sci. 1995;36(7):1327–35.
9. Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. Invest Ophthalmol Vis Sci. 1997;38:2290–9.
10. Smolek MK, Klyce SD. Screening of prior refractive surgery by a wavelet-based neural network. J Cataract Refract Surg. 2001;27:1926–31.

11. Vieira de Carvalho LA, Barbosa MS. Neural networks and statistical analysis for classification of corneal videokeratography maps based on Zernike coefficients: a quantitative comparison. Arq Bras Oftalmol. 2008;71:337–41.
12. Belin MW, Khachikian SS. Keratoconus: it is hard to define, but. Am J Ophthalmol. 2007;143:500–3.
13. Cairns G, McGhee CN. Orbscan computerized topography: attributes, applications, and limitations. J Cataract Refract Surg. 2005;31:205–20.
14. Souza MB, Medeiros FW, Souza DB, Garcia R, Alves MR. Evaluation of machine learning classifiers in keratoconus detection from orbscan II examinations. Clinics. 2010;65:1223–8.
15. Nabney YT. Netlab: algorithms for pattern recognition. 4th ed. London: Springer; 2004.
16. Accardo PA, Pensiero S. Neural network-based system for early keratoconus detection from corneal topography. J Biomed Inform. 2002;35:151–9.
17. Kovács I, Mihåltz K, Kránitz K, et al. Accuracy of machine learning classifiers using bilateral data from a Scheimpflug camera for identifying eyes with preclinical signs of keratoconus. J Cataract Refract Surg. 2016;42(2):275–83.
18. Lopes BT, Ramos IC, Salomão MQ, et al. Enhanced tomographic assessment to detect corneal ectasia based on artificial intelligence. Am J Ophthalmol. 2018;195:223–32.
19. Ambrosio R Jr, Lopes BT, Faria-Correia F, et al. Integration of Scheimpflug-based corneal tomography and biomechanical assessments for enhancing ectasia detection. J Refract Surg. 2017;33(7):434–43.
20. Vinciguerra R, Ambrósio R Jr, Elsheikh A, et al. Detection of keratoconus with a new biomechanical index. J Refract Surg. 2016;32:803–10.
21. Lavric A, Valentin P. KeratoDetect: Keratoconus detection algorithm using convolutional neural networks. Comput Intell Neurosci. 2019;2019:8162567.
22. Xie Y, Zhao L, Yang X, Wu X, Yang Y, Huang X, Liu F, Xu J, Lin L, Lin H, Feng Q, Lin H, Liu Q. Screening candidates for refractive surgery with corneal tomographic-based deep learning. JAMA Ophthalmol. 2020;138(5):519–26.
23. Yoo TK, Ryu IH, Lee G, et al. Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. NPJ Digit Med. 2019;2:59.
24. Yoo TK, Ryu IH, Choi H, Kim JK, Lee IS, Kim JS, Lee G, Rim TH. Explainable machine learning approach as a tool to understand factors used to select the refractive surgery technique on the expert level. Trans Vis Sci Tech. 2020;9(2):8.
25. Koh YW, Celik T, Lee HK, Petznick A, Tong L. Detection of meibomian glands and classification of meibography images. J Biomed Opt. 2012;17(8):086008.
26. Remeseiro B, Penas M, Mosquera A, Novo J, Penedo MG, Yebra-Pimentel E. Statistical comparison of classifiers applied to the interferential tear film lipid layer automatic classification. Comput Math Methods Med. 2012;2012:207315.
27. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10.

5 Artificial Intelligence and Cataract

Sahil Thakur, Jocelyn Hui Lin Goh, and Yih-Chung Tham

5.1 The Magnitude of the Problem

Accounting for 65.2 million cases of vision impairment and blindness globally, cataract is the leading cause of avoidable blindness in the world [1]. These cases are projected to increase to 70.5 million by 2020 [1]. Despite considerable advancement in safe management of cataract, diagnosis and case finding remains a substantial problem, especially in countries with weak public health infrastructure [2–4]. It has been estimated that eliminating cataract-related vision impairment in a country like India would cost $2.6 billion and would yield a net societal benefit of $13.5 billion [5]. This indicates that there is a significant potential for investment and innovation in developing tools that can address this public health problem.

5.2 Limitations of Current Clinical Practice

We can classify these into broadly four categories.

S. Thakur · J. H. L. Goh
Ocular Epidemiology Research Group, Singapore Eye Research Institute, Singapore
e-mail: jocelyn.goh.h.l@seri.com.sg

Y.-C. Tham (✉)
Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore

Duke-NUS Medical School, Singapore, Singapore
e-mail: tham.yih.chung@seri.com.sg

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_5

5.2.1 Case Detection/Screening Programs

Currently cataracts are clinically graded by ophthalmologists on slit-lamp examination. The Lens Opacities Classification System (LOCS) III criteria is one such standard image-based criteria that is widely used [6]. In addition to clinical classification, this system has also been used to plan the type of intervention needed. This grading process requires clinical expertise, training and expensive ancillary equipment for making clinical decisions about patient management. This poses a significant challenge for developing countries or rural communities, where there is shortage of these resources [7]. Additionally, image-based grading scales are subjective and may be affected significantly by inter and intra grader variability [8]. These challenges thus make screening for cataract a time-dependent and financially expensive endeavour. Thus, there is an unmet need to develop new methods that can address these limitations and potentially aid in efficient and effective cataract screening.

5.2.2 Intraocular Lens (IOL) Power Calculation

The standard of care for cataract management is surgical removal of the human lens followed by IOL implantation for restoration of vision [9]. Over the past 70 years, numerous advancements have studded the history of cataract management, but the single most important component of cataract surgery planning is the calculation of IOL power [9]. This significantly affects the visual outcome and prognosis of the surgery. IOL power calculation depends on several factors like axial length, corneal curvature, effective lens position and type of IOL selected [10–14]. These days these calculations can be made quickly using biometry machines that are vital parts of a cataract surgeon's workflow. Nevertheless, due to the wide variation of ocular biometry profiles across individuals, there is currently no single formula that can be used in all patients [15]. Existing formulae were created for eyes with a typical range of normative biometric measurements and often cannot be used for extremes of axial length [16]. They also cannot be applied in eyes with atypical corneal profiles like eyes with history of refractive surgery, keratoplasty, keratoconus, microcornea and significant astigmatism [17–19]. By using AI, ML and large-scale registries, novel IOL formulae have been developed that can effectively be used for a larger subset of patients [20]. Though we are still away from the one formula fits all scenarios, in the coming years we can expect to see major developments in this regard.

5.2.3 Manpower Training and Surgical Review

In order to meet the expected need of the ageing population, the current graduation rate of ophthalmologists would have to increase by 75%–100% and even this increase may not be able to meet all outpatient requirements over the next 20 years [21]. It has also been noted in an Australian study that the supply of ophthalmologists to remote regions is usually 19 times lower than national average. This also translates to reduced cataract surgery rates, up to ten times lower than national average [22]. One of the most important limitations in current training programs for ophthalmic surgeons is the need for extensive patient exposure. Often this exposure may not be available, and the graduated surgeon may have to invest in a fellowship program to further develop their skills. This translates to significant investment in terms of training costs, time and skill development [23]. It is also important to know that for most training centres techniques such as virtual reality (VR) simulators, wet labs, didactic training and watching videos are the primary tools to teach advanced surgical techniques [24]. This is where applications of AI, like force sensing in VR surgical simulators and automatic annotation of surgical videos, can help in streamlining the training and accreditation process [25–27]. The ICO-OSCAR guidelines are currently used to assess the competency of the surgeons but need close mentor supervision in the patient setting. In the future, software-assisted assessment may become the standard in wet-lab setting allowing the trainees to gain more confidence and proficiency before actual patient exposure [28, 29]. DL-assisted speaker systems that confirm surgical sites in high-volume settings are another novel use case scenario for optimized patient outcomes [30]. Such applications can prevent wrong-site surgery and potential high litigation costs [31].

5.2.4 Post-Operative Care and Quality of Life (QoL)

Good post-operative care is one of the most important determinants of any surgical procedure. With increased use of digital technology in patient identification, consultation and medication management, role of artificial intelligence will increase over the coming years. Live teleconsultations, chatbots and simulated tele attendants are being explored for teleconsultation and feedback collection indicating trends that digital health monitoring will take over the next few years [32, 33]. These technologies can be subsequently adopted into mobile applications that can guide patients during their post-operative period. It is also possible to predict the risk of complications like posterior capsular opacification post-cataract surgery using artificial neural networks [34]. These applications indicate the potential of artificial intelligence in significantly improving the quality of life and surgical outcomes in patients with cataract.

5.3 Artificial Intelligence and Cataract Detection

There are several publications that report algorithms for automated detection and grading of cataract. These algorithms differ in terms of approach, input type and potential use case scenario. We have briefly described below some of the most popular techniques used for detection and classification of cataracts.

5.3.1 Based on Slit-Lamp Photographs

These studies solely focus on using slit-lamp photographs as training data for algorithm development for automated detection and grading of nuclear cataract. Li and colleagues applied a modified Active Shape Model (ASM) to first identify the lens and its nucleus on 5820 slit-lamp photographs from the Singapore Malay Eye Study (SiMES) [35]. The algorithm had 95% success rate in correctly identifying the location of lens. Subsequently, 100 slit-lamp photographs were used to develop a severity grading algorithm using the support vector machine (SVM) regression model. When compared to the reference standard (Wisconsin cataract grading system), the algorithm showed a mean difference of 0.36 for nuclear cataract grading in a set of 5490 photographs.

Xu et al. [36] reported a mean absolute error (MAE) of 0.336 in their study. They used a modified ASM for identification of lens location, but then applied bag-of-features (BOF) model for feature extraction, and group sparsity regression (GSR) for feature selection and nuclear cataract severity grading. Using the same data set as Xu et al., Gao et al. [37] demonstrated an MAE of 0.304 by using a convolutional-recursive neural network (CRNN) for lens detection and feature learning. The incremental improvement in performance that we note in these studies is due to use of more sophisticated machine learning (GSR) and deep learning (CRNN) techniques which have greater ability in terms of feature extraction and learning [38].

Wu et al. have demonstrated use of residual neural network (ResNet) algorithm to comprehensively diagnose and refer cataract patients. Their algorithm can differentiate between different input images like dilated/undilated/optical section/diffuse slit-lamp images. Then the algorithm classifies the images as normal, cataractous or post-cataract surgery. If the algorithm identifies an image with cataract, the type and severity of the cataract is identified based on the LOCS II scale. Finally, the algorithm arrives at a decision whether to follow-up or refer the patient for tertiary care. The results from their study are extremely encouraging and consistently show area under the receiver operating characteristic curve (AUC) of more than 0.90 for the different stages of evaluation. (Table 5.1) It is interesting to note that the dilated eye images with optical sections were the best for evaluating cataract status (AUC 0.9915) while undilated images with diffuse illumination (AUC 0.9328) were least optimal [39]. This AI algorithm was further pilot tested as a web-based platform while incorporating a smartphone app for subject engagement. The subjects can 'self-report' symptoms of decreased vision or blurred vision using the app. These cases then went to community-based healthcare facilities, where undilated slit-lamp images were taken by nurses or technicians. The images were then evaluated by the AI algorithm. The sensitivity and specificity of the algorithm for cataract detection were 92% and 83.85% when compared to ophthalmologist opinion. These findings are a proof of concept that patients with cataract can be identified at community-based healthcare facilities and primary care centres where AI algorithms can be deployed. These innovative digital care models can potentially allow the ophthalmologists to serve more patients than current healthcare models.

5.3.2 Based on Colour Fundus Photographs

Over the last few years, the use of fundus photographs for diabetic retinopathy screening has become a part of several primary healthcare systems [41, 42]. This offers an opportunity to potentially use the fundus images to screen for other causes of vision impairment like cataract as well (Table 5.2).

Table 5.1 Previous studies on automated detection and grading of cataract based on slit-lamp photographs

Year	Author	Data source	Method	Definition of gold standard/ ground truth	Performance
2009	Li et al. [40]	Singapore Malay eye study (SiMES) 100 training, 5490 testing	Modified ASM: Lens structure detection HSV model: Feature extraction SVM: Automatic grading	Wisconsin cataract grading system	Accuracy: 95% MAE (grading): 0.36
2013	Xu et al. [36]	ACHIKO-NC (subset of SiMES) 100 training, 5278 testing	Modified ASM: Lens structure detection BOF model: Feature extraction GSR: Automatic grading	Wisconsin cataract grading system	MAE: 0.336
2015	Gao. et al. [37]	ACHIKO-NC (subset of SiMES) 100 training, 5278 testing	CRNN: Feature learning SVM regression: Automatic grading	Wisconsin cataract grading system	MAE: 0.304

(continued)

Table 5.1 (continued)

Year	Author	Data source	Method	Definition of gold standard/ground truth	Performance
2019	Wu X. et al. [39].	Chinese medical Alliance for artificial intelligence (CMAAI) 30,132 training, 7506 testing	ResNet	LOCS II	1. Capture mode recognition AUC = 99.36%[a] AUC = 99.28%[b] AUC = 99.68%[c] AUC = 99.71%[d] 2. Cataract diagnosis Cataract AUC = 99.93%[a] AUC = 99.96%[b] AUC = 99.19%[c] AUC = 99.38%[d] Post-operative eye AUC = 99.93% [a] AUC = 99.93% [b] AUC = 98.99%[c] AUC = 99.74%[d] 3. Detection of referable cataracts a. Adult cataract: AUC = 94.88% b. Paediatric cataract with VA involvement: AUC = 100% c. PCO with VA involvement: AUC = 91.90%

ASM active shape model, *HSV* hue, saturation, value, *SVM* support vector machine, *BOF* bags-of-features, *GSR* group sparsity regression, *CRNN* convolutional residual neural network, *ResNet* residual neural network, *MAE* mean absolute error, *PCO* posterior capsular opacification, *VA* visual axis
[a]Dilated-diffuse
[b]Dilated-Slit-lamp
[c]Undilated diffuse
[d]Undilated-Slit-lamp

Dong et al. used 5495 fundus images and a deep learning network (Caffe software based) for feature extraction followed by training for cataract detection and grading using machine learning (SoftMax function). The gold standard was appearance of fundus images as graded by ophthalmologists into normal, mild, moderate and severely affected by the cataract. The algorithm was validated in a set of 2355

Table 5.2 Previous studies on automated detection and grading of cataract based on fundus photographs

Year	Author	Data source	Method	Definition of gold standard/ground truth	Performance
2017	Dong et al. [43]	5495 training, 2355 testing	Caffe: Feature extraction SoftMax: Detection and grading	Labelled fundus images by ophthalmologists	Accuracy[a] = 94.07%[b] Accuracy[a] = 90.82%[c]
2017	Zhang et al. [47]	Beijing Tongren eye Centre's clinical database. 4004 training, 1606 testing	DCNN: Detection and grading	Labelled fundus images by graders	AUC = 0.935[b] AUC = 0.867[c]
2018	Ran et al. [44]	Not described	DCNN: Feature extraction Random forest: Detection and grading	Labelled fundus images by ophthalmologists crosschecked by graders	AUC = 0.970[b] Sensitivity = 97.26%[b] Specificity = 96.92%[b]
2018	Li et al. [48]	Beijing Tongren eye Centre's clinical database. 7030 training, 1000 testing	ResNet 18: Detection ResNet 50: Grading	Labelled fundus images by graders	AUC = 0.972[b] AUC = 0.877[c]
2019	Pratap and Kokil [45]	Multiple online databases 400 training, 400 testing	Pre-trained CNN: Feature extraction SVM: Detection and grading	Labelled fundus images by ophthalmologists	Accuracy[a] = 100%[b] Accuracy[a] = 92.91%[c]

DCNN deep convolutional neural network, *CNN* convolutional neural network, *SVM* support vector machine, *ResNet* residual neural network

[a]Denotes proportion of images being correctly classified among total of images tested

[b]2 class (non-cataract versus cataract)

[c]4 class (non-cataract, mild, moderate and severe)

images, with 94.07% images correctly identifying cataract and 90.82% images correctly classifying different severity levels [43].

Ran et al. alternatively used a deep convoluted neural network (DCNN) for feature extraction, and a random forest (RF) machine learning model for identification of cataract. They used a gold standard based on 'haziness' of fundus image as determined by ophthalmologists. They reported excellent performance as well with an AUC of 0.970 and a sensitivity/specificity of 97.26%/96.92%, respectively [44].

Pratap and Kokil demonstrated the use of transfer learning by using a combination of pre-trained CNN and SVM in their AI system. They fine-tuned the pre-trained CNN (trained used millions of non-medical images), using 400 fundus images for relevant feature extraction [38]. Using a similar haziness-based ground truth like the previous studies, they managed to improve the detection performance to 100% and severity classification to 92.91% when tested in another 400 images [45]. Though the algorithm performance is excellent, it is always important to keep in mind the small sample of images and lack of external validation which is vital for real-world deployment of any such algorithm.

Another approach described by Zhang et al. uses a DCNN-based deep learning system trained with 4004 fundus images. A green channel (G-channel) filter was used to enhance the contrast of image and visibility of retinal vessels [46]. Using a similar haziness-based ground truth like the previous studies, the algorithm demonstrated an AUC of 93.52% for cataract detection, and 86.69% for severity grading in a test set of 1606 images [47].

Li et al. subsequently used both ResNet-18 and ResNet-50 for detection and classification of cataract grading [48]. The gold standard in their study was four classes of cataract severity (non-cataract, mild, moderate and severe cataract) based on grading by professional graders. They improved the performance of the algorithm (1000 images test set) to an AUC of 97.2% for cataract detection and 87.7% for severity grading. The interesting feature in their study was the generation of saliency maps (i.e., heatmaps) in order to highlight regions used by the algorithm for decision-making. This is an important step in unlocking the black box of artificial intelligence and increasing end-user understanding of algorithm referral decisions. However, heatmaps shown in the publication are not entirely correlating with degree of cataract severity indicating the potential of improvement in cataract feature detection and classification.

When we look at these studies, we can easily highlight the need for a more objective method of cataract detection and severity classification. The studies employed graders to classify the fundus images on basis of 'clarity' or haziness and assumed it was due to cataract. However, it is imperative to note that factors like ocular surface status, refractive power, anterior chamber status and vitreous humour status can significantly affect the fundus images. It is also well known that reflections due to lashes, posterior vitreous detachment, misalignment during acquisition and inadequate dilation are other real-world issues that affect image quality in the clinical setting. These miscellaneous factors need to be carefully considered by the developers when planning generalizability studies for algorithms that will be potentially deployed in the population.

5.4 Artificial Intelligence and IOL Power Calculation

IOL power calculation has seen significant improvement over the years. From the highly simplified SRK formula to the complex Barrett Universal II/Barrett Toric formula, targeting 20/20 vision has become possible in a majority of patients undergoing cataract surgery. However, the formulae are still largely based on conventional linear assumptions between the biometric measures and inclusion of parameters like surgeon induced astigmatism (SIA) which is often subjective and difficult to assess on individual basis. With the use of artificial intelligence, it is possible to explore complex non-linear relationships among ocular parameters and generate IOL powers that account for every individual's eye profile. It is also possible to integrate in real time the impact of post-operative outcomes like refractive status post-surgery to calibrate surgeon's performance parameters like SIA.

The Ladas Super Formula (LSF) is an example of an artificial intelligence-derived formula that uses the 'ideal portions' of Hoffer Q, Holladay-1, Holladay-1 with Koch adjustment, Haigis and SRK/T formulae to aid in formula selection [20]. Though exact details of the formula are proprietary and not disclosed in literature, it derives the 'ideal biometric components' of each existing formula and potentially automate the selection process for inexperienced surgeons. It has however been highlighted that LSF has its own limitations. Like traditional formulas, it relies on keratometry to estimate corneal refractive power, which assumes the ratio between a 'uniformly spherical' anterior corneal and the posterior corneal curvature remains unchanged. However, this assumption does not hold true, especially in patients with previous refractive surgery, keratoplasty or even patients with prolonged history of contact lens use [49].

Two additional formulae that use artificial intelligence are the Hill-Radial Basis Function (RBF) and the Kane formula. The Hill-RBF method has been derived using 12,000 eyes with measurements obtained from the Haag-Streit Lenstar optical biometer [50]. The Kane formula has been derived using high-performance cloud-based computing, where regression models and machine learning were used for refinement of IOL power predictions [51]. The accuracy of Hill-RBF method and the Kane formula has been assessed in a wide variety of eyes [52]. The results when comparing formula-estimated and actual post-operative refractive errors are presented in Table 5.3. Though the performance of these formulae is quite good, there is room for improvement in eyes with short axial length. However large-scale studies in eyes with a variety of difficult refractive situations are needed to validate these formulas for clinical use.

The Hill-RBF method has also been compared to Barrett Universal II and the SRK/T formula. In this study, the artificial intelligence-based formula outperformed both the Barrett and SRK/T formula [53]. For the performance estimates, the percentage of post-operative target of eyes within ±0.5 D was 83.62% using Hill-RBF method, 79.66% using Barrett Universal II and 74.01% with the SRK/T formula, indicating the utility of Hill-RBF method in cataract surgery planning. It is also

Table 5.3 The performance metrics (MAE: mean absolute error) of Kane formula and Hill-RBF method of IOL power prediction [52]

	Kane formula	Hill-RBF
Short axial length (≤22.0 mm)	0.441	0.440
Intermediate axial length (>22.0 to <26.0 mm)	0.322	0.340
Long axial length (≥26.0 mm)	0.326	0.358

important to understand that with the increased safety of cataract surgery, the focus has now shifted to implanting premium IOLs. These offer less margin of error especially in demanding patients who expect perfect vision. This industry need has to be addressed by having more reliable and highly flexible formulae that perform well in a multitude of settings.

Another issue is to address the increasing ageing population that has undergone refractive surgery. This is especially relevant in Asian countries with large burden of myopia. Existing formulae were not designed to cater for eyes with past refractive surgery, and it is difficult to account for such patients especially if previous records are not available [54, 55].

Lastly, large, robust and well-defined clinical datasets are essential for development of any new algorithm or refinement of existing algorithm. Often in resource-limited settings where the need of the algorithm is the highest, the medical records, imaging modalities and technical expertise to adapt or develop the algorithm are absent. Thus, it is exceedingly difficult to develop algorithms that have pan regional generalizability. It is also essential to ensure that issues like patient privacy, data ownership and intellectual property rights are adequately addressed and regulated while allowing the steady flow of data, which is the real 'fuel' in the algorithm-driven world.

5.5 Artificial Intelligence and Manpower Training

Recent work (Table 5.4) has shown that artificial intelligence can be used for recognizing different phases and steps of cataract surgery [25, 56, 57]. This can lead to optimized surgical workflow with real-time monitoring and assistance [56, 58]. Systems like these would be potentially useful for training surgeons especially during early career. However, there is much work that needs to be done in this field as traditional machine learning has typically struggled with phase and step detection in cataract surgery. This has been attributed to similar instruments being used during the course of surgery making identification of the exact step difficult for the algorithm. There has been incremental improvement with the use of CNNs but still there is a lot of potential for development of better algorithms [59, 60]. In the current form, these algorithms can be used for assisted annotation that can significantly reduce labelling time and costs [59]. Even with the current performance limitations, innovative algorithms like the V_EBIRD (Video-Based Intelligent Recognition and Decision system) have potential to be implemented as a part of robotic surgery for calibration of ultrasound power during phacoemulsification [61]. The current bottleneck to algorithm development is the lack of availability of high-quality annotated surgical data. However, with the increased availability of recording high-definition video on surgical microscopes, this data is expected to multiply over the next few years. The CATARACTS dataset is one such currently available dataset that can be used by the developers to test their algorithms [56]. In addition to surgical training, these algorithms can be used for surgical review and potentially for

Table 5.4 Previous studies on cataract surgery video annotation and phase recognition

Year	Author	Data source	Method	Definition of gold standard/ground truth	Performance
2013	Lalys et al. [62]	20 cataract surgery videos (19 training, 1 testing)	SVM	Manual annotation	Accuracy: 64.5%
2015	Tian et al. [61]	For eye detection and cataract grading: 1000 video frames for training, 1000 for testing For tracking: 1 video training, 4 videos testing	V_EBIRD 1. Eye detection: Improved Hough transform approach 2. Tracking: Adaptive TLD 3. Cataract identification and grading: SVM	Compared with: 1. Randomized Hough transform 2. Conventional TLD 3. SVM vs kNN (Emery-little classification for cataract grading)	Detection accuracy: 3000 samples; 77.2% 10,000 samples: 92.3% Tracking precision: Adaptive TLD: 87.57% TLD: 55.75% Cataract grading: SVM: 99.2% identification, 96.3% classification kNN: 92.5% classification
2018	Hajj et al. [26]	50 cataract surgery videos, 80 cholecystectomy videos	CNN (feature extraction) + RNN (temporal association)	Manually labelled cataract surgery (A*cs*) videos for 21 tools and 7 tools in cholecystectomy (A*cc*) videos	AUC: Offline mode: (using past + present + future information) A*cs*: 0.996 A*cc*: 0.994 Online mode: (using past+ present information) A*cs*: 0.996 A*cc*: 0.994
2019	Morita et al. [60]	303 surgery videos (245 for training, 10 for verification and 48 for testing)	CNN (inception V3)	Manually labelled for 3 types of forceps for CCC, 4 techniques for nucleus extraction and 2 types of lighting for CCC and nucleus extraction	Accuracy: CCC: 90.7% Nucleus extraction: 94.5% Others: 97.9% MAE: (vs ophthalmologist extracted surgery phase start/end time) CCC start: 3.34 s CCC end: 4.43 s Nucleus extraction start: 7.21 s Nucleus extraction end: 6.04 s

(continued)

Table 5.4 (continued)

Year	Author	Data source	Method	Definition of gold standard/ground truth	Performance
2019	Yu et al. [25]	100 surgery videos with 5-fold cross validation design (threefold for training, 1 for tuning and 1 for testing)	Comparison of 5 algorithms (1. SVM with cross sectional instrument labels, 2. RNN with time series instrument labels, 3. CNN with cross sectional image data, 4. CNN-RNN with time series images and 5. CNN-RNN with time series of images and instrument labels)	Manually annotated 10 labels for surgery phase and 14 labels for instrument use	Accuracy: 1: 0.938 2: 0.959 3: 0.956 4: 0.921 5: 0.915 AUC: 1: 0.737 2: 0.773 3: 0.712 4: 0.752 5: 0.737
2020	Lecuyer et al. [59]	CATARACTS dataset (50 videos)	3 CNNs (VGG19, inception V3, ResNet50)	Manually annotated for different phases and steps of cataract surgery	Accuracy: (cataract phase) VGG19: 79.1% Inception V3: 80.0% ResNet50: 83.3% Accuracy: (cataract step) VGG19: 45.4% Inception V3: 60.5% ResNet50:62.3% User study (phase detection) Assisted system: 97.7% Manual: 96.8% User study (step detection) Assisted system: 77.2% Manual: 70.9% Speed Assisted system: 1767.8 s Manual: 2401.6 s

AUC area under the receiver operating characteristic curve, *CCC* continuous curvilinear capsulorhexis, *CNN* convolutional neural network, *kNN K* nearest neighbour classifiers, *MAE* mean absolute error, *RNN* recurrent neural network, *SVM* support vector machine, *TLD* tracking learning detection accreditation of surgical training using standard guidelines like the ICO-OSCAR guidelines [28, 29].

5.6 Artificial Intelligence and Post-Operative Care/Quality of Life (QoL)

Unified healthcare systems that integrate smartphones and AI-based algorithms are currently pilot tested across the world [39, 63]. These systems like the CC-Guardian offer individualized patient predictions and can become a single point of care for management of chronic conditions [63]. CC-Guardian focuses on management of patients with congenital cataract. These patients are at risk of two most common types of complications, i.e., high intraocular pressure (IOP) and visual axis opacification (VAO). The platform consists of three components: 1. prediction module to identify risk, 2. scheduling/dispatch module to schedule follow-up visit when risk identified, 3. telehealth module for intervention decisions based on follow-up visit. The training dataset included clinical records of 594 congenital cataract patients and 4881 follow-up images (2615 follow-ups, 2266 interventions). Validation was performed in 142 patients with clinical records (61 VAO, 81 non VAO; 79 high IOP, 63 normal) and 1220 follow-up images (671 follow-up, 549 interventions). Ground truth was based on expert panel grading. For performance metrics, the authors reported an AUC of 0.991 for VAO and 0.979 for high IOP. For the telehealth module, the AUC was 0.996. The authors also used another dataset of 79 patients (33 VAO, 46 non VAO; 28 high IOP, 51 normal) for external validation and report an AUC of 0.944 for VAO and 0.961 for high IOP. The authors also performed a cost-effectiveness analysis in another retrospective self-controlled test in 141 patients (93 VOA, 105 high IOP). Accuracy in this subset was 96.8% for VAO and 96.2% for high IOP. Further analysis demonstrated that the patients had 1579 tele health visits (instead of 987 distant visits), reduced travel of 928.6 miles/year and reduced expenditure of $1324/year. These results are a startling example of the profound impact of novel models of healthcare on our current standards of care. While the algorithm may not perform that well in different populations, the design of the study is definitely adaptable for local use. The study is also significant as it not only demonstrates the proof-of-concept but also examines the social and financial impact of real-world implementation. More such studies will be vital in establishing the long-term safety and reliability of AI-based algorithms for public health use. The same group has also performed a randomized control trial comparing CC Cruiser (cataract detection, risk stratification and treatment recommendation platform very similar to CC Guardian) with ophthalmologists in real-world setting [64]. The study consisted of 350 participants randomized to CC Cruiser or senior consultants. The model accuracy for cataract detection and treatment determination was 87.4% and 70.8% as compared to 99.1% and 96.7% for the senior consultant. However, the mean time for diagnosis was 2.79 min for CC Cruiser as compared 8.53 min for senior consultants. The authors additionally

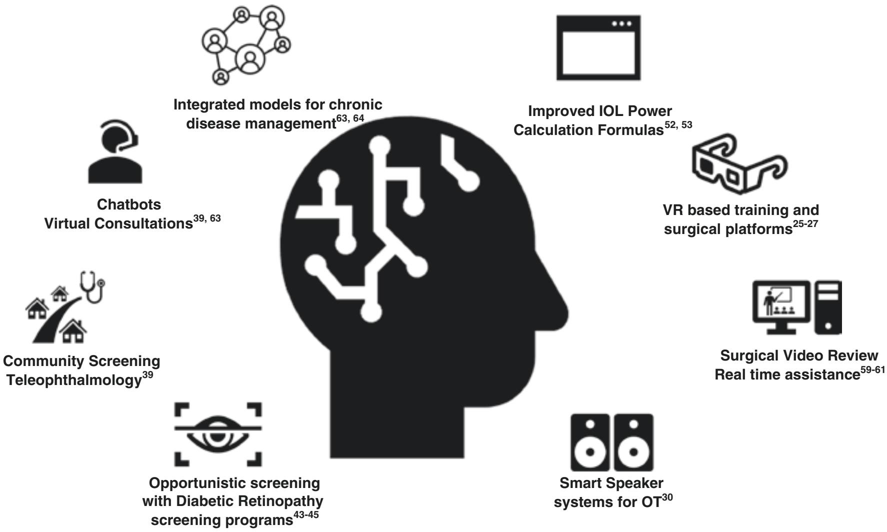


Fig. 5.1 Summary of applications of artificial intelligence in the detection and management of cataracts

conducted a satisfaction survey where mean rating of CC cruiser was 3.47 as compared to 3.38 for the doctors indicating patient acceptance of the AI-based platform. These results show that when algorithms are tested in real world their performance usually drops; however, the benefits of objective reliability, optimum resource utilization, cost effectiveness and improved workflow are upsides that need careful consideration before large-scale deployment.

5.7 Conclusion

With significant improvements in computing power and increased availability of big data via biobanks and disease registries, we are poised to see a seismic shift in how we detect and manage cataracts in the coming years (Fig. 5.1). The next goal in mind would be to translate and integrate these innovations into clinical practice.

Acknowledgement YCT is supported by the National Medical Research Council, Singapore [NMRC/MOH-TA18nov-0002]. The funding organization had no role in the design or conduct of this research.

References

1. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. Lancet Glob Health. 2017;5(12):e1221-e34.

2. Chua J, Lim B, Fenwick EK, et al. Prevalence, risk factors, and impact of undiagnosed visually significant cataract: the Singapore epidemiology of eye diseases study. PLoS One. 2017;12(1):e0170804.
3. Varma R, Mohanty SA, Deneen J, Wu J, Azen SP. Burden and predictors of undetected eye disease in Mexican-Americans: the Los Angeles Latino Eye Study. Med Care. 2008;46(5):497–506.
4. Keel S, McGuiness MB, Foreman J, Taylor HR, Dirani M. The prevalence of visually significant cataract in the Australian National Eye Health Survey. Eye (Lond). 2019;33(6):957–64.
5. Le HG, Ehrlich JR, Venkatesh R, et al. A sustainable model for delivering high-quality, efficient cataract surgery in southern India. Health Aff (Millwood). 2016;35(10):1783–90.
6. Chylack L, Wolfe J, Singer D, et al. The lens opacities classification system III. The longitudinal study of cataract study group. Arch Ophthalmol. 1993;111:831–6.
7. Resnikoff S, Lansingh VC, Washburn L, et al. Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs? Br J Ophthalmol. 2020;104(4):588–92.
8. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. Invest Ophthalmol Vis Sci. 1991;32(2):422–32.
9. Liu Y-C, Wilkins M, Kim T, Malyugin B, Mehta JS. Cataracts. Lancet. 2017;390(10094):600–12.
10. Aristodemou P, Cartwright NEK, Sparrow JM, Johnston RL. Formula choice: Hoffer Q, Holladay 1, or SRK/T and refractive outcomes in 8108 eyes after cataract surgery with biometry by partial coherence interferometry. J Cataract Refract Surg. 2011;37(1):63–71.
11. Chen C, Xu X, Miao Y, et al. Accuracy of intraocular lens power formulas involving 148 eyes with long axial lengths: a retrospective chart-review study. J Ophthalmol. 2015;2015:976847.
12. Zhang Y, Liang XY, Liu S, et al. Accuracy of intraocular lens power calculation formulas for highly myopic eyes. J Ophthalmol. 2016;2016:1917268.
13. Rong X, He W, Zhu Q, et al. Intraocular lens power calculation in eyes with extreme myopia: comparison of Barrett Universal II, Haigis, and Olsen formulas. J Cataract Refract Surg. 2019;45(6):732–7.
14. Kane JX, Van Heerden A, Atik A, Petsoglou C. Intraocular lens power formula accuracy: comparison of 7 formulas. J Cataract Refract Surg. 2016;42(10):1490–500.
15. Olsen T. Calculation of intraocular lens power: a review. Acta Ophthalmol Scand. 2007;85(5):472–85.
16. Wang L, Shirayama M, Ma XJ, Kohnen T, Koch DD. Optimizing intraocular lens power calculations in eyes with axial lengths above 25.0 mm. J Cataract Refract Surg. 2011;37(11):2018–27.
17. Siddiqui AA, Devgan U. Intraocular lens calculations in atypical eyes. Indian J Ophthalmol. 2017;65(12):1289.
18. Chen X, Yuan F, Wu L. Metaanalysis of intraocular lens power calculation after laser refractive surgery in myopic eyes. J Cataract Refract Surg. 2016;42(1):163–70.
19. Wang L, Tang M, Huang D, Weikert MP, Koch DD. Comparison of newer intraocular lens power calculation methods for eyes after corneal refractive surgery. Ophthalmology. 2015;122(12):2443–9.
20. Ladas J, Siddiqui A, Devgan U, Jun A. A 3-D "super surface" combining modern intraocular lens formulas to generate a "super formula" and maximize accuracy. JAMA Ophthalmol. 2015;133:1–6.
21. Tuulonen A, Salminen H, Linna M, Perkola M. The need and total cost of Finnish eyecare services: a simulation model for 2005-2040. Acta Ophthalmol. 2009;87(8):820–9.
22. Turner AW, Mulholland W, Taylor HR. Funding models for outreach ophthalmology services. Clin Exp Ophthalmol. 2011;39(4):350–7.
23. Benzekri R, Marie-Louise J, Chahed S. Cost of teaching cataract surgery in a public hospital. J Fr Ophtalmol. 2017;40(10):860–4.
24. Yen AJ, Ramanathan S. Advanced cataract learning experience in United States ophthalmology residency programs. J Cataract Refract Surg. 2017;43(10):1350–5.

25. Yu F, Silva Croso G, Kim TS, et al. Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. JAMA Netw Open. 2019;2(4):e191860-e.
26. Al Hajj H, Lamard M, Conze PH, Cochener B, Quellec G. Monitoring tool usage in surgery videos using boosted convolutional and recurrent neural networks. Med Image Anal. 2018;47:203–18.
27. Hubschman JP, Son J, Allen B, Schwartz SD, Bourges JL. Evaluation of the motion of surgical instruments during intraocular surgery. Eye (Lond). 2011;25(7):947–53.
28. Golnik KC, Beaver H, Gauba V, et al. Cataract surgical skill assessment. Ophthalmology. 2011;118(2):427.e1-5.
29. Farooqui JH, Jaramillo A, Sharma M, Gomaa A. Use of modified international council of ophthalmology- ophthalmology surgical competency assessment rubric (ICO- OSCAR) for phacoemulsification- wet lab training in residency program. Indian J Ophthalmol. 2017;65(9):898–9.
30. Yoo TK, Oh E, Kim HK, et al. Deep learning-based smart speaker to confirm surgical sites for cataract surgeries: a pilot study. PLoS One. 2020;15(4):e0231322.
31. Maloley L, Morgan LA, High R, Suh DW. Wrong-site surgery in pediatric ophthalmology. J Pediatr Ophthalmol Strabismus. 2018;55(3):152–8.
32. Bian Y, Xiang Y, Tong B, Feng B, Weng X. Artificial intelligence-assisted system in postoperative follow-up of orthopedic patients: exploratory quantitative and qualitative study. J Med Internet Res. 2020;22(5):e16896.
33. Wu X, Chen J, Yun D, et al. Effectiveness of an ophthalmic hospital-based virtual service during COVID-19. Ophthalmology. 2020:S0161-6420(20)31010-1.
34. Mohammadi SF, Sabbaghi M, H ZM, et al. Using artificial intelligence to predict the risk for posterior capsule opacification after phacoemulsification. J Cataract Refract Surg. 2012;38(3):403–8.
35. Li H, Chutatape O. Boundary detection of optic disk by a modified ASM method. Pattern Recogn. 2003;36(9):2093–104.
36. Xu Y, Gao X, Lin S, et al., editors. Automatic grading of nuclear cataracts from slit-lamp Lens images using group sparsity regression. Berlin, Heidelberg: Springer; 2013.
37. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. IEEE Trans Biomed Eng. 2015;62(11):2693-701.
38. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719–31.
39. Wu X, Huang Y, Liu Z, et al. Universal artificial intelligence platform for collaborative management of cataracts. Br J Ophthalmol. 2019;103(11):1553.
40. Li H, Lim H, Liu J, et al. An automatic diagnosis system of nuclear cataract using slit-lamp images. In: Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference 2009; p. 3693–6.
41. Lian JX, Gangwani RA, McGhee SM, et al. Systematic screening for diabetic retinopathy (DR) in Hong Kong: prevalence of DR and visual impairment among diabetic population. Br J Ophthalmol. 2016;100(2):151.
42. Prescott G, Sharp P, Goatman K, et al. Improving the cost-effectiveness of photographic screening for diabetic macular oedema: a prospective, multi-centre, UK study. Br J Ophthalmol. 2014;98(8):1042.
43. Dong Y, Zhang Q, Qiao Z, Yang J, editors. Classification of cataract fundus image based on deep learning. In: 2017 IEEE international conference on imaging systems and techniques (IST); 18–20 Oct 2017.
44. Ran J, Niu K, He Z, Zhang H, Song H, editors. Cataract detection and grading based on combination of deep convolutional neural network and random forests. In: 2018 international conference on network infrastructure and digital content (IC-NIDC); 22–24 Aug 2018.
45. Pratap T, Kokil P. Computer-aided diagnosis of cataract using deep transfer learning. Biomed Signal Proc Contr. 2019;53:101533.

46. Yitao L, Lianlian H, Chao F, Feng W, Wei L, editors. Preprocessing study of retinal image based on component extraction. 2008 IEEE international symposium on it in medicine and education; 12–14 Dec 2008.
47. Linglin Z, Jianqiang L, et al., editors. Automatic cataract detection and grading using Deep Convolutional Neural Network. In: 2017 IEEE 14th international conference on networking, sensing and control (ICNSC); May 2017. p. 16–8.
48. Li J, Xu X, Guan Y, et al., editors. Automatic cataract diagnosis by image-based interpretability. In: 2018 IEEE international conference on systems, man, and cybernetics (SMC); 7–10 Oct 2018.
49. Hee MR. State-of-the-art of intraocular lens power formulas. JAMA Ophthalmol. 2015;133(12):1436–7.
50. Forman D, Newell DG, Fullerton F, et al. Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation. BMJ. 1991;302(6788):1302–5.
51. Melles RB, Kane JX, Olsen T, Chang WJ. Update on intraocular lens calculation formulas. Ophthalmology. 2019;126(9):1334–5.
52. Connell BJ, Kane JX. Comparison of the Kane formula with existing formulas for intraocular lens power selection. BMJ Open Ophthalmol. 2019;4(1):e000251.
53. Nemeth G, Modis L Jr. Accuracy of the Hill-radial basis function method and the Barrett Universal II formula. Eur J Ophthalmol. 2020:1120672120902952.
54. Hoffer KJ. Intraocular lens power calculation after previous laser refractive surgery. J Cataract Refract Surg. 2009;35(4):759–65.
55. Alio JL, Abdelghany AA, Abdou AA, Maldonado M. Cataract surgery on the previous corneal refractive surgery patient. Surv Ophthalmol. 2016;61(6):769–77.
56. Al Hajj H, Lamard M, Conze P-H, et al. CATARACTS: challenge on automatic tool annotation for cataRACT surgery. Med Image Anal. 2019;52:24–41.
57. Zisimopoulos O, Flouty E, Luengo I, et al. DeepPhase: surgical phase recognition in CATARACTS videos. arXivorg. 2018.
58. Maier-Hein L, Speidel S, Navab N, et al. Surgical data science: enabling next-generation surgery. arXivorg. 2017;1(9).
59. Lecuyer G, Ragot M, Martin N, Launay L, Jannin P. Assisted phase and step annotation for surgical videos. Int J Comput Assist Radiol Surg. 2020;15(4):673–80.
60. Morita S, Tabuchi H, Masumoto H, Yamauchi T, Kamiura N. Real-time extraction of important surgical phases in cataract surgery videos. Sci Rep. 2019;9(1):16590.
61. Tian S, Yin X-C, Wang Z-B, Zhou F, Hao H-WA. VidEo-based intelligent recognition and decision system for the phacoemulsification cataract surgery. Comput Math Methods Med. 2015;2015:202934.
62. Lalys F, Bouget D, Riffaud L, Jannin P. Automatic knowledge-based recognition of low-level tasks in ophthalmological procedures. Int J Comput Assist Radiol Surg. 2013;8(1):39–49.
63. Long E, Chen J, Wu X, et al. Artificial intelligence manages congenital cataract with individualized prediction and telehealth computing. NPJ Digit Med. 2020;3:112.
64. Lin H, Li R, Liu Z, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. EClinicalMedicine. 2019;9:52–9.

Artificial Intelligence and Glaucoma

6

Sidong Liu, Yuyi You, and Stuart L. Graham

6.1 Introduction

Glaucoma is a group of optic neuropathies characterized by progressive optic nerve degeneration and loss of retinal ganglion cells (RGC) [1]. Glaucoma progresses without causing symptoms until the disease is advanced with substantial neural damage; when symptoms do occur, as many as 30–50% of retinal ganglion cells may be lost [2], resulting in irreversible visual field (VF) loss with concomitant reduction in quality of life. Glaucoma is the second leading cause of irreversible blindness worldwide, currently affecting more than 80 million people globally and being predicted to affect 110 million in 2040 [3]. Risk factors for glaucoma include increased intraocular pressure (IOP), a family history of the disease, older age (>50), African or Asian descent, myopia and use of systemic or topical corticosteroids [4]. Primary open-angle glaucoma (POAG) is the most common type with the highest prevalence (3.05%), followed by primary angle-closure glaucoma (PACG, 0.5%) worldwide [3]. Currently there is no cure for glaucoma, and it is important to detect the disease as early as possible, so that IOP-lowering treatment can be initiated to avoid irreversible visual functional loss.

Recent advances in artificial intelligence (AI), especially the advent of deep learning (DL), have shown transformative impact on the healthcare industry,

S. Liu
Centre for Health Informatics, Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, NSW, Australia
e-mail: sidong.liu@mq.edu.au

Y. You (✉) · S. L. Graham
Department of Clinical Medicine, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, NSW, Australia
e-mail: yuyi.you@mq.edu.au; stuart.graham@mq.edu.au

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_6

demonstrating outstanding performance in skin cancer classification [5], early diagnosis of Alzheimer's disease [6], glioma prognosis [7], diabetic retinopathy detection [8], and most recently COVID-19 severity assessment [9], as well as many other applications. Several AI systems have been proposed for detecting signs of glaucoma-related structural and functional damage and for the diagnosis and clinical evaluation of glaucoma. This chapter aims to provide an overview of the applications of AI in glaucoma with a focus on the DL models, then discuss their clinical and technical challenges, and finally project potential development of AI in glaucoma in future work.

6.2 Overview of AI Systems in Glaucoma

AI systems in glaucoma are predominately based on imaging data derived from VF tests (perimetry), fundus photography, and optical coherence tomography (OCT), as these modalities provide highly structured data that are suited for training the AI models. The glaucoma AI systems can be broadly grouped into two categories based on how the models process the data to obtain classification/prediction results.

6.2.1 Classic Machine Learning Models

The first category systems are using classic machine learning methodology, which usually consists of three steps in the workflow, as shown in Fig. 6.1, to transform the input data into feature vectors and then make a prediction based on those feature vectors. The first step is to preprocess the input images, including removal of artifact presented in the image and enhance segmentation of the region of interest, such as the optic cup (OC) and the optic disc (OD). The next step is to extract distinctive features, including both clinical indicators (e.g., cup-to-disc ratio, CDR) and visual features (e.g., spectral, morphological, and texture features), from the preprocessed images. Then a machine learning model, such as support vector machine [10] and naïve Bayes [11], can be trained using the extracted feature vectors for the classification/prediction tasks. The preprocessing techniques, features, and classification algorithms used in the first category of systems for glaucoma have been comprehensively reviewed in previous studies [12–14].

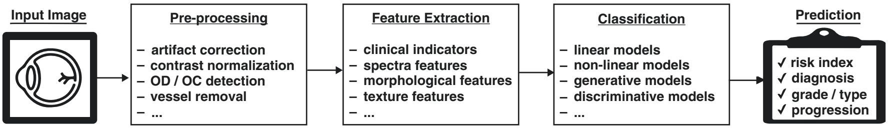


Fig. 6.1 Diagram of a classic machine learning methodology for glaucoma classification

6.2.2 Deep Learning Models

The second category of AI systems is based on deep learning (DL), which is a subset of machine learning and more sophisticated than classic machine learning algorithms. DL methods show that good representations of data can be learned automatically using a multilayer neural network with multiple levels of abstraction. A typical deep convolutional neural network (CNN) model uses only pixel values as inputs and can be trained end to end from images to categorical labels or continuous variables directly. For example, Fig. 6.2 presents a multilayer CNN model, which takes the cropped OD photographs as input and then predicts the risk of glaucoma [15]. DL models can discover intricate pattern in the data by automatically adjusting the internal parameters of the models that are used to compute the representations of the data, with no need for feature engineering or conducting feature extraction and classification in a two-stage process.

6.3 Applications of AI in Glaucoma

AI has shown a great potential for the clinical management of glaucoma, with an increasing number of AI models proposed for detecting glaucoma-related structural and functional changes, as well as for the clinical diagnosis of glaucoma and prognosis prediction. In this section, we will summarize the roles of AI in three typical applications, including (1) detection, (2) diagnosis, and (3) prognosis.

6.3.1 Detection of Glaucomatous Signs

AI-based glaucoma detection intends to identify, mark, highlight, or direct attention to portions of the input data that may have glaucoma-related structural and functional abnormalities, and/or to extract measurements and features that can quantify such abnormalities, so that healthcare providers can use the visual and quantitative information to detect glaucoma or monitor changes during patient

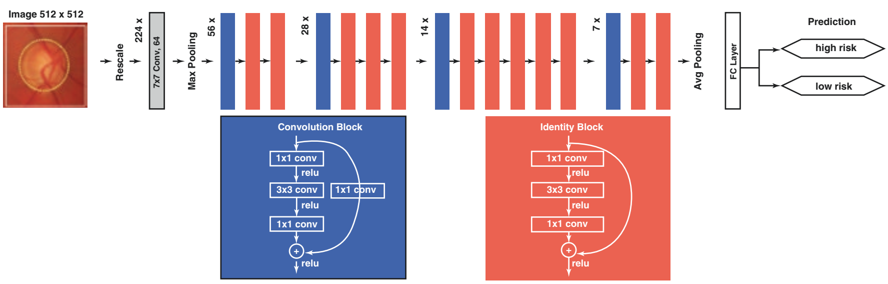


Fig. 6.2 Architecture of the DL model that was derived from an glaucoma screening system described by Liu et al. [15]

follow-up. AI-based detection methods are usually designed for mono-modal data, e.g., OD/OC segmentation methods are predominantly used in fundus photograph analysis.

OD/OC Segmentation Examination of the optic nerve head in retinal images is very important in detecting the glaucoma-related damage. Various AI-based techniques have been developed to segment the OD and OC in retinal images and to calculate CDR. Prior to the DL methods, contour-based methods [16, 17] and superpixel classification [18] were among the most common techniques for OD and OC segmentation. These prior art OD/OC segmentation methodologies for glaucoma image detection have been comprehensively reviewed [12, 13]. With the advances in DL algorithms, the concurrent high-performance computing facilities, and the emerging large-scale public imaging datasets, DL-based models have been increasingly used in OD/OC segmentation, with the U-Net representing the state-of-the-art [19, 20]. In addition to local datasets used in different studies, a few public datasets, such as RIM-ONE [21], Drishti-GS [22], ORIGA [23], and HRF [24], are often used for development and validation of OC/OD segmentation algorithms. A recent effort along this line was the international Retinal Fundus Glaucoma Challenge (REFUGE) [25], which proposed a standardized evaluation framework to compare different algorithms in OD/OC segmentation and glaucoma classification.

VF Loss Detection Computerized automated VF testing is a cornerstone in detecting the functional changes induced by glaucoma. Compared to fundus photographs and OCT scans, VFs are low-dimensional psychophysical data which usually include reliability parameters, age-matched sensitivity arrays across visual space, and global indices that summarize visual function. In a pioneering work by Goldbaum et al. [26], the first neural network for VF analysis was proposed, which used the VF position values as input and derived a topological VF defect pattern for glaucoma patients. More recently, Elze et al. [27] developed an unsupervised machine learning method, known as archetypal analysis, to further leverage regional sensitivity data contained in VFs. Archetypal analysis provides a regional stratification of VF and assigns a weighting coefficient to each of these regional patterns. As demonstrated in a subsequent study [28], some archetypes were found to have higher correlation with glaucoma than others; patients with high weighting coefficients for a glaucoma-correlated archetype were more likely to have high CDRs. Mayro et al. also summarized the roles of AI in glaucoma with an emphasis on VF loss detection [29].

Other applications of AI for glaucoma detection may include tissue segmentation [30], retinal vessel segmentation [31], visualization [32], abnormality detection [33], and risk estimation [34]. A typical use case of these models is population-based screening by general healthcare providers to achieve early glaucoma detection and specialist referral.

6.3.2 Diagnosis of Glaucoma

AI-based diagnostic models for glaucoma intend to aid in the classification of the parameters and features extracted from the input data and to provide diagnostic or phenotype recommendations. These diagnostic models, either mono-modal or multimodal, are typically used to provide a second opinion to clinicians in most scenarios, although there are autonomous AI-based diagnostic systems that can work without clinical users' involvement, such as the IDx-DR system for diabetic retinopathy detection [35].

Mono-Modal Data-Based Diagnosis The majority of AI-based diagnostic models for glaucoma use retinal imaging data, as they provide rich information, such as color, texture, and morphology, of the optic nerve. Previous methodologies that were used to extract these features through feature engineering and to train the diagnostic models have been extensively reviewed [12–14]. DL models further leverage the structural details in imaging data without feature engineering and represent the state-of-the-art in glaucoma diagnosis. In one of the initial studies, Chen et al. [36] proposed a glaucoma classification model based on a simple neural network for fundus photographs, and achieved an area under the curve (AUC) of 0.898. More sophisticated models were then built as the neural network architecture became deeper. VGG [37–39] and ResNet [15, 40–42] architectures are among the most popular neural network models for structural imaging data analysis. A number of selected diagnostic models are presented in Table 6.1. Although various dataset were used to train and test these models, the models' performance were consistently high. For example, for the task of differentiating glaucoma and non-glaucoma cases using fundus photos, most models achieved an AUC of 0.9 and above, and the best reported performance so far is an AUC of 0.996, with sensitivity of 0.962 and specificity of 0.977, respectively, achieved by a ResNet variant which was trained and tested using a total of 241,032 fundus photos [42].

Goldbaum et al. [26] proposed the first neural network model for VF-based glaucoma diagnosis, which achieved comparable performance to two glaucoma specialists. DL has further exploited the regional sensitivity data in VF to assist glaucoma diagnosis. For example, Asaoka et al. [43] developed a four-layer fully connected neural network (FNN) for detecting pre-perimetric glaucoma and achieved an AUC of 0.926, which is higher than the compared machine learning methods. In another study, Li et al. [44] proposed a DL model based on VGG-16 architecture for differentiating glaucomatous from non-glaucomatous cases using their VF data. This model demonstrated a high potential for clinical application with an AUC of 0.966, sensitivity of 0.932, and specificity of 0.826, tested on 300 subjects. Table 6.1 listed the datasets, models, and the reported performance in the abovementioned studies. Devalla et al. [14] provided a summary of the AI studies using functional data for glaucoma diagnosis, including both classic machine learning models and DL models.

Multimodal Data-Based Diagnosis Different modalities may provide complementary information to each other. As demonstrated in an early study by Brigatti et al.

Table 6.1 Selected studies on glaucoma diagnosis and prognosis using neural network models (in chronological order)

Authors	Year	Dataset	Modality	Model	Application	AUC	Sensitivity	Specificity
Goldbaum et al. [26]	1994	Local dataset	Visual fields	A two-layer FNN	Diagnosis	–	0.65	0.72
Brigatti et al. [45]	1996	Local dataset	Visual fields; OD and RNFL measurements	A four-layer BPNN	Diagnosis: Functional	–	0.84	0.86
					Diagnosis: Structural	–	0.87	0.56
					Diagnosis: Multimodal	–	0.90	0.84
Chen et al. [36]	2015	ORIGA [23]; SCES*	Fundus photos	A six-layer CNN	Diagnosis	0.898	–	–
Orlando et al. [37]	2016	Drishti-GS1 [22]	Fundus photos	OverFeat; VGG-S	Diagnosis	0.763	–	–
Asaoka et al. [43]	2016	Local dataset	Visual fields	A four-layer FNN	Diagnosis	0.926	0.749	1.000
Ting et al. [38]	2017	SiDRP 14–15#	Fundus photos	VGG-19	Diagnosis	0.942	0.964	0.932
Liu et al. [15]	2018	Local dataset; HRF [31]; RIM-ONE [21]	Fundus photos	ResNet50	Diagnosis: Local dataset	0.970	0.893	0.971
					Diagnosis: HRF	0.890	0.867	0.867
Li et al. [50]	2018	LabelMe [51]	Fundus photos	Inception_v3	Diagnosis	0.986	0.956	0.920
Shibata et al. [40]	2018	Local dataset	Fundus photos	ResNet	Diagnosis	0.965	–	–
Li et al. [44]	2018	Local dataset	Visual fields	VGG-16	Diagnosis	0.966	0.932	0.826
Christopher et al. [41]	2018	ADAGES$; DIGS~	Fundus photos	ResNet50	Diagnosis	0.910	0.840	0.830
Medeiros et al. [46]	2019	Local dataset	OCT scans; fundus photos	ResNet34	Diagnosis	0.944	0.900	0.800
Liu et al. [42]	2019	CGSA^	Fundus photos	ResNet	Diagnosis	0.996	0.962	0.977

Asaoka et al. [52]	2019	Local dataset	OCT scans	A 12-layer CNN	Diagnosis	0.937	0.825	0.939
Fu et al. [39]	2019	Local dataset	AS-OCT scans	VGG-16	Diagnosis	0.96	0.90	0.92
Normando et al. [33]	2020	Local dataset	OCT scans	MobileNet_v2	Diagnosis	–	0.911	0.971
					Prognosis	0.890	0.857	0.917
Thakur et al. [49]	2020	OHTS+	Fundus photos	MobileNet_v2	Diagnosis	0.94	–	–
					Prognosis: 1–3 years	0.88	–	–
					Prognosis: 4–7 years	0.77	–	–

SiDRP14–15#: Singapore Integrated Diabetic Retinopathy Screening Programme; SCES*: Singapore Chinese Eye Study; CGSA^: Chinese Glaucoma Study Alliance; ADAGES$: The African Descent and Glaucoma Evaluation Study; DIGS~:Diagnostic Innovation in Glaucoma Study; OHTS+: The Ocular Hypertension Treatment Study

[45], when the same back propagation neural network (BPNN) model was applied to mono-modal data (functional: VF; structural: OD/RNFL measurements) and multimodal data, the model achieved better performance on multi-modal data than either functional or structural data alone. Another benefit of multimodal approach is that a model can be trained to predict one modality using another, so that the model can overcome certain defects. For example, Mederiros et al. [46] proposed a new approach to train a DL model for quantification of structural damage on OD photographs. This model, using spectral-domain OCT measurements as the reference standard rather than highly variable assessment by human raters, therefore, can objectively evaluate the degree of neural damage and predict glaucoma with higher reliability. A few more examples of the models using multimodal approaches can be found in a recent review [14].

6.3.3 Progression and Prognosis of Glaucoma

Once diagnosed, it is important to predict the likely progression of glaucoma to avoid overtreatment or undertreatment. However, glaucoma progresses non-linearly and can be affected by multiple factors, which poses a great challenge on prognosis evaluation of glaucoma. There is no widely accepted clinical tests that could predict glaucoma progression, and the assessment depends heavily on clinicians' expertise and experience and usually requires multiple clinic visits.

Trajectory-Based Prediction A few prediction models have been proposed for glaucoma prognosis, which require serial data of glaucoma patients to build progression trajectories. For example, Kalman filtering, a widely used forecasting method, has helped increase efficiency (29% less tests) and reduce delay (57% sooner than fixed interval monitoring system) in identifying glaucoma progression [47], and predict mean VF and IOP measurements for normal tension glaucoma [48]. Using archetypal analysis, Wang et al. [28] developed a method to calculate the change rate in the weighting coefficients associated with different archetypes and built a prediction model based on a large cohort with serial tests. This model achieved an accuracy of 0.77 in glaucoma progression prediction. Mayro et al. [29] provided a comparison of these trajectory-based prediction models.

Single Time-Point Prediction More recently, a few DL models have been developed to predict glaucoma progression using the data at a single time-point, which, if achievable, are more cost-effective than the models that require serial data. However, it is very challenging to develop such models, as predictions need to be made prior to the presence of clinical manifestation of glaucoma. Two examples are listed in Table 6.1. Thakur et al. [49] attempted to develop a DL model for prediction of OD or VF abnormalities from fundus photographs. This model achieved an AUC of 0.77 when predicting glaucoma progression 4–7 years prior to disease onset, and the AUC was higher (0.88) when predicting progression 1–3 years prior to disease onset. In another study, Normando et al. [33] proposed a CNN-aided method to predict glaucoma progression using a Detection of Apoptosing Retinal Cells (DARC) test, which can visualize apoptotic retinal cells in the retina in humans. The CNN model in this study was used not to predict progression but to detect DARC positively stained cells in the retinal fluorescent image. Using OCT RNFL measurements at 18 months as reference standard, the CNN-aided DARC test demonstrated 0.857 sensitivity and 0.917 specificity with an AUC of 0.89 in differentiating the rapid-progressing eyes from stable eyes.

6.4 Challenges of AI in Glaucoma

6.4.1 Dataset Dependence

Training datasets have an enormous impact on AI models, especially DL models which require a large number of labeled training samples. It is usually a complex endeavor to collect, select, and annotate/label many samples, which may take years. It is not surprising to see that an AI model could not maintain good performance on samples taken from different ethnic groups, using different machines and protocols and with different qualities. The dependence on dataset may lead to compromised performance on heterogeneous samples hence the potential clinical application of the AI models for glaucoma.

6.4.2 Disagreement Between Clinicians

Glaucoma is not like diseases such as diabetic retinopathy with well-understood reference standards. There can be disagreement between clinicians—ophthalmologists or even glaucoma subspecialists—in the assessment of glaucoma patients. Diagnosis and prognosis of glaucoma depend heavily on clinicians' expertise and the clinical tests available. Even if the samples are highly homogeneous, disagreement between different clinicians may induce variation to the ground truth labeling, which will then propagate during training and lead to bias in the model's output.

6.4.3 Early Glaucoma

From the pathologic perspective, about 50% glaucoma cases are undiagnosed until a relatively late state as glaucoma progresses without causing symptoms in its early stage [53]. Early diagnosis is important, so that the treatment can be escalated. However, it can be more difficult for AI systems to detect cases with less-severe disease manifestations, such as glaucoma suspect and pre-perimetric glaucoma, compared to severe and advanced glaucoma. There is no widely accepted method for confirmed diagnosis of early glaucoma although the World Glaucoma Association has published a consensus document that defines the main features [54]. The model proposed by Thakur et al. [49] shows promising results in predicting glaucoma conversion a couple of years prior to disease onset, but it is still not clear how their model recognizes the early signs of glaucoma.

6.4.4 Comorbid Eye Conditions

AI systems also face the challenge of analyzing images with multiple comorbid eye conditions. As pointed out by Li et al. [50], coexistence of high or pathologic myopia is the most common cause of false-negative results produced by their model, which was trained using fundus photographs of a large Chinese cohort. Another example is that VF represents a functional assay of the entire visual pathway, so VF loss associated with other pathologies, such as lesions in the posterior visual pathway [55], may pose a challenge in differentiating glaucomatous from non-glaucomatous damage. It is undoubtly important but also very challenging to take these confounding factors into consideration when designing AI systems for glaucoma.

6.5 Potential Future Development

6.5.1 Portable Equipment and Cloud Platform

As people are living longer, there will be increased demand of ophthalmic services in the community due to higher prevalence of glaucoma and other age-related ocular disorders. Efficient glaucoma services, such as screening, therefore, are an essential part of that future landscape. Currently glaucoma screening is still expensive to implement at both community and hospital levels. Portable equipment and cloud platform might be a feasible solution, as demonstrated by a few regional and national diabetic retinopathy screening projects detailed in [56]. The same idea could be applied to glaucoma screening. We developed a platform that can detect glaucomatous signs using portable fundus cameras and an AI model on the cloud [57], as illustrated in Fig. 6.3. After the examinee's fundus photographs are taken by

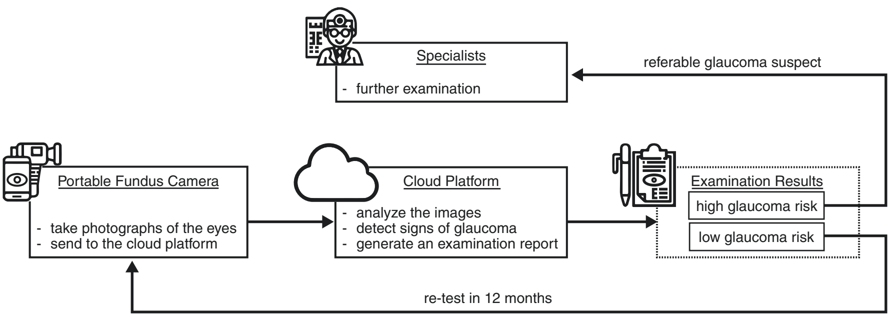


Fig. 6.3 Schematic diagram of the AI platform for glaucoma screening [57]

an intended user, e.g., optometrists, general practitioners, or other healthcare providers, the photographs will be sent to a cloud platform for analysis. The cloud platform will then generate an examination report. People with a result of high glaucoma likelihood can be referred to specialists for further examination, otherwise retest will be suggested in 12 months.

6.5.2 Augmented Intelligence

AI can help clinicians with decision-making to reduce errors in diagnosis and to improve patient outcome. As pointed out by A. Di Ieva [58], there is “underlying fear of a dystopic challenge wherein AI is in competition with human experts” and such fear can be overcome by viewing AI as a means to enhance the expertise of human. This alternative view would shift the paradigm from human versus AI to human with AI, i.e., augmented intelligence.

From the clinicians’ perspective, given more and more clinicians are expected to use AI in the future, the way to prepare the new-generation clinicians for the coming AI times is to adapt medical education to the digital world and to build the capacity for enhanced decision-making and prognostication by means of AI [59]. It is equally important that clinicians need to stay vigilant about the incompetence of AI to avoid blindly following the decision made by the machine. From the AI engineers’ perspective, we should use “human-in-the-loop” approach to actively engage clinicians in the full course of development, testing, and implementation of AI systems. With the clinicians’ input, AI models are likely to learn much more rapidly. It is also necessary for clinicians to continuously provide feedback to AI engineers to prevent system-wide failure.

6.5.3 Explainable AI

Many AI models, particularly DL models, are having the interpretability problem, i.e., the models’ predictive mechanism is unknown and we cannot explain how AI arrives at a specific decision. The benefits of explainable AI are multifold; for example, it can increase our trust in the models; the way AI perceives data may enlighten clinicians about new diagnostic and prognostic biomarkers which may lead to new findings about pathological mechanisms for the disease.

There is an ongoing progress in explainable AI in glaucoma. In a pioneering work, Goldbaum et al. [26] attempted to interpret the perimetry results of glaucoma by mapping feature weights to VF regions. A more recent example is archetype analysis, which outputs the coefficient for each characteristic archetypes. Notably, many advanced AI models for fundus photographs and OCT scans are able to visualize the suspect pathologies or saliency areas in the images [32, 33]. Further research will be needed to continuously improve the explainable AI models.

6.5.4 Converged Technologies

Another potential future development lies in the combination of different technologies with the help of AI. Imaging genetics, for example, is a very promising area for the future, where the aim is to identify the genetic basis of anatomical and functional abnormalities and to show how this relates to glaucoma. Khawaja et al. [60] identified 112 genomic loci associated with IOP and the development of glaucoma, and a regression model based on these loci achieved an AUC of 0.76 in glaucoma classification. In a recent work, Margeta et al. [61] found that the APOE ε4 allele is associated with a reduced risk of POAG, suggesting a protective effect for APOE ε4 in glaucoma. With the converted technologies, AI will help identify more efficiently and effectively the diagnostic, prognostic, and therapeutic biomarkers for glaucoma.

6.6 Conclusion

AI is playing several roles in the management of glaucoma, such as detecting signs of structural and functional damage, and assisting in disease diagnosis, but it is still far from reaching its potential. It remains dependant on the accuracy of the training data sets. While there are clinical and technical challenges in using AI in real-life settings, future research will likely accelerate emergence of effective AI systems in glaucoma practice.

References

1. Weinreb RN, Khaw PT. Primary open-angle glaucoma. Lancet (London). 2004;363(9422):1711–20.
2. Quigley HA, Addicks EM, Green WR, Maumenee AE. Optic nerve damage in human glaucoma. II. The site of injury and susceptibility to damage. Arch Ophthalmol. 1981;99(4):635–49.
3. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology. 2014;121(11):2081–90.
4. Hollands H, Johnson D, Hollands S, et al. Do findings on routine examination identify patients at risk for primary open-angle glaucoma? The rational clinical examination systematic review. JAMA. 2013;309(19):2035–42.
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8.
6. Liu S, Liu S, Cai W, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans Biomed Eng. 2015;62(4):1132–40.
7. Liu S, Shah Z, Sav A, et al. Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. Sci Rep. 2020;10:7733.
8. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10.

9. Feng YZ, Liu S, Cheng ZY, Quiroz JC, Rezazadegan D, Chen PK, et al. Severity assessment and progression prediction of COVID-19 patients based on the lesionencoder framework and chest CT. 2020. https://www.medrxiv.org/content/10.1101/2020.08.03.20167007v2.full.pdf.
10. Acharya UR, Dua S, Du XVSS, Chua CK. Automated diagnosis of glaucoma using texture and higher order spectra features. IEEE Trans Inf Technol Biomed. 2011;15(3):449–55.
11. Noronha KP, Acharya UR, Nayak KP, Martis RJ, Bhandary SV. Automated classification of glaucoma stages using higher order cumulant features. Biomed Signal Proc Control. 2014;10:174–83.
12. Almazroa A, Burman R, Raahemifar K, Lakshminarayanan V. Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey. J Ophthalmol. 2015;2015:180972.
13. Hagiwara Y, Koh JEW, Tan JH, Bhandary SV. Computer-aided diagnosis of glaucoma using fundus images: a review. Comput Methods Prog Biomed. 2018;165:1–12.
14. Devalla SK, Liang Z, Pham TH, Boote C, Strouthidis NG, Thiery AH, Girard MJA. Glaucoma management in the era of artificial intelligence. Br J Ophthalmol. 2020;104(3):301–11.
15. Liu S, Graham SL, Schulz A, Kalloniatis M, Zangerl B, Cai W, Gao Y, Chua B, Arvind H, Grigg J, Chu D, Klistorner A, You Y. A deep learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs. Ophthalmol Glaucoma. 2018;1(1):15–22.
16. Joshi GD, Sivaswamy J, Krishnadas SR. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Trans Med Imaging. 2011;30(6):1192–205.
17. Lu S. Accurate and efficient optic disc detection and segmentation by a circular transformation. IEEE Trans Med Imaging. 2011;30(2):2126–33.
18. Cheng J, et al. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans Med Imaging. 2013;32(6):1019–32.
19. Sevastopolsky A. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. Pattern Recognit Image Anal. 2017;27(3):618–24.
20. Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans Med Imaging. 2018;37(7):1597–605.
21. Fumero F, Alayon S, Sanchez JL, Sigut J, Gonzalez-Hernandez M. RIM-ONE: an open retinal image database for optic nerve evaluation. In: 2011 24th international symposium on computer-based medical systems (CBMS). 2011. p. 1–6.
22. Sivaswamy J, Krishnadas SR, Joshi GD, Ujjwal MJ, Tabish S. Drishti GS: Retinal image dataset for optic nerve head(ONH) segmentation. In: 2014 IEEE 11th international symposium on biomedical imaging, ISBI 2014. 2014. p. 53–6.
23. Zhang Z, Yin FS, Liu J. ORIGA(-light): an online retinal fundus image database for glaucoma analysis and research. In: Conference proceedings of annual international conference of the IEEE engineering in medicine and biology society. 2010. p. 3065–68.
24. Budai A, Bock R, Maier A, Hornegger J, Michelson G. Robust vessel segmentation in fundus images. Int J Biomed Imaging. 2013;2013:154860.
25. Orlando JI, Fu H, Barbosa Breda J, van Keer K, Bathula DR, Diaz-Pinto A, et al. REFUGE challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Med Image Anal. 2020;101570:59.
26. Goldbaum MH, Sample PA, White H, Côlt B, Raphaelian P, Fechtner RD, Weinreb RN. Interpretation of automated perimetry for glaucoma by neural network. Invest Ophthalmol Vis Sci. 1994;35(9):3362–73.
27. Elze T, Pasquale LR, Shen LQ, Chen TC, Wiggs JL, Bex PJ. Patterns of functional vision loss in glaucoma determined with archetypal analysis. J R Soc Interface. 2015;12(103):20141118.
28. Wang M, Pasquale LR, Shen LQ, Boland MV, Wellik SR, De Moraes CG, et al. Reversal of glaucoma hemifield test results and visual field features in glaucoma. Ophthalmology. 2018;125(3):352–60.
29. Mayro EL, Wang M, Elze T, Pasquale LR. The impact of artificial intelligence in the diagnosis and management of glaucoma. Eye (Lond). 2020;34(1):1–11.

30. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342–50.
31. Odstrcilik J, Kolar R, Budai A, Hornegger J, Jan J, Gazarek J, et al. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. IET Image Process. 2013;7(4):373–83.
32. Keel S, Wu J, Lee PY, Scheetz J, He M. Visualizing deep learning models for the detection of referable diabetic retinopathy and glaucoma. JAMA Ophthalmol. 2019;137(3):288–92.
33. Normando EM, Yap TE, Maddison J, Miodragovic S, Bonetti P, Almonte M, Mohammad NG, Ameen S, Crawley L, Ahmed F, Bloom PA, Cordeiro MF. A CNN-aided method to predict glaucoma progression using DARC (Detection of Apoptosing Retinal Cells). Expert Rev Mol Diagn. 2020;20(7):737–48.
34. Bock R, Meier J, Nyúl LG, Hornegger J, Michelson G. Glaucoma risk index: automated glaucoma detection from color fundus images. Med Image Anal. 2010;14(3):471–81.
35. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018;1:39.
36. Chen X, Xu Y, Yan S, Wong DWK, Wong TY, Liu J. Automatic feature learning for glaucoma detection based on deep learning. BT Med Imag Comput Comput-Assist Intervent MICCAI. 2015;2015:669–77.
37. Orlando JI, Prokofyeva E, del Fresno M, Blaschko MB. Convolutional neural network transfer for automated glaucoma identification. In: Proceedings of the SPIE. 2017, vol. 10160.
38. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 2017;318(22):2211–23.
39. Fu H, Baskaran M, Xu Y, Lin S, Wong DWK, Liu J, Tun TA, Mahesh M, Perera SA, Aung T. A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images. Am J Ophthalmol. 2019;203:37–45.
40. Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, Asaoka R. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Sci Rep. 2018;8(1):14665.
41. Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, Girkin CA, Liebmann JM, Zangwill LM. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci Rep. 2018;8(1):16685.
42. Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. JAMA Ophthalmol. 2019;137(12):1353–60.
43. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. Ophthalmology. 2016;123(9):1974–80.
44. Li F, Wang Z, Qu G, Song D, Yuan Y, Xu Y, Gao K, Luo G, Xiao Z, Lam DSC, Zhong H, Qiao Y, Zhang X. Automatic differentiation of Glaucoma visual field from non-glaucoma visual filed using deep convolutional neural network. BMC Med Imaging. 2018;18(1):35.
45. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. Am J Ophthalmol. 1996;121(5):511–21.
46. Medeiros FA, Jammal AA, Thompson AC. From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. Ophthalmology. 2019;126(4):513–21.
47. Schell GJ, Lavieri MS, Helm JE, Liu X, Musch DC, Van Oyen MP, Stein JD. Using filtered forecasting techniques to determine personalized monitoring schedules for patients with open-angle glaucoma. Ophthalmology. 2014;121(8):1539–46.
48. Garcia GP, Nitta K, Lavieri MS, Andrews C, Liu X, Lobaza E, Van Oyen MP, Sugiyama K, Stein JD. Using kalman filtering to forecast disease trajectory for patients with normal tension glaucoma. Am J Ophthalmol. 2019;199:111–9.

49. Thakur A, Goldbaum M, Yousefi S. Predicting glaucoma before onset using deep learning. Ophthalmol Glaucoma. 2020;3(4):262–8.
50. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology. 2018;125(8):1199–206.
51. Healgoo Ltd. LabelMe. [Online]. Available: http://www.labelme.org. Accessed: 16 Feb 2020.
52. Asaoka R, Murata H, Hirasawa K, Fujino Y, Matsuura M, Miki A, Kanamoto T, Ikeda Y, Mori K, Iwase A, Shoji N, Inoue K, Yamagami J, Araie M. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. Am J Ophthalmol. 2019;198:136–45.
53. Quigley HA, West SK, Rodriguez J, Munoz B, Klein R, Snyder R. The prevalence of glaucoma in a population-based study of Hispanic subjects: proyecto VER. Arch Ophthalmol. 2001;119(12):1819–26.
54. Weinreb RN, Garway-Heath DF, Leung C, Medeiros FA, Liebmann J, editors. 10th consensus meeting: diagnosis of primary open angle glaucoma. Kugler Publications; 2016.
55. You Y, Graham SL. Bilateral optic atrophy from a silent occipital lesion. Ophthalmology. 2019;126(7):979.
56. Fenner BJ, Wong RLM, Lam WC, Tan GSW, Cheung GCM. Advances in retinal imaging and applications in diabetic retinopathy screening: a review. Ophthalmol Ther. 2018;7(2):333–46.
57. Liu S,Yuyi Y, Lelin Z. Sydney genesis program enlight. AI System. sydney morning herald, 2019. [Online]. Available: https://researchers.mq.edu.au/en/clippings/sydney-genesis-program-enlightai-system.
58. Di Ieva A. AI-augmented multidisciplinary teams: hype or hope? Lancet. 2019;394(10211):1801.
59. Coiera E. The fate of medicine in the time of AI. Lancet. 2018;392(10162):2331–2.
60. Khawaja AP, Cooke Bailey JN, Wareham NJ, et al. Genome-wide analyses identify 68 new loci associated with intraocular pressure and improve risk prediction for primary open-angle glaucoma. Nat Genet. 2018;50(6):778–82.
61. Margeta MA, Letcher SM, Igo RP, Cooke Bailey JN, Pasquale LR, Haines JL, Butovsky O, Wiggs JL, for the NEIGHBORHOOD Consortium. Association of APOE with primary open-angle glaucoma suggests a protective effect for APOE ε4. Invest Ophthalmol Vis Sci. 2020;61(8):3.

7 Artificial Intelligence in Retinal Diseases

Aman Kumar, Nitin Kumar Menia, and Aniruddha Agarwal

7.1 Introduction

Artificial intelligence (AI) enables computers to function independently and intelligently to perform tasks usually done by humans. It involves machine learning incorporating various algorithms which are known as neural networks that allow these computers to learn and edit from the provided data sets and subsequently make accurate future predictions. AI in ophthalmology has focused most in the field of retina since it involves large amounts of images which are capable of diagnosing a condition [1, 2].

Retinal diseases are often investigated using multimodal imaging including fundus photography, retinal angiography and optical coherence tomography (OCT). In addition, retinal diseases often have similar, overlapping phenotypes making pattern recognition an important aspect of diagnosis and management. Retinal vascular diseases such as diabetic retinopathy (DR) and retinal vein occlusions may have features such as microaneurysms, and complications such as macular oedema, detection of which has an important role in the screening and early treatment of these patients.

An enormous interest has been generated among the medical fraternity due to the introduction and application of deep learning (DL) and AI in medical services. DL is a new addition to the AI learning technique [3, 4]. DL, as has been previously told in the book, is based on learning from large volumes of data, its analyses, processing of the data and extracting meaningful patterns from the analyses [5, 6]. The algorithm most commonly used in DL is convoluted neural networks (CNNs) [6]. Repetition of tasks and self-learning forms the core of DL, using the CNN. DL utilizing CNN is most suitable for analysis of image-based data [7].

A. Kumar · N. K. Menia · A. Agarwal (✉)
Advanced Eye Center, Post Graduate Institute of Medical Education and Research, Chandigarh, India

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_7

AI is especially useful in the medical sub-specialities in which imaging plays an important role, as in ophthalmology. The basic elements for use of AI in various algorithms analyzing images include enhancement of the images, identification of interest region in the images, descriptor computation and screening classification [8]. Recently, there have been a number of path-breaking advances in the field of AI specifically related to retinal diseases. One such example is the approval of the first AI-enabled screening device for diagnosing an ophthalmic disease, the IDx-DR by the United States Food and Drug Administration (US-FDA) in April 2018 [9]. In this chapter, we have summarized various applications of AI in the field of retinal diseases. The potential limitations of this technology and the future applications of AI have been provided.

7.2 Applications of AI in Retinal Diseases

7.2.1 AI in Diabetic Retinopathy

The basic aim of application of AI in diabetic retinopathy (DR) is to screen for proliferative diabetic retinopathy (PDR) and diabetic macular oedema (DME), the two major causes of severe visual loss in patients with DR. The most important predictor for majority of algorithms is identification of referable diabetic retinopathy (RDR). RDR includes moderate non-proliferative DR (NPDR) or higher and clinically significant macular oedema (CSME). The presence of severe NPDR, proliferative PDR and/or DME constitute sight threatening DR (STDR) [10].

Screening of diabetic patients for DR changes has evolved as an effective method for prevention of blindness. AI-based screening protocols have been validated for screening and found to be quite reliable in differentiating between referable DR and non-referable DR. Gulshan and colleagues from Google AI Healthcare reported a DL system with excellent diagnostic performance. Their system used 128,175 retinal images and was graded for DR and DME by a panel of 54 US-licensed ophthalmologists and ophthalmology residents. A test data set of approximately 10,000 images retrieved from publicly available databases were analyzed by seven certified ophthalmologists. The area under the receiver operating characteristic curve (AUC) was close to 0.991 for both the databases [11].

A study using another DL system achieved an AUC of 0.980, with sensitivity and specificity of 96.8% and 87.0%, respectively, in the detection of referable DR. [12] These studies highlight the potential use of DL in early detection of referable DR. A major study on the validation of DL was performed by Ting et al. [13] in Singapore with multiple retinal images taken with conventional fundus cameras. This study showed a high sensitivity and specificity for identifying DR. The potential challenges and uncertainties include the testing of these DL systems in real-world DR screening programmes and accessing the generalizability of applying these systems to populations of different ethnicities, and using retinal images captured by different fundus cameras.

IDx is the first US-FDA–approved AI device used for screening in DR in 2018. Topcon NW400 fundus camera is coupled to this device, and the camera is used to upload images on the software. The software was programmed in such a way, if the DR is more than mild, then the case is referred to an ophthalmologist. If it was not 'more than mild DR' then the software calls the patient for rescreening after 12 months. Sensitivity and specificity of 87.3% and 89.5% was detected in a multicentre trial in which 900 adults with diabetes were registered. As the device can take a screening decision, it can be used by non-ophthalmologists.

EyeArt™ by Eyenuk was used EyePACS tele-screening system to train the AI algorithms for screening DR and showed a sensitivity and specificity of 90% and 63.2%, respectively. It also detected microaneurysms with a sensitivity of 100%. The system used 40,542 images which were taken from 5084 patients. Another study by Tufail et al. showed the sensitivity of EyeArt was 94.7% for any DR, 93.8% for referable retinopathy and 99.6% for PDR. It also evaluated the results of Retmarker showing sensitivities of 73.0% for any retinopathy, 85.0% for RDR and 97.9% for PDR. Google Health has also reported creating a dataset of 128,000 images which were incorporated by scientists in order to train a DL network for diabetic retinopathy.

OCT angiography is a new technology that has immense applications in the field of DR and DME. A number of research manuscripts have focused on the use of OCT angiography in determining quantitative parameters such as foveal avascular zone (FAZ) areas, retinal microangiopathy changes (such as capillary tortuosity and dropouts) and retinal vascular density indices. The use of AI to OCT angiography images is in its infancy. There are only few published reports that describe the application of DL algorithms in determining the vascular changes on OCT angiography. Guo et al. have proposed a DL algorithm that can automatically segment and quantify the capillary density of the superficial FAZ. The correlation coefficient between the area calculated by the DL algorithm and that calculated by manual segmentation was 0.997 [14].

Heisler et al. used ensemble learning techniques along with DL in classifying DR on OCT angiography. The authors analyzed 380 eyes and concluded that ensemble learning increases the predictive accuracy of CNNs for classifying RDR on OCT angiography [15]. Lo et al. have also employed OCT angiography images in assessing the superficial and deep retinal capillary plexus using CNNs. The algorithm provided an accurate retinal microvasculature assessment using CNNs [16]. Thus, OCT angiography is a useful tool and its segmentation using CNNs is a promising area of research.

7.2.2 AI in Age-Related Macular Degeneration

Age-related macular degeneration (AMD) is among the most common cause of vision impairment specially in elderly patients. The American Academy of Ophthalmology (AAO) recommends subjects with intermediate and late AMD to undergo at least two yearly follow up. With increasing ageing population, it is needed that we have a system to efficiently screen and follow-up these patients [18]. Ting et al. published an accepted DL system to detect and refer patients with AMD. In their study, fovea-centred images without macular segmentation were used [13].

DL systems have been developed using the AREDS data set with a high number of referable AMD (intermediate AMD or worse). Using a fivefold cross-validation, Burlina et al. [19] reported a diagnostic accuracy of between 88.4% and 91.6%, with an AUC of between 0.94 and 0.96. Unlike Ting et al. [13], the authors pre-segmented the macula region prior to training and testing, with an 80/20 split between the training and testing in each fold [19]. In terms of the DL architecture, both AlexNet® and OverFeat® have been used (both of these are CNNs), with AlexNet yielding a better performance. Grassmann et al. [20] reposted a sensitivity of 84.2% in detecting any stage of AMD. Here, six CNNs were used to train different models. The authors concluded that their algorithm was suitable to classify AMD fundus images in other datasets using individuals >55 years of age.

Since the ultimate goal in the application of AI in AMD is detection of neovascular AMD, and determination of cases that need treatment with intravitreal anti-vascular endothelial growth factor (anti-VEGF) agents, there have been attempts made to develop models that predict requirement of anti-VEGF in AMD by DL techniques. Bogunovic et al. [21] have evaluated a machine learning algorithm to predict anti-VEGF treatment needs from OCT scans taken during treatment initiation. However, this is a pilot study and needs further research.

7.2.3 AI in Choroidal Neovascularization and Macular Diseases

Optical coherence tomography has transformed the diagnosis and management of retinal/macular diseases. OCT provides a microscopic view of retina comparable to its histological structure. As far as DL is considered, OCTs, especially macular ones, are suitable for the development of a learning algorithm. The special attributes of this tool are an increased number of OCTs performed around the world, widespread availability, non-invasive nature and a three-dimensional structural information captured with the macular OCTs [1]. The large quantity of macular OCTs can provide a very large database set which can be used to train the DL systems. Compared to colour fundus photographs, the fixation is consistent among the acquired serial automated OCT scans. The accuracy in the acquisition of the OCT scans, however, can lower the complexity of the data and allows deep learning systems to extract meaningful data from a smaller data set [1]. As the macular OCTs provide ultrastructural details of retinal layers, DL can be also used to identify novel biomarkers for the macular diseases.

The first application of DL in interpreting macular OCTs was to automatically classify AMD. Lee et al. used more than 100,000 OCT images to train a DL system to classify AMD, with AUC of 0.97 [22]. Most of the initial studies highlighting the use of macular OCTs in DL have utilized the single OCT B-scans rather than three-dimensional volume-based scans which is a potential barrier to the applicability of DL in OCT-based algorithms [1]. DL using CNNs has been used to successfully segment the retinal anatomical boundaries, intra-retinal fluid cysts and subretinal fluid on OCT B-scans as compared to traditional methods of segmentation [23, 24].

A novel AI framework was used by De Fauw et al. [25] to segment as well as classify OCT images. First, a segmentation network delineated almost 15 retinal morphological features and OCT acquisition artefacts. A classification network then classified the output of the segmentation network into ten different OCT pathologies (including choroidal neovascular membrane; CNV, macular oedema, drusen, geographic atrophy, full-thickness macular hole, partial thickness macular hole, epiretinal membrane, vitreomacular traction, central serous retinopathy and 'normal'). These ten different OCT pathologies were categorized into urgent, semi-urgent, routine and observation. They concluded that the system classified the OCTs at par with the experts [25]. The DL system can be implemented for setting up rapid access 'virtual clinics' and help in triaging patients with macular diseases strengthening the referral system and thus decreasing the undue load on the tertiary-level health institutes [26]. These triaging systems can be used in optometrists in rural settings and help in referrals [27, 28].

7.2.4 Retinopathy of Prematurity

Retinopathy of prematurity (ROP) is an important cause of childhood blindness with increasing annual incidence to 32,000 worldwide [29]. The risk factors being increasing preterm births due to better neonatal intensive care unit (NICU) managements, low birth weights, oxygen supplementation to neonates, among others. A simple but timely indirect ophthalmoscopic examination or digital fundus photo can detect the stage and progression of the disease and prevent blindness. The major barriers are that the diagnosis of ROP is subjective being examiner dependent, and there are very few trained to screen worldwide [30]. Recently, Brown et al. [17] reported the results of a fully automated DL system that could diagnose plus disease, the most important feature of severe ROP, with an AUC of 0.98 compared with a consensus reference standard diagnosis combining image-based diagnosis and ophthalmoscopy. This was compared with ROP experts worldwide, and it was found that six out of eight experts agreed with the i-ROP DL system diagnosis. It showed promising results in disease progression, regression and the response to treatment.

Currently, the DL algorithms focusing on ROP are being designed to distinguish plus disease from non-plus disease. However, more intuitive systems that will enable detection of different grades of ROP, and other conditions such as aggressive posterior ROP, are yet to be developed.

7.2.5 Retinal Vein Occlusions

After DR, retinal vein occlusion is the commonest cause of vision impairment among retinal vascular causes [31]. It leads to retinal haemorrhages, macular oedema and exudation. This condition is more common in the older age group with hypertension, atherosclerosis and cardiac disease being major risk factors

[32]. At present, the use of AI has not been widely explored for retinal vein occlusions. One group has reported that the use of CNN combined with patch-based and image-based vote methods to recognize the fundus image of branch retinal vein occlusion automatically. The authors have reported a high accuracy of over 97% [1].

Nagasato et al. [33] have applied DL algorithms in detecting retinal non-perfusion areas in retinal vein occlusions by using a novel imaging modality, OCT angiography. OCT angiography has revolutionized the field of non-invasive retinal angiography, and application of DL algorithms to this imaging modality certainly opens up newer avenues of research and clinical applicability. In their study, the authors generated heat maps to detect areas of retinal non-perfusion, and the AUC in distinguishing retinal vein occlusion OCT angiography images from normal control subjects was encouraging (0.986). The sensitivity, specificity and average required time for distinguishing the images were 93.7%, 97.3% and 176.9 s, respectively, and the DL algorithm outperformed ophthalmologists in all parameters [33].

Table 7.1 reviews the performance of artificial intelligence algorithms to detect retinal diseases using fundus images.

Table 7.1 Review of performance of artificial intelligence algorithms to detect retinal diseases using fundus images

Author	Study type	AI algorithm/fundus camera	Dataset	Sensitivity (%)	Specificity (%)
Abràmoff et al. [12] (2016)	Retrospective (DR)	Topcon TRC NW6 nonmydriatic fundus camera/ IDx-DR X2	MESSIDOR-2	96.8	87
Gulshan et al. [11] (2016)	Retrospective (DR)	Topcon TRC NW6 nonmydriatic camera/ inception-V3	MESSIDOR-2	87	98.50
Ting et al. (2017) [13]	Retrospective (DR)	FundusVue, canon, Topcon and Carl Zeiss/VCG-19	SiDRP 14–15	90.5	91.6
			Guangdong	98.7	81.6
			SIMES	97.1	82.0
			SINDI	99.3	73.3
			SCES	100	76.3
			BES	94.4	88.5
			AFEDS	98.8	86.5
			RVEEH	98.9	92.2
			Mexican	91.8	84.8
			CUHK	99.3	83.1
Brown et al. [17] (2018)	Retrospective (ROP)	Inception-V1 and U-net	AREDS	100	94

7.3 Pearls and Pitfalls in the Applications of AI

AI is a potential game-changer in the medical field, including ophthalmology. AI is considered one of the most significant of information technology revolutions of recent past [34]. It is a complex field of research, application in medical field is a challenge as it involves two inherently different branches, medical and computer sciences. Challenges exist for AI in ophthalmology practices as the exact number of training images in the dataset and validation set is not standardized [35]. Most datasets use a huge number of images, as more numbers are considered better. Excessive datasets, datasets with homogenous populations and machines of different brands may alter the accuracy assessments. However, including wide demographic profiles, reducing dataset size and algorithm complexity and restricting the number of classifications within a program can improve accuracy and have significant prognostic relevance [35].

As ophthalmological practices are mostly outpatient-based services and most of the sub-specialities in ophthalmology use imaging modalities, AI can be implied to improve patient care. AI is a major boon to the health sector. However, the cost of the machines may be high and the medical companies manufacturing these machines may have conflicts of interest. It is of utmost importance to look into this potential issue and make sure that the technology is available to all including the populations with poor affordability.

AI is not a fail-proof technology, and there could be patients with severe disease whose detection could be missed out. A false-negative report may have a serious bearing on the visual function in these patients. The algorithms having great accuracies do have relatively high false-negative rates in disease detection which could lead to decreased diagnosis. Certain disease manifestations of diabetes, such as featureless retina, associated glaucoma or macular degeneration, may be missed. This is because a current limitation of AI is that separate individual programs need to be created for each individual task, referred to as weak or narrow AI. Thus, until these issues are sorted, the gold standard still remains clinical examination. Although AI has a great promise, it comes with its sets of limitations and risks. There is a risk of reducing the skill of the human workforce to the point that clinicians lose their diagnostic abilities.

In future, AI will be incorporated in the computer-based diagnostic and management tools. This will be particularly useful in rural, underprivileged populations, who have limited access to the health care. Moreover, AI-associated systems have been looked upon as an important way to reduce the social inequalities in the health sector.

7.4 Conclusions

AI and machine learning are enabling enhanced screening and prognostication in retinal diseases specially DR and ROP. This advancement has the potential to increase the patient access to clinical care and reduce health care costs. In a developing country such as India, this is a promising tool which can benefit millions of patients who may not have direct access to ophthalmologists and retina specialists. Future research is surely required for the clinical implementation and cost effectiveness, but DL is likely to impact the practice of medicine and in particular ophthalmology in the coming times.

References

1. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019;103(2):167–75.
2. Padhy SK, Takkar B, Chawla R, Kumar A. Artificial intelligence in diabetic retinopathy: a natural step to the future. Indian J Ophthalmol. 2019;67(7):1004–9.
3. Yadav D, Garg RK, Chhabra D, Yadav R, Kumar A, Shukla P. Smart diagnostics devices through artificial intelligence and mechanobiological approaches. 3 Biotech. 2020;10(8):351.
4. Angehrn Z, Haldna L, Zandvliet AS, et al. Artificial intelligence and machine learning applied at the point of care. Front Pharmacol. 2020;11:759.
5. Yu B, Wang Y, Wang L, Shen D, Zhou L. Medical image synthesis via deep learning. Adv Exp Med Biol. 2020;1213:23–44.
6. Ahmad BU, Kim JE, Rahimy E. Fundamentals of artificial intelligence for ophthalmologists. Curr Opin Ophthalmol. 2020;31(5):303–11.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
8. Raman R, Srinivasan S, Virmani S, Sivaprasad S, Rao C, Rajalakshmi R. Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. Eye (Lond). 2019;33(1):97–109.
9. Armstrong GW, Lorch AC. A(eye): a review of current applications of artificial intelligence and machine learning in ophthalmology. Int Ophthalmol Clin. 2020;60(1):57–71.
10. Wilkinson CP, Ferris FL, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology. 2003;110(9):1677–82.
11. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10.
12. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci. 2016;57(13):5200–6.
13. Ting DSW, Cheung CY-L, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 2017;318(22):2211–23.
14. Guo M, Zhao M, Cheong AMY, Dai H, Lam AKC, Zhou Y. Automatic quantification of superficial foveal avascular zone in optical coherence tomography angiography implemented with deep learning. Vis Comput Ind Biomed Art. 2019;2(1):21.
15. Heisler M, Karst S, Lo J, et al. Ensemble deep learning for diabetic retinopathy detection using optical coherence tomography angiography. Transl Vis Sci Technol. 2020;9(2):20.
16. Lo J, Heisler M, Vanzan V, et al. Microvasculature segmentation and intercapillary area quantification of the deep vascular complex using transfer learning. Transl Vis Sci Technol. 2020;9(2):38.
17. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol. 2018;136(7):803–10.
18. Chew EY, Clemons TE, Agrón E, et al. Effect of omega-3 fatty acids, lutein/zeaxanthin, or other nutrient supplementation on cognitive function: the areds2 randomized clinical trial. JAMA. 2015;314(8):791–801.

19. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA Ophthalmol. 2017;135(11):1170–6.
20. Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology. 2018;125(9):1410–20.
21. Bogunovic H, Waldstein SM, Schlegl T, et al. Prediction of anti-VEGF treatment requirements in neovascular AMD using a machine learning approach. Invest Ophthalmol Vis Sci. 2017;58(7):3240–8.
22. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. Ophthalmol Retina. 2017;1(4):322–7.
23. Venhuizen FG, van Ginneken B, Liefers B, et al. Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography. Biomed Opt Express. 2018;9(4):1545–69.
24. Hamwood J, Alonso-Caneiro D, Read SA, Vincent SJ, Collins MJ. Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers. Biomed Opt Express. 2018;9(7):3049–66.
25. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342–50.
26. Buchan JC, Amoaku W, Barnes B, et al. How to defuse a demographic time bomb: the way forward? Eye (Lond). 2017;31(11):1519–22.
27. McLaughlin CR, Biehl M, Chan BJ, et al. Ophthalmology referrals from optometry: a comparative study (the R.O.C.S study). Can J Ophthalmol. 2018;53(5):491–6.
28. Kaplan RI, Chen M, Gupta M, Rosen RB. Impact of automated OCT in a high-volume eye urgent care setting. BMJ Open Ophthalmol. 2019;4(1):e000187.
29. Blencowe H, Moxon S, Gilbert C. Update on blindness due to retinopathy of prematurity globally and in India. Indian Pediatr. 2016;53(Suppl 2):S89–92.
30. Fleck BW, Williams C, Juszczak E, et al. An international comparison of retinopathy of prematurity grading performance within the benefits of oxygen saturation targeting II trials. Eye (Lond). 2018;32(1):74–80.
31. Rogers SL, McIntosh RL, Lim L, et al. Natural history of branch retinal vein occlusion: an evidence-based systematic review. Ophthalmology. 2010;117(6):1094–1101.e5.
32. Klein R, Klein BE, Moss SE, Meuer SM. The epidemiology of retinal vein occlusion: the beaver dam eye study. Trans Am Ophthalmol Soc. 2000;98:133–41. discussion 141–143
33. Nagasato D, Tabuchi H, Masumoto H, et al. Automated detection of a nonperfusion area caused by retinal vein occlusion in optical coherence tomography angiography images using deep learning. PLoS One. 2019;14(11):e0223965.
34. Du X-L, Li W-B, Hu B-J. Application of artificial intelligence in ophthalmology. Int J Ophthalmol. 2018;11(9):1555–61.
35. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. Clin Exp Ophthalmol. 2019;47(1):128–39.

8 Artificial Intelligence in Neuro-Ophthalmology

Raymond P. Najjar, Caroline Vasseneix, and Dan Milea

8.1 Introduction

Artificial intelligence (AI) is an entity capable of receiving, interpreting and learning from single or various inputs before displaying a flexible task to achieve a particular task. Within the past decade, AI has provided ophthalmologists with new, fast, accurate and automated means for diagnosing and sometimes treating ocular diseases, opening new avenues for modern eye care [1]. Among AI techniques, machine learning (ML) and deep learning (DL) have been successful in diagnosing several ocular conditions ranging from the anterior to the most posterior segment of the eye. Several subspecialties of ophthalmology have, however, benefitted from AI to a larger extent than others. Among the most common ocular diseases automatically detected by DL algorithms is diabetic retinopathy, glaucoma and age-related macular degeneration [2–6]. Conversely, subspecialties like neuro-ophthalmology have, until recently, been deprived from major advances in the AI-driven detection, let alone treatment, of neuro-ophthalmological conditions.

R. P. Najjar (✉)
Visual Neuroscience Group, Singapore Eye Research Institute, Singapore, Singapore

Duke-NUS Medical School, Singapore, Singapore
e-mail: raymond.najjar@seri.com.sg

C. Vasseneix
Visual Neuroscience Group, Singapore Eye Research Institute, Singapore, Singapore

D. Milea
Visual Neuroscience Group, Singapore Eye Research Institute, Singapore, Singapore

Duke-NUS Medical School, Singapore, Singapore

Singapore National Eye Centre, Singapore, Singapore
e-mail: dan.milea@singhealth.com.sg

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_8

8.2 Neuro-Ophthalmology

The structural arrangement of the visual system extends beyond the eyes to the most posterior segments of the brain (i.e., the occipital cortex). Consequently, patients with intracranial pathologies often complain of visual disruptions and end up consulting an ophthalmologist [7]. Neuro-ophthalmology is a medical sub-specialty, at the intersection of ophthalmology and neurology, dealing with conditions that affect specifically afferent (i.e., vision) and/or efferent (i.e., eye movements, pupillary responses) pathways connecting the eyes with the rest of the nervous system. As such, modern neuro-ophthalmology has become an integrative medical discipline, connecting not only ophthalmologists and neurologists but also neuro-radiologists, neurosurgeons, neuro-otologists, neuro-immunologists, geneticists, and neuropathologists. Such multidisciplinary approaches are becoming the norm for the often-difficult diagnosis and management of neuro-ophthalmic conditions which remain relatively rare and complex, compared other, more common ophthalmic and neurologic diseases. Practicing neuro-ophthalmology requires expertise not only in diagnosing/treating ocular diseases but also in managing conditions affecting the brain, nerve, and muscle.

Very briefly, the most common neuro-ophthalmic conditions can be organized in,

1. conditions affecting the afferent visual system (from the retina to the optic nerves, chiasm, retro-chiasmal pathways, and occipital lobes) and causing higher order visual dysfunctions,
2. conditions affecting the efferent pathways, causing central ocular motor disorders (at cortical, brainstem level), gaze instability, ocular motor cranial neuropathies, and pupillary disorders in addition to more peripheral dysfunctions affecting the neuromuscular junction or the muscles themselves.

   The range of conditions that can specifically affect one or more of these anatomical structures is very large, including autoimmune, inflammatory, ischemic, infectious, compressive, traumatic, congenital, and degenerative diseases. It is not uncommon that an isolated, relatively benign neuro-ophthalmic dysfunction (i.e., inflammatory optic neuropathy) heralds other, more serious neurological diseases (multiple sclerosis, for example). Similarly, acute squint due to acquired ocular misalignment (or swelling of optic nerve heads) can be the first and only manifestation of life-threatening conditions (due to aneurysms, tumors, systemic metabolic diseases, etc), requiring urgent detection and care.

During the recent decades, clinical neuro-ophthalmology has largely benefitted from fundamental scientific discoveries and their applications, performed by neuroscientists, data scientists, biomedical engineers, etc. Interestingly, compared to other ophthalmic disciplines, neuro-ophthalmology has not benefitted, until recently, from significant advances in the area of AI and telemedicine. Some of the main reasons for such a delay are (1) the low prevalence and heterogeneity of neuro-ophthalmological conditions and consequently the scarcity of data necessary to efficiently train an AI algorithm; (2) the neuro-ophthalmologist community is relatively small compared to other ophthalmic specialties; (3) in some neuro-ophthalmological conditions, neurologists provide the final diagnosis which may lead to a loss of follow-up data and reliable ground truth necessary to train AI algorithms; and (4) ground truth can be heterogeneous from center to center, while multiple centers and large sample sizes are often required to efficiently train AI algorithms to detect rare neuro-ophthalmic conditions. Nevertheless, the field is not barren of advancements in the field of AI. In the following paragraphs, we summarize the most prominent investigations utilizing AI to detect prominent neuro-ophthalmic conditions affecting the optic nerve head and eye movements.

8.3 Artificial Intelligence in Optic Nerve Head (Disc) Abnormalities

The optic disc or optic nerve head is the proximal end of the optic nerve. In a clinical setting, the optic nerve head (optic disc) integrity can be evaluated in real time using direct ophthalmoscopy or fundus photography. While some optic nerve lesions cause visible disc changes including swelling, pallor, cupping (e.g., in glaucoma) or infiltration, other lesions that are distal from the disc, such as retrobulbar optic neuritis, are not associated, at the acute stage, with disc abnormalities (only one-third of optic neuritis patients have optic disc edema) [7].

Optic disc abnormalities associated with neuro-ophthalmic conditions are relatively rare compared to optic disc changes seen in glaucoma and affecting 3.5% of individuals aged between 40 and 80, worldwide [8]. Given the high availability of optic disc images in glaucomatous eyes, in addition to the clinical need for a cost-effective screening method for the disease, artificial multiple intelligence methods including DL neural network algorithms, notoriously dependent on large datasets, have attempted to automatically detect glaucoma on digital fundus images [9], based on the optic disc appearance alone, or in combination with optical coherence tomography (OCT) findings [1, 10]. Conversely, a limited number of studies have aimed to automatically determine the optic disc laterality using DL and transfer learning [11], and detect neuro-ophthalmic optic nerve head abnormalities [12].

Compared to sight-threatening ocular conditions such as glaucoma and diabetic retinopathy, some neuro-ophthalmic manifestations, alike papilledema defined as bilateral optic disc edema (swelling) from intracranial hypertension (ICH), can be life-threatening. Whether idiopathic or due to a brain tumor, venous sinus thrombosis or medication, papilledema detection is a medical emergency that requires prompt diagnosis and clinical intervention. Failure to detect optic nerve edema can cause devastating diagnostic errors leading to permanent visual loss, neurologic dysfunction, or even death [13, 14]. Conversely, false diagnosis of optic disc swelling can lead to unnecessary, invasive, and expensive diagnostic investigations including neuroimaging (e.g., CT scans and MRIs) as well as invasive examination of the cerebrospinal fluid [15]. While trained ophthalmologists are capable of identifying most acquired optic disc abnormalities using ophthalmoscopy, non-ophthalmic healthcare providers are less confident in their aptitudes to visualize the appearance of the optic disc using this technique [16]. This is particularly true in emergency departments (ED) where patients with symptoms of ICH are most likely to turn up. Ocular fundus digital cameras providing high-quality photographs of the optic nerve and retina offer a reliable alternative to direct ophthalmoscopy [17]. In a study conducted at an ED, 8.5% of patients presenting with headache, neurologic deficit, visual loss, or elevated diastolic blood pressure had abnormal findings on digital fundus photographs and 2.6% (1 out of 38) had optic nerve head edema [18]. However, fundus photographs required interpretation either by physicians on-site or by ophthalmologists or by other experts through tele-ophthalmology platforms [19, 20].

Studies using fundus images showed that using feature extraction along with ML (e.g., support vector machine, tree-based models) and advanced statistical procedures (e.g., gray-level co-occurrence matrix) can not only discriminate discs with papilledema from normal discs with high accuracy (93%) [21] but can also grade the severity of the condition with substantial agreement with neuro-ophthalmologist (Kappa score = 0.71) [22] and spectral domain optical coherence tomography features such as optic nerve head volume [23]. In a real-life setting, however, the identification of the optic disc appearance yields several diagnostic possibilities (multiclass classification). In a recent study, Ahn et al. used machine learning to discriminate between normal discs, swollen discs due to various optic neuropathies, and pseudopapilledema [24]. Using data augmentation to overcome the risk of overfitting and a classical CNN with TensorFlow and transfer learning, the authors were able to differentiate true optic disc swelling from pseudo-swelling with an accuracy of ~95%. Unfortunately, this study had numerous methodological limitations, including the lack of rigorous clinical inclusion criteria. Nevertheless, these studies paved the way for the utilization of AI for papilledema detection yet had limited sample sizes that were inadequate for DL approaches.

In a recent large collaborative effort, Milea and colleagues managed to train a deep learning system (DLS) on 14,341 photographs collected from 6779 patients from 19 neuro-ophthalmology centers worldwide (BONSAI consortium), including 9156 images of normal optic discs, 2148 of discs with confirmed papilledema, and 3037 of discs with other abnormalities [25]. When evaluated on a separate set of 1505 fundus photographs retrospectively collected from five other centers including a wide range of ethnic groups and fundus cameras, the BONSAI-DLS yielded a high accuracy for the classification of normal discs, discs with papilledema, and discs with other abnormalities (e.g., non-arteritic ischemic optic neuropathy, optic disc atrophy, optic disc drusen) with AUCs of 0.98, 0.96, and 0.90, respectively [25]. Interestingly, the BONSAI-DLS was capable of correcting labeling errors within the reference standard and, in a subsequent study, showed a performance that was at least as good as two expert neuro-ophthalmologists (Classification accuracy of the DLS = 85%; accuracies of the two experts: 80% and 84%) [26]. It is important to mention that the classification by the DLS and experts was, however, purely performed based on fundoscopic images, without consideration of other important clinical features of intracranial hypertension and consequently papilledema such as visual loss, headache, tinnitus. The high classification performance of the BONSAI-DLS is promising for the automated and fast interpretation of optic disc abnormalities in clinical settings where expert opinion is not readily available [27], especially with the emergence of user-friendly handheld fundus cameras. Furthermore, DL in general and these findings in particular provide a unique opportunity for non-ophthalmologic healthcare personnel who are not skilled in performing ophthalmoscopy, to automatically and successfully identify sight- and life-threatening conditions on fundus images. Neurologists and emergency department doctors are most likely to benefit from the integration of automated optic disc appearance classification into their routine evaluations of certain patients, especially with the limited access to neuro-ophthalmologists [27]. A prospective validation of the BONSAI-DLS in real-life settings is, however, essential before recommendations can be made regarding its use in various clinical scenarios.

In addition to optic disc swelling, optic disc atrophy resulting in optic disc pallor on fundoscopy is an important clinical manifestation of retinal nerve fiber degeneration and glial reorganization. Optic disc pallor is often associated with various types and advanced stages of optic neuropathies including, but not limited to compressive, ischemic, inflammatory, or hereditary optic neuropathies. As such, the diagnosis of optic disc pallor and more importantly its underlying medical condition remains challenging. This is in part due to the subjective nature of the optic disc evaluation leading to high interindividual variability, in addition to various masking anatomic differences between patients' optic nerve heads including physiologic temporal pallor [28], pseudophakia, peripapillary atrophy, and tilted discs. Even though intraocular axonal loss can nowadays be quantified by advanced and costly imaging modalities such as OCT, the detection of pallor and underlying conditions using ophthalmoscopy or fundus images remains more cost-effective, especially for primary eye care services. In 2018, Yang and colleagues designed a computer-aided detection system (CAD) to automatically segment the optic disc, enhance fundus images, and extract features and parameters of disc pallor [29]. The parameters used by the authors were (1) brightness correction defined as "the ratio of the mean brightness intensity of the 'cup depth' compared to the 'background region'" and (2) the temporal-to-nasal ratio defined as "the mean brightness intensity of pixels in the temporal region divided by the mean intensity of pixels in the nasal region of the clinically significant neuroretinal rim." A logistic regression model for pallor risk classification integrating the latter parameters yielded high accuracy (i.e., overall accuracy of 96% and sensitivity and specificity >95%) for the automatic detection of optic disc pallor from normal discs on fundus images. A performance that surpassed that of two ophthalmologists who manually graded the images [29].

Optic disc cupping is a feature of glaucomatous optic neuropathy [30], but can be found in other optic neuropathies. Optic nerve head notching is also indicative, yet not specific to glaucoma, being often sectorial and congruent with the corresponding visual field loss. In clinical practice, it may be difficult to differentiate glaucomatous from non-glaucomatous optic disc features, especially when comparing hereditary optic neuropathies and normal tension glaucoma, which share a few of these features. In addition, glaucomatous eyes may also display secondary optic disc pallor in advanced stages [30, 31]. Distinguishing glaucomatous from non-glaucomatous causes of optic neuropathies is important yet elusive if solely based on the morphologic assessment of the disc [32]. Using color fundus photographs, Yang and colleagues recently evaluated the performance of a DL algorithm based upon the convolutional neural network (CNN) of the ResNet-50 architecture for differentiating glaucomatous and other, non-glaucomatous optic neuropathies with disc pallor (including compression of the anterior visual pathway, demyelination, inflammation, ischemic, toxic, and traumatic optic neuropathy). The model was trained on 18,000 images (after a 20-fold data augmentation) including 6000 images of normal discs, and tested on 2675 images including 2503 normal discs. The DL algorithm showed a high overall accuracy (99%) for the classification of glaucomatous optic neuropathies, non-glaucomatous optic neuropathies, and normal discs on fundus images. The model also yielded an area under the precision-recall curve of 0.87, a sensitivity of 93.4%, and a specificity of 81.8% for the differentiation of glaucomatous and non-glaucomatous optic neuropathies [33].

8.4 Artificial Intelligence in Eye Movement Disorders

Eye movement disorders include ocular motor nerve palsies (third, fourth, and sixth cranial nerves) and other causes of diplopia and ocular misalignment, conjugate gaze abnormalities, and nystagmus or other abnormal eye movements. There is a paucity of literature about AI applications in this domain. Some authors used machine learning techniques to evaluate conjugate gaze limitations as ocular biomarkers for neurodegenerative diseases (Parkinson [34, 35], Alzheimer [36], Huntington disease [37]) or even neuropsychiatric diseases [38]. In addition, AI has also been used with neuroimaging modalities for multiple tasks including segmentation, classification, diagnosis, prognosis, prediction of outcome, and risk assessment [39]. We will not discuss the implications of AI in neuroimaging or neurological conditions as these topics are beyond the scope of this chapter and pertain more to neurology than neuro-ophthalmology.

AI techniques were described to model ocular motor data [40], or predict features related to congenital nystagmus [41]. More recent and advanced techniques, using DL, are described for strabismus detection or recognition in pediatric ophthalmology [42–48], which could potentially be used for cranial palsies and telemedicine applications.

8.4.1 Artificial Intelligence and Ocular Motor Features

Few studies describe the use of machine learning techniques to study conjugate gaze abnormalities and nystagmus. Conjugate gaze abnormalities include vertical or horizontal conjugate gaze limitations, pursuit or saccadic deficits, or involuntary conjugate gaze deviations. In one study [40], published in 2001, the authors used a decision tree induction to model the relationships between ocular motor test parameters and lesion sites in a dataset of operated cerebellopontine angle tumor, operated hemangioblastoma, infarction of cerebello-brainstem, Menière's disease (total of 137 patients) and 78 controls. Ocular motor evaluation included pursuit eye movements and saccadic eye movements, and best results were obtained with a combination of the two types of movements. The classification system with three classes (control subject, central lesion, and peripheral lesion) yielded a mean accuracy of 91% and the system with five classes (control subject, brainstem lesion, cerebellar lesion, cerebello-brainstem lesion, and peripheral lesion) yielded a mean accuracy of 88%.

Nystagmus is an involuntary, rhythmical oscillation of the eyes, and can be pendular or have jerk characteristics (slow and fast phases). Nystagmus can be physiologic or pathologic, congenital, or acquired, due to central nervous system dysfunction, peripheral vestibular disease, or visual loss. In order to investigate relationships between different parameters of congenital nystagmus, D'Addio et al. [41], used electrooculography to record eye movements in 20 patients and extracted some parameters from the signals through a custom-made software. Predictive models were built using two algorithms, random forests, and logistic regression tree. The models was capable of predicting visual acuity and variability of eye positioning according to nystagmus features (such as baseline oscillations and nystagmus foveation periods) with a coefficient of determination over 0.72 and over 0.70, respectively [41].

The use of AI for the detection of strabismus is subject to a growing interest. Several clinical applications are described in Reid and Eaton's review [48]. Among these, the detection of strabismus by various AI systems, which could be extrapolated to the detection of cranial nerve palsies and other ocular misalignment. Lu et al.[46] developed a tele-strabismus dataset composed of 5685 facial photographs taken by the patients themselves and used the dataset to train and test a new algorithm in two stages; first a region-based fully convolutional network (R-FCN) performed eye region segmentation, then a deep CNN classified the segmented eye regions as strabismus or normal. The classification performance of Lu and colleagues' model were encouraging with a sensitivity of 93.3%, specificity of 96.2%, and accuracy of 93.9%. Chen et al. [45] used in-office application of detection of strabismus by a CNN previously trained on ImageNet, which was able to detect fixation deviations from eye tracking data in nine directions of gaze. Experiments were performed on 17 strabismus patients and 25 controls. The best performance achieved was 95% accuracy, 94% sensitivity, and 96% specificity. Gramatikov et al. [47] used artificial neural networks (ANN) to detect central fixation determined by retinal birefringence scanning (RBS), a technique that uses changes in polarization of light returning from the eye. When applied to both eyes simultaneously, this technique allows to measure the alignment of the eyes. The classification performance of the ANN in 39 subjects (19 strabismus and 20 controls) yielded 98.5% sensitivity and 100% specificity. Van Eenwyk et al. [42] used a photorefractive video-based system combining Bruckner pupil red reflex imaging (i.e., a deviated eye has a brighter reflex) and eccentric photorefraction to screen 610 children aged 6 months to 6 years old for amblyogenic factors (including strabismus) and need for ophthalmic referral "refer/

do not refer." Four AI techniques were used, case-based reasoning (template matching, with comparison of videos of known conditions with new cases), case-based fuzzy logic, ANNs (connected layers of neurons, with input, hidden and output layers), and decision tree (program that classify attributes of a case in a specific order). The best classification performance ("refer/do not refer") was obtained using the decision tree algorithm with an accuracy of 77%, compared to the "gold standard" specialist examination. Finally, Fisher and Chandna [43, 44] studied the diagnosis and patterns of deviation in vertical strabismus using an expert system (StrabNet) approach employing ANNs. Data included measurements of vertical deviations by prism cover test by ophthalmologists and orthoptists. Four hypertrophic and four hypotrophic patterns of deviations associated with a diagnosis were chosen based on clinical experience, and the authors used 50 sets of measurement for each of the eight diagnoses for training and evaluation of ANN. The performance of StrabNet was good, with an accuracy of 94% and specificity of 100%, and 84% match with an expert orthoptist. Finally, in pediatric ophthalmology, as in other medical fields, Reid and Eaton [48] emphasize some limitations or hurdles facing AI systems, which are (1) the disagreement on reference standard among experts, (2) the poor reproducibility and comparability if authors do not use publicly available datasets, (3) the lack of temporal evaluation, and (4) uninterpretable black box models in DL systems which render them as difficult to trust for some healthcare providers.

8.5 Conclusion

AI systems have proven to be useful for the screening and characterization of optic disc health and to a lesser extent eye movement disorders, as well as early detection and prediction of stage disease in patients with certain neurological and neuro-ophthalmologic conditions. Such systems provide a unique opportunity to make complex diagnostic procedures affordable, accessible, objective, and accurate. Furthermore, in urgent settings, AI has the potential to improve the care for patients who need neuro-ophthalmic assessment when experts are not readily available. However, more studies are needed to evaluate the real-life clinical application of AI systems in neuro-ophthalmic conditions, as already done for other fields of medicine and ophthalmology.

The acceptance of telemedicine and AI-based medical interventions by providers, patients, and regulators was accelerated by the COVID-19 pandemic [49, 50]. Nevertheless, the surge in AI applications requires complementation with digital innovations allowing for long-distance clinical investigations/self-investigations (e.g., digital fundus photography, vision testing applications) [51] to promote the recognition of tele-neuro-ophthalmology as a viable healthcare delivery system.

Acknowledgements Supported by the Singapore National Medical Research Council (Clinician Scientist Individual Research grant CIRG18Nov-0013), and the Duke-NUS Medical School, Ophthalmology and Visual Sciences Academic Clinical Program grant (05/FY2019/P2/06-A60).

References

1. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Tan GSW, Schmetterer L, Keane PA, Wong TY. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019;103:167–75.
2. Bellemo V, Lim G, Rim TH, et al. Artificial intelligence screening for diabetic retinopathy: the real-world emerging application. Curr Diab Rep. 2019;19:72.
3. Ting DSW, Cheung CY-L, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 2017;318:2211–23.
4. Medeiros FA. Deep learning in glaucoma: progress, but still lots to do. Lancet Digit Health. 2019;1:e151–2.
5. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. Ophthalmol Retina. 2017;1:322–7.
6. Yan Q, Weeks DE, Xin H, Swaroop A, Chew EY, Huang H, Ding Y, Chen W. Deep-learning-based prediction of late age-related macular degeneration progression. Nat Mach Intellig. 2020;2:141–50.
7. Martin TJ. Chapter 12 - neuro-ophthalmology. In: Palay DA, Krachmer JH, editors. Primary care ophthalmology. 2nd ed. Philadelphia: Mosby; 2005. p. 199–227.
8. Tham Y-C, Li X, Wong TY, Quigley HA, Aung T, Cheng C-Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology. 2014;121:2081–90.
9. Devalla SK, Liang Z, Pham TH, Boote C, Strouthidis NG, Thiery AH, Girard MJA. Glaucoma management in the era of artificial intelligence. Br J Ophthalmol. 2020;104:301–11.
10. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. Prog Retin Eye Res. 2019;72:100759.
11. Liu TYA, Ting DSW, Yi PH, Wei J, Zhu H, Subramanian PS, Li T, Hui FK, Hager GD, Miller NR. Deep learning and transfer learning for optic disc laterality detection: implications for machine learning in neuro-ophthalmology. J Neuroophthalmol. 2020;40:178–84.
12. Milea D, Singhal S, Najjar RP. Artificial intelligence for detection of optic disc abnormalities. Curr Opin Neurol. 2020;33:106–10.
13. Woodward EG. Clinical negligence. Ophthalmic Physiol Opt. 2006;26:215–6.
14. Rawlinson K. Optometrist wins appeal against conviction for manslaughter of boy, 8. The Guardian. https://www.theguardian.com/uk-news/2017/jul/31/optometrist-honey-rose-wins-appeal-against-conviction-manslaughter-boy-8; Accessed 1 Nov 2020.
15. Poostchi A, Awad M, Wilde C, Dineen RA, Gruener AM. Spike in neuroimaging requests following the conviction of the optometrist Honey Rose. Eye. 2018;32:489–90.
16. Biousse V, Bruce BB, Newman NJ. Ophthalmoscopy in the 21st century. Neurology. 2018;90:167–75.
17. Bruce BB, Lamirel C, Wright DW, Ward A, Heilpern KL, Biousse V, Newman NJ. Nonmydriatic ocular fundus photography in the emergency department. N Engl J Med. 2011;364:387–9.
18. Sachdeva V, Vasseneix C, Hage R, Bidot S, Clough LC, Wright DW, Newman NJ, Biousse V, Bruce BB. Optic nerve head edema among patients presenting to the emergency department. Neurology. 2018;90:e373–9.
19. Bruce BB, Thulasi P, Fraser CL, Keadey MT, Ward A, Heilpern KL, Wright DW, Newman NJ, Biousse V. Diagnostic accuracy and use of nonmydriatic ocular fundus photography by emergency physicians: phase II of the FOTO-ED study. Ann Emerg Med. 2013;62:28–33.e1.
20. Irani NK, Bidot S, Peragallo JH, Esper GJ, Newman NJ, Biousse V. Feasibility of a nonmydriatic ocular fundus camera in an outpatient neurology clinic. Neurologist. 2020;25:19–23.
21. Akbar S, Akram MU, Sharif M, Tariq A, Yasin UU. Decision support system for detection of papilledema through fundus retinal images. J Med Syst. 2017;41:66.
22. Echegaray S, Zamora G, Yu H, Luo W, Soliz P, Kardon R. Automated analysis of optic nerve images for detection and staging of papilledema. Invest Ophthalmol Vis Sci. 2011;52:7470–8.

23. Agne J, Wang J-K, Kardon RH, Garvin MK. Determining degree of optic nerve edema from color fundus photography. In: Medical imaging 2015: computer-aided diagnosis. International Society for Optics and Photonics. p 94140F.
24. Ahn JM, Kim S, Ahn K-S, Cho S-H, Kim US. Accuracy of machine learning for differentiation between optic neuropathies and pseudopapilledema. BMC Ophthalmol. 2019;19:1–7.
25. Milea D, Najjar RP, Jiang Z, et al. Artificial intelligence to detect papilledema from ocular fundus photographs. N Engl J Med. 2020;382:1687–95.
26. Biousse V, Newman NJ, Najjar RP, et al. Optic disc classification by deep learning versus expert neuro-ophthalmologists. Ann Neurol. 2020;88:785–95.
27. Stunkel L, Mackay DD, Bruce BB, Newman NJ, Biousse V. Referral patterns in neuro-ophthalmology. J Neuroophthalmol. 2020;40(4):485–93.
28. Osaguona VB. Differential diagnoses of the pale/white/atrophic disc. Commun Eye Health. 2016;29:71–4.
29. Yang HK, Oh JE, Han SB, Kim KG, Hwang J-M. Automatic computer-aided analysis of optic disc pallor in fundus photographs. Acta Ophthalmol. 2019;97:e519–25.
30. Ramm L, Schwab B, Stodtmeister R, Hammer M, Sauer L, Spörl E, Pillunat LE, Terai N. Assessment of optic nerve head pallor in primary open-angle glaucoma patients and healthy subjects. Curr Eye Res. 2017;42:1313–8.
31. Hata M, Miyamoto K, Oishi A, Makiyama Y, Gotoh N, Kimura Y, Akagi T, Yoshimura N. Comparison of optic disc morphology of optic nerve atrophy between compressive optic neuropathy and glaucomatous optic neuropathy. PLoS One. 2014;9:e112403.
32. O'Neill EC, Danesh-Meyer HV, Kong GXY, Hewitt AW, Coote MA, Mackey DA, Crowston JG, Optic Nerve Study Group. Optic disc evaluation in optic neuropathies: the optic disc assessment project. Ophthalmology. 2011;118:964–70.
33. Yang HK, Kim YJ, Sung JY, Kim DH, Kim KG, Hwang J-M. Efficacy for differentiating nonglaucomatous versus glaucomatous optic neuropathy using deep learning systems. Am J Ophthalmol. 2020;216:140–6.
34. Prashanth R, Dutta Roy S, Mandal PK, Ghosh S. High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. Int J Med Inform. 2016;90:13–21.
35. Przybyszewski A, Kon M, Szlufik S, Szymanski A, Habela P, Koziorowski D. Multimodal learning and intelligent prediction of symptom development in individual Parkinson's patients. Sensors. 2016;16:1498.
36. Nam U, Lee K, Ko H, Lee J-Y, Lee EC. Analyzing facial and eye movements to screen for Alzheimer's disease. Sensors. 2020;20:5349.
37. Miranda Â, Lavrador R, Júlio F, Januário C, Castelo-Branco M, Caetano G. Classification of Huntington's disease stage with support vector machines: a study on oculomotor performance. Behav Res. 2016;48:1667–77.
38. Mao Y, He Y, Liu L, Chen X. Disease classification based on eye movement features with decision tree and random forest. Front Neurosci. 2020;14:798.
39. Zhu G, Jiang B, Tong L, Xie Y, Zaharchuk G, Wintermark M. Applications of deep learning to neuro-imaging techniques. Front Neurol. 2019;10:869.
40. Viikki K, Isotalo E, Juhola M, Pyykkö I. Using decision tree induction to model oculomotor data. Scand Audiol. 2001;30:103–5.
41. D'Addio G, Ricciardi C, Improta G, Bifulco P, Cesarelli M. Feasibility of machine learning in predicting features related to congenital nystagmus. In: Henriques J, Neves N, de Carvalho P, editors. XV Mediterranean conference on medical and biological engineering and computing – MEDICON 2019. Cham: Springer; 2020. p. 907–13.
42. Van Eenwyk J, Agah A, Giangiacomo J, Cibis G. Artificial intelligence techniques for automatic screening of amblyogenic factors. Trans Am Ophthalmol Soc. 2008;106:64–73. discussion 73–74
43. Fisher AC, Chandna A, Cunningham IP. The differential diagnosis of vertical strabismus from prism cover test data using an artificially intelligent expert system. Med Biol Eng Comput. 2007;45:689–93.

44. Chandna A, Fisher AC, Cunningham I, Stone D, Mitchell M. Pattern recognition of vertical strabismus using an artificial neural network (StrabNet \textsuperscript©). Strabismus. 2009;17:131–8.
45. Chen Z, Fu H, Lo W-L, Chi Z. Strabismus recognition using eye-tracking data and convolutional neural networks. J Healthcare Eng. 2018;2018:e7692198.
46. Lu J, Fan Z, Zheng C, Feng J, Huang L, Li W, et al. Automated strabismus detection for telemedicine applications. 2018. arXiv:1809.02940.
47. Gramatikov BI. Detecting central fixation by means of artificial neural networks in a pediatric vision screener using retinal birefringence scanning. Biomed Eng Online. 2017;16:52.
48. Reid JE, Eaton E. Artificial intelligence for pediatric ophthalmology. Curr Opin Ophthalmol. 2019;30:337–46.
49. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill. 2020;6:e18810.
50. Bloem BR, Dorsey ER, Okun MS. The coronavirus disease 2019 crisis as catalyst for telemedicine for chronic neurological disorders. JAMA Neurol. 2020;77:927–8.
51. Ko MW, Busis NA. Tele-neuro-ophthalmology: vision for 20/20 and beyond. J Neuroophthalmol. 2020;40(3):378–84.

9 Artificial Intelligence and Other Applications in Ophthalmology and Beyond

Stephanie Wangyu (Chiang) and Lama A. Al-Aswad

9.1 Introduction

Ophthalmology lends itself perfectly to artificial intelligence applications due to its high dependency on imaging. Virtually no ophthalmic field is left untouched by artificial intelligence (AI), as already discussed in the prior chapters on the anterior segment, glaucoma, and retinal diseases.

Machine learning (ML), a subset of AI, has been frequently explored for clinical predictive applications in ophthalmology. ML has the ability to learn from "inputs," often images, without being specifically coded. However, many of these approaches require significant preprocessing to manually label features or classify inputs. More recently, deep learning (DL) methods, such as convolutional neural networks (CNNs), have been developed to yield impressive outcomes. In this approach, DL algorithms are able to bypass manual preprocessing and learn in an unsupervised approach, enabling faster, more efficient, and, oftentimes, more accurate analyses. Thus, many of the emerging AI techniques developed for ophthalmic use are DL-based.

Here, we discuss the role of AI in revolutionizing ocular pathology, oncology, genetics, and pediatric ophthalmology. Furthermore, AI platforms have expanded and elucidated prior hypotheses that the eye is truly the window to the rest of the body. The retina holds predictive power in evaluating cardiovascular health, anemia, multiple sclerosis, and other neurodegenerative disorders such as Alzheimer's and Parkinson's disease. Finally, we will explore AI in tele-ophthalmology, the next frontier of eye care delivery. While AI development in these areas is still in its nascence, its transformative power cannot be understated.

S. Wangyu (Chiang) · L. A. Al-Aswad (✉)
Department of Ophthalmology, NYU Langone Health, NYU Grossman School of Medicine, New York, NY, USA
e-mail: Lama.Al-Aswad@nyulangone.org

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_9

9.2 AI in Pathology

9.2.1 Precedents

Digitized pathology slides provide a robust data source for DL algorithmic training and use. Already, CNNs have demonstrated capability of diagnosing breast cancer nodal metastasis better than a panel of pathologists when compared to an independent reference standard [1, 2]. A deep learning algorithm classified whole-slide images with an AUC of 0.996, significantly better than pathologist performance with time constraints (AUC 0.810). Notably, the algorithms were able to better detect cases with only micrometastases with an AUC of 0.885, whereas pathologists achieved an AUC of 0.808, with even the best-performing pathologist on the panel missing 37% of those cases [2]. Additionally, automated prostate adenocarcinoma grading based on H&E-stained slides demonstrate a 75% match rate between DL and pathologists [1].

AI is being explored as a potential method to derive genotype and phenotype details based on pathology slides. As ocular pathology slides are increasingly digitized, DL algorithms can help automate diagnostics in a time efficient and effective manner, as well provide additional information on disease characteristics. Coudray and his team trained a deep CNN to accurately classify frozen lung tissue samples into normal, adenocarcinoma, and squamous cell carcinoma in a fraction of the time it took pathologists [3]. Their model achieved comparable sensitivity and specificity to those of pathologists, 89% and 93%, respectively. Around 50% of the images misclassified by the algorithm were also misclassified by pathologists, but conversely, over 80% of the images misclassified by at least one pathologist were correctly classified by the algorithm, suggesting a supplementary role for the network [3]. Additionally, their model was able to predict six commonly mutated genes in lung adenocarcinoma from pathology images. They also tested their neural network model on independent cohorts, reflecting the generalizability of their algorithm to any cancer type. The model is now openly available to the public but needs FDA approval prior to clinical use.

9.2.2 Future Direction

Specific, validated applications of AI in ocular pathology have not yet been widely described. However, as described above, pathology images naturally lend itself to deep learning techniques and uses. As the dataset for ophthalmic histopathology slides expands, it is only a matter of time before AI plays an integral role in ocular pathology as well.

9.3 AI in Ocular Oncology

AI has already shown significant promise in detecting, diagnosing, and predicting oncological conditions with precision, and in some cases even going beyond the reach of clinicians [1, 4]. Given its reliance on radiographic imaging techniques, oncology naturally lends itself to effective AI applications.

9.3.1 Precedents

Use of DL for lung, colorectal, breast, and bladder cancers are all areas of active investigation, with outcome data demonstrating high sensitivities thus far. Notably, not only CNNs have been used to detect and classify various cancers but they have also been used to identify malignant features, such as nodal metastasis, that are typically unreachable and undiagnosable by clinicians. In head and neck squamous cell carcinoma, nodal metastasis and lymph node extranodal extension are notoriously difficult for clinicians to detect [4]. Yet a developed CNN was able to detect them with greater than 85% accuracy, a promising application for prognosis and treatment determination. Additionally, AI methods are being explored in both radiogenomics, where underlying genotypic traits can be predicted from imaging data, and treatment response and toxicity prediction [1, 4]. Specifically, AI programs have been described that predict genitourinary toxicity from prostate radiation, hepatobiliary toxicity after liver short beam radiotherapy, and rectal toxicity from cervical external beam radiotherapy and brachytherapy to varying degrees of accuracy [5–9]. Nevertheless, these networks demonstrate the feasibility of AI to further explore and predict specific treatment response. Furthermore, DL-based natural language processing techniques have potential to predict disease development based on electronic health records (EHRs) [1]. Based on unsupervised processing of patient features from aggregated EHRs of over 700,000 patients, Mt. Sinai was able to predict prostate, rectum, and liver cancers with 93% overall accuracy, with severe diabetes, schizophrenia, and other cancers among the top performing predictive outputs [1, 10].

9.3.2 Choroidal Melanoma

These AI-based techniques will inevitably be useful diagnostic and clinical management tools in ocular oncology as well. Choroidal melanoma outcomes have been a focus for AI applications. Kaiserman et al. were able to forecast five-year mortality from choroidal melanoma based on input variables of patient demographic data and ultrasound tumor details using an artificial neural network (ANN) that performed with greater accuracy (86%) compared to that of an ocular oncologist (70%) [11]. Additionally, another ANN-based method was able to model survival prognosis in patients with choroidal melanoma and determined relevant risk factors to be age, sex, clinical tumor stage, cytogenetic melanoma type, and histologic malignancy grade [12].

9.3.3 Carcinoma Reconstructive Surgery

Habibalahi et al. demonstrated the feasibility of using ML techniques to detect and demarcate boundaries of ocular surface squamous neoplasia based on multispectral imaging [13]. Correlation between spectral image analysis and histology assessment, based on H&E sections, was 94%. Additionally, this technique was able to reliably differentiate neoplastic tissue from normal tissue both intrapatient

($p < 0.0005$) and interpatients ($p < 0.001$). Impressively, the imaging system used does not require contact with the ocular surface and can capture multiple images and predict in less than 3 min, providing a real-time clinical tool. Furthermore, the technique described can be built into slit lamps in clinic or surgical equipment in ORs, easily integrating into existing workflow processes [13]. Tan et al. designed a risk stratification system that predicts complexity of reconstructive surgery after periocular basal cell carcinoma excision [14]. The group discovered three significant predictive variables: preoperative assessment of complexity, surgical delays, and tumor size. Based on their findings, they proposed a new algorithmic approach to timing basal cell carcinoma excisions.

9.3.4 Retinoblastoma and Leukocoria

Retinoblastoma is the most common cause of primary intraocular malignancy of childhood and accounts for 10–15% of cancers that occur in the first year of life [15]. This malignancy can grow and metastasize rapidly. Therefore, early diagnosis and treatment is critical to improve chance of survival, vision preservation, and minimize the need for toxic treatment. A red pupillary reflex test is typically performed using a handheld direct ophthalmoscope to screen children for leukocoria, the most common presenting symptom of retinoblastoma. In addition to retinoblastoma, leukocoria can be a symptom for multiple pediatric eye disorders including pediatric cataract, retinopathy of prematurity (ROP), Coats' disease, persistent fetal vasculature, and strabismus. In patients with leukocoria, rather than eliciting a red-colored reflection, the test reveals a white or yellow reflection. Though this screening test is required in regular pediatric checkups, studies have reported that signs of retinoblastoma are first detected by pediatricians in only 8% of cases compared to 80% by family and friends [16]. Of importance, the reflex can also be detected when patients' eyes are photographed with a camera flash.

AI can be used to improve screening and detection of retinoblastoma. Munson and colleagues developed CRADLE (Computer-Assisted Detector of Leukocoria), a smartphone application designed to screen children for leukocoria [17]. Affected patients had retinoblastoma, Coats' disease, amblyopia, hyperopia, and cataracts. CRADLE remodeled and now uses an embedded CNN that was previously designed to detect leukocoria in nonclinical settings [18]. Importantly, the 52,982 longitudinal photographs used to train and test the algorithm are all collected by the patients' parents in casual settings. A retrospective study on CRADLE's performance demonstrated that not only was the application able to detect leukocoria in 80% of patients with eye disorders, but it also detected leukocoria in photographs that were taken on average 1.3 years before diagnosis [17]. For a deadly disease like retinoblastoma, even shaving mere months off time to diagnosis is crucial to a patient's clinical course. Vision loss and need for interventions like chemotherapy and radiation therapy that have long-term toxicities can be reduced. The added benefit (and technical complexity) of the CRADLE app is its use of photographs taken by parents in various settings, significantly augmenting detection of potentially referral-warranted eye conditions.

9.3.5 Future Direction

Despite the examples described above, ocular tumors are not often seen. Therefore, the datasets on which algorithms must be trained are not robust enough for accuracy, reliability, and reproducibility [19]. Nonetheless, there is tremendous promise that AI will transform the way oncology as a field operates—malignancies can be screened and detected earlier, neoplasia characteristics described more in more granular phenotypic and genotypic details, treatment response and potential toxicity predicted, and prognosis estimated.

9.4 AI in Ocular Genetics

AI applications in ocular genetics remain relatively untouched, although ophthalmology's heavy use of imaging technology provides a wealth of source material for genetic prediction. Genetics play various roles in AI—they can be both inputs (e.g., genetic risk factors) and outputs.

9.4.1 Precedents

As discussed above, preceding work within general oncology predicting genetic information from radiographic and histopathological images highlights the potential of ocular image-based genetic prediction. To elaborate further, one study developed and trained a CNN to evaluate brain MRIs of patients with both low- and high-grade glioma in order to predict underlying genotypic traits [20]. Specifically, the system was able to independently predict both IDH mutation and MGMT methylation status with 94% and 83% clinical accuracy, respectively. For a disease such as glioma, determination of genotype is critical for appropriate therapy. Similarly, six different gene mutations in lung adenocarcinoma—*STK11, EGFR, FAT1, SETBP1, KRAS*, and *TP53*—were able to be predicted with AUCs from 0.733 to 0.856 based on pathology image inputs [3]. This model can be applied to any cancer type, including ocular cancers. Both studies demonstrate the value of creating algorithms that can expedite oncological workup and inform clinical decision-making.

9.4.2 Inherited Retinal Disorders (IRDs)

IRD is a burgeoning area of research and development of deep learning applications. Through next-generation sequencing, over 250 genes and 300 genes and loci involved in retinal dystrophies have already been identified, opening the door for emerging gene therapies and pharmacological agents, among others [21]. Additional large cohort studies have also greatly expanded knowledge about IRD clinical manifestations, including fundus and morphological appearances. Combined, both genetic and clinical studies have been able to associate characteristic morphologic features to specific genes in various retinal dystrophies, such as *ABCA4*, *RP1L1,* and

EYS. Fujinami-Yokokawa and colleagues trained and developed deep neural networks to accurately predict the causative genes in macular dystrophy (*ABCA4* and *RP1L1*) in comparison to those in retinitis pigmentosa (*EYS*) based off of spectral-domain optical coherence tomographic (SD-OCT) images. The model yielded a mean overall test accuracy of 90.9%, with individual category accuracies of 100% for *ABCA4*, 78% for *RP1L1*, and 89.8% for *EYS*, and 93.4% for normal subjects.

9.4.3 Future Direction

Ultimately, the purpose of genetic studies is to inform clinical diagnosis and management. AI applications in IRDs and other specialties demonstrate the value in linking genetic prediction with clinical management. These applications become especially valuable for conditions in which specific expertise is scarce, such as in IRD, and incidence low leading to limited information. There is no doubt that genetics will also be incorporated into general screening to better augment clinical diagnoses.

9.5 AI in Pediatric Ophthalmology

9.5.1 Retinopathy of Prematurity

Retinopathy of prematurity (ROP) is one of the leading causes of childhood blindness worldwide, particularly in middle-income countries [22]. ROP incidence was 19.9% in the United States and accounts for 6–18% of childhood blindness [19]. ROP occurs when there is abnormal retinal neovascularization, which can lead to retinal edema, hemorrhage, scarring, and retinal detachment [22]. If the condition is severe enough and progresses to plus disease and subsequent retinal detachment, permanent visual impairment and blindness can result. However, ROP is treatable if intervention, such as laser photocoagulation and anti-vascular endothelial growth factors (anti-VEGF), is staged in timely manner. Early treatment has been shown to improve chance of better retinal structure outcome and visual acuity in patients with high-risk prethreshold ROP, though 9% still resulted in blindness [19, 23]. Significant risk factors include prematurity and low birth weight. Thus, as survival rates of premature infants increase, the necessity of early ROP screening and intervention becomes critical.

Currently, ROP screening and diagnostic exams are difficult to perform on infants, subjective, and time consuming for the following reasons [22]. Primarily, diagnostic classification differs significantly among experts [19]. The International Classification for Retinopathy of Prematurity (ICROP) provides a standard system to assess disease extent and severity—it looks at four main features: zone, stage, extent, and presence of plus disease [24]. Plus disease is the most important indicator of disease severity and is defined as abnormal blood vessel dilation and tortuosity [25]. It is the most severe form of ROP and is usually quickly followed by retinal detachment. Previously, patients were treated when they had reached "threshold ROP," but now patients are treated when they develop high-risk prethreshold ROP or type 1 ROP. Variable diagnostic classification becomes problematic especially when evaluating for plus disease. Plus disease diagnosis has traditionally been a binary choice—plus vs pre-plus [25]. However, research has indicated that retinal vasculature changes occur on a continuous spectrum, making it difficult to standardize a specific "cutoff" in order to be plus disease [26]. Thus, experts have differed on multiple components of diagnostic classification including even which retinal vascular features to consider or whether additional factors should be included, all leading to inconsistent and unreliable end classification. Furthermore, there has also been inconsistent expert diagnoses on pre-plus disease [27].

AI has the potential to overcome these challenges and remove barriers to scalable ROP screening. AI-powered systems use retinal fundus photos to detect and grade ROP or plus disease. Traditional algorithms utilize "feature engineering" in which images are preprocessed and manually classified with explicit rules [22]. For ROP images, these include vessel dilation, tortuosity, and venular features—however, ROP is challenging in that understanding of symptoms and presentation is still lacking. Plus disease is more easily discernable. Thus, recently, CNNs, a branch of deep learning, have been used to yield improved results. CNNs are fed large-scale ROP datasets and automatically learn latent features, thus bypassing the preprocessing required in other image classification algorithms. DeepROP and i-ROP-DL are two recent systems that have yielded consensus results with expert opinion and even improved disease detection over some experts.

DeepROP is trained on a large database constructed with images labeled "normal," "minor ROP," and "severe ROP"—two CNNs for identification and subsequent grading are used on a cloud-based platform [28]. When clinically implemented, DeepROP identification performed with high sensitivity and accuracy, though grading remained a challenge. When tested against a panel of clinicians, DeepROP outperformed only one of the three experts. The differences between "minor ROP" and "severe ROP" are less easily discernable compared to those between "normal" and any ROP cases. However, DeepROP results do demonstrate that if CNNs are trained on large-scale datasets without manual classification, ROP can be detected with high fidelity.

i-ROP-DL uses deep learning to detect plus disease and quantify severity of retinal vascular abnormality on a 1–9 scale [29]. It was trained on a set of posterior pole fundus images labeled with (1) "no ROP," (2) "mild ROP" defined as ROP less than type 2, (3) "type 2 ROP" defined as zone I, stage 1 or 2, without plus disease or zone II, stage 3, without plus disease, and (4) "type 1 or treatment-requiring ROP," defined as zone I, any stage, with plus disease; zone I, stage 3, without plus disease; or zone II, stage 2 or 3, with plus disease. Additionally, an additional category "clinically significant ROP" sought to capture referral-warranted cases—these include type 1, type 2, and pre-plus disease categories as described above. i-ROP-DL also involves two CNNs trained for retinal vessel segmentation and plus disease detection. i-ROP-DL was able to accurately detect clinically significant ROP with 94% sensitivity for type 1, and furthermore, demonstrate that its retinal vascular severity scale is strongly correlated with expert assessment of disease severity. When tested against eight ROP experts, i-ROP outperformed six of the eight experts: i-ROP diagnosed 91% of the images accurately whereas experts had a mean accuracy of 82% [30].

DeepROP and i-ROP-DL are only two systems utilizing CNNs to improve ROP screening and diagnosis. While study results of high sensitivity and accuracy have been encouraging, in order for these to be implemented as screening tools, sensitivity and negative predictive value need to further increase. Furthermore, additional outputs needed include zones and stages, both of which have been challenging to differentiate. Importantly, aside from being a helpful clinical tool, DL systems can shed light on objective measures to evaluate ROP classification and progression [19].

However, as is the case for many DL applications, in order to improve model robustness, training datasets need to significantly expand to include a greater number of high-quality images that encompasses a broad set of clinical features. There is software in development that automatically recognizes whether individual images are of sufficient quality [31].

9.5.2 Pediatric Cataracts

Pediatric cataracts are a common and preventable cause of visual impairment and possible permanent visual loss worldwide [32]. Cataracts interfere with normal visual development, and if left untreated, especially during the critical period of birth to 5 years of age, may lead to irreversible amblyopia. Thus, early detection and surgical intervention are critical. However, cataract removal is also complicated by posterior capsule opacification (PCO) and secondary glaucoma. Cataracts are currently visualized and diagnosed using a slit-lamp exam, which can often be difficult to perform on pediatric patients, subjective and of poor quality.

CC-Cruiser is a cloud-based AI platform that screens, stratifies, and recommends treatment based on slit-lamp images from patients who may have congenital cataracts (Long, Lin—Nature). It consists of three CNNs that: (1) screens and identifies potential patients with cataracts, (2) evaluates disease severity (lens opacity) based on opacity area, density, and location and risk stratify, and (3) provides treatment recommendation—surgery or follow up—based on risk stratification. A multicenter randomized trial using CC-Cruiser demonstrated that the platform's accuracies in diagnosing cataract presence and recommending appropriate treatment were 87.4% and 70.8%, respectively, significantly underperforming experts that obtained accuracies of 99.1% and 96.7% [33]. However, CC-Cruiser was able to deliver a diagnosis within 3 min on average compared to the nearly 9 min for experts, resulting in high patient satisfaction.

Postoperative complication predictive systems are also in development. Random forest and Naïve Bayesian classifier have been used to predict severe lens proliferation into the visual axis (SLPVA) and abnormal high intraocular pressure (AHIP) with accuracies above 76% and 75%, respectively, for the two programs [34]. Zhang and colleagues were also able to study which specific factors affected accuracy levels—specifically, prediction improved for composite complications (SLPVA and AHIP) when gender and age were excluded, for SLPVA only when secondary IOL placement, operation mode, age, and area of cataracts were excluded, and finally for AHIP only when gender, operation mode, and laterality were removed. Similarly, another CNN system has been able to detect PCO severe enough to warrant surgery with an accuracy of 92% [22, 35].

Though cataracts are seen commonly, cataract imaging is lacking. Therefore, similar to ocular tumors, algorithms cannot be reliably trained on such a small sample. Additionally, pediatric cataracts have different characteristic and risk profiles, and so, existing models and datasets for adult cataracts cannot be used.

9.5.3 Strabismus

Strabismus is a condition of ocular misalignment and is screened with simple clinical exams such as the Hirschberg test and cover test [36]. In clinical settings with specialized equipment, CNNs can detect strabismus from fixation deviations in eye-tracking data (accuracy of 95%) as well as from retinal birefringence scanning (accuracy of 100%) with high sensitivity and specificity [37, 38]. Van Eenwyk et al. have also described applying AI techniques to analyze Bruckner pupil red reflex imaging and photorefraction videos for amblyogenic factors [39]. The designated output for the algorithm was whether to refer or not refer to an ophthalmologist. The decision tree made the same refer/do not refer decision in 77% of cases as clinicians, who were considered the "gold standard."

However, in settings where there is poor access to ophthalmic clinics and equipment, a model has been proposed to develop a telemedicine-purposed CNN from images gathered in telemedicine settings. Aiding both tele-ophthalmology and clinic strabismus diagnosis, an AI-powered mobile platform has been described that locates and classifies eye versions in nine gaze positions [40].

9.6 Additional AI Applications in Ophthalmology

The retina not only houses information about pathological eye disorders, but it is also the window into the rest of the body. Currently, however, clinicians do not possess any tools that link together the eye and other systemic conditions and enable accurate interpretation of retinal imaging. Retinal images provide a wealth of data to evaluate cardiovascular health, anemia, central nervous system (CNS) disorders including multiple sclerosis, Alzheimer disease, Parkinson's disease, among many other ailments.

Ophthalmic applications of AI in predicting the above conditions go beyond the existing abilities of providers and any clinical tool. These methods can and will form the basis of noninvasive, inexpensive methods of screening, diagnosing, and monitoring various systemic conditions.

9.6.1 Cardiovascular Risk Factors

Cardiovascular disease (CVD) remains the leading cause of death globally [41].The healthcare community faces a shortage of manpower and deficient infrastructure to effectively address this growing global demand. Accurate, time efficient, and scalable approaches to assess CVD risk in patients continue to be a critical and urgent need in order to predict CV outcomes including heart disease, stroke, chronic kidney disease (CKD), and mortality, as well as prevent adverse CV events such as heart attacks and strokes [42]. Frequently used risk calculators such as Framingham, SCORE (systematic coronary risk evaluation model), and Pooled Cohort all require invasive blood tests and pre-draw fasts.

The retina, which enables the only noninvasive approach to microvasculature visualization, is a promising source of information on a patient's CV risk [42]. Ocular manifestations of CVDs include hypertensive retinopathy, cholesterol emboli and occlusions, flame hemorrhages, cotton-wool spots, and ischemic events. Prior work uncovered that retinal vasculature measurements (artery and vein diameters) in fundus photos are associated with CVD outcomes. For example, CVD risk is higher in patients with narrower retinal arterioles and wider venules. For patients with diabetes, predictive power increases when retinal imaging details are supplemented with additional measurements including established risk factors (blood pressure) and CRP level [42]. Furthermore, investigation has shown that retinal imaging improves stroke risk prediction beyond that of typical risk factors (age, sex, blood pressure, total cholesterol, LDL cholesterol, glycosylated hemoglobin (HbA1c), and antihypertensive medication use) [43].

However, interpreting and evaluating retinal images is a tedious and time-consuming task for clinicians, even with semiautomated software [19].

In a study by Verily, Google's research organization, Poplin and colleagues have developed and validated a DL-based method that is able to predict multiple CV risk factors better than baseline solely based on retinal fundus images—these risk factors include age, gender, smoking status, systolic blood pressure, and body mass index [41, 44]. The model was trained using retinal fundus images from 48,101 patients from the UK biobank study and 236,234 patients from the eyePACS (picture archive and communication system) population. The study also created an additional algorithm that predicted the onset of major adverse cardiovascular events (MACE) within 5 years using fundus photographs alone. MACE included unstable angina, myocardial infarction, or stroke or death from CV causes. The model performed fairly accurately with an AUC of 0.70 (95% CI: 0.65, 0.74), especially when compared to the SCORE risk calculator's AUC of 0.72 (0.67, 0.76).

The study also generated soft attention maps to understand the anatomical regions upon which the model was basing its predictions, and compared them to clinician blinded assessment. In predicting age, smoking, and systolic blood pressure, the algorithm drew from blood vessel patterns. HbA1c predictions were primarily based on perivascular surroundings, and gender was predicted using optic disc, vessels, and macula, with signal across the retina as well [41]. This technique already surpasses current physician abilities to distinguish gender, smoking status, blood pressure measurements from fundus photos.

Despite significant progress and encouraging results, these DL methods must be clinically validated and trained on larger, more diverse datasets. Though retinal fundus imaging may be additive to current CV risk assessments, further exploration on whether it can also replace certain risk markers is needed.

9.6.2 Anemia

DL-powered systems can also be noninvasive screening tools for anemia based on retinal fundus imaging. Anemia is a widespread public health issue affecting millions and can be a symptom of a dangerous underlying condition. A DL model developed by Google uses a combination of images and patient demographic metadata (e.g., sex, age) to predict hemoglobin (Hb) concentration and anemia [45]. Notably, the system demonstrated that the fundus-only model was more accurate than metadata-only model for both hb and anemia predictions, though the combination model of both fundus photos and metadata yielded the most accurate results of the three. Specifically, compared to the metadata-only model, DL improved anemia detection by 14%.

Another research group validated Google's model on an independent dataset from Asia (Google's dataset was primarily Caucasian), suggesting the system's generalizability [46].

However, anemia has a number of subtypes and diverse etiologies, making diagnosis and management challenging for AI-powered systems. While the model described above has not yet been able to predict subtype and/or etiology pathophysiology from fundus images, that information could be vital if AI were to be a possible tool [45]. Furthermore, there is potential that these systems can develop to not only estimate Hb concentration, but also the levels of other elements in a traditional complete blood count (CBC) panel. Such a tool could potentially augment or replace standard blood tests, and would be invaluable.

9.6.3 Multiple Sclerosis

Multiple sclerosis (MS) is a neurodegenerative disorder marked by demyelination of the CNS, leading to progressive clinical disability. MS is characterized by areas of demyelination, loss of oligodendrocytes, astroglial scarring, and axonal injury. Diagnosis is based on the McDonald criteria which encompasses clinical attacks and objective evidence including brain and spinal cord MRIs and cerebrospinal fluid exams. While MS course and progression greatly vary, patients commonly first present with episodes of optic neuritis or vision loss and diplopia due to internuclear ophthalmoplegia.

Diagnosis Recently, studies have found a correlation between MS and both axonal loss in the optic nerve and macular thickness loss and have thus begun investigating the use of retinal OCTs to supplement existing MS diagnostic processes [47, 48]. The retina is composed of retinal ganglion cells and their axons, which form the retinal nerve fiber layer (RNFL). Based on OCT image captures, many studies have found and confirmed that patients with MS have thinner RNFLs compared to healthy controls, both with and without optic neuritis [49]. Furthermore, another study demonstrated that in MS patients without optic neuritis, OCT-measured RNFL thickness was a marker for subsequent neurological disability in MS. [50] Specifically, patients with RNFL thicknesses <88 micrometers had twice the risk of worsened disability compared to those with thicker RNFLs.

Work on creating machine learning applications to improve MS diagnosis using retinal OCT images is well underway. Del Palomar and colleagues compared three different machine learning algorithms—decision trees, multilayer perceptron, and support vector machine—to evaluate the diagnostic value of RNFL and ganglion cell layer (GCL) thickness loss, as measured by SS-OCT. [47] The study found that the decision trees yielded the best prediction (97.4%) using RNFL data. Additionally, across the three machine learning techniques, RNFL thickness loss data was better in classifying MS or healthy compared to GCL thickness loss data.

Supporting this study, Cavaliere and colleagues demonstrated the efficacy of using support vector machine to detect MS in patients without a history of optic neuritis [49]. In the first stage of the study, the authors manually identified OCT features with the greatest discriminant capacity to be used in the SVM. These features were global GCL thickness at the peripapillary area (which includes both RNFL and GCL), macular retina thickness in the nasal quadrant of the inner ring, and macular retinal thickness in the nasal quadrant of the outer ring. Subsequently, these selected features were used as inputs in the automatic classifier in the SVM. The diagnostic results from the SVM yielded a sensitivity of 0.89, specificity of 0.92, accuracy of 0.91, and AUC of 0.97.

While the McDonald's criteria for MS diagnosis primarily relies on neuroimaging and CSF exams, retinal degeneration and its easy visualization can be a future diagnostic tool as demonstrated by the studies described above. Additional studies further honing not only algorithmic accuracy and specificity but also specific markers within the retina are needed first.

Progression Monitoring Monitoring and predicting MS course as well as evaluating response to therapy are crucial to providing appropriate interventions and disease-modifying therapies. However, validated specific biomarkers and objective tools are lacking. There is promise that machine learning, in addition to having diagnostic capabilities, also is able to predict course of disease, response to therapy, and improve understanding of MS progression.

A recent study based its focus on findings that a greater number of fixational microsaccades, small eye motions, are correlated with higher MS disability metrics [51]. Additionally, patients with MS had significantly larger vertical amplitude components and lower peak accelerations. Models predicting disability based on these features obtained greater than 80% accuracy. C. Light Technologies created a retinal imaging instrument that tracks micro eye motions and when paired with its machine learning algorithm is able to provide insight into MS progression [52].

Another area of active AI development builds upon prior research suggesting that the retinal inner nuclear layer volume reflects severity of disease [53] and that monitoring this layer via OCT can be an effective form of evaluating immunotherapy response [54]. Knier et al. studied MS patients not on therapy, on first-line therapy, or on second-line therapy, and found that thinning of the inner nuclear layer and thickening of the macula were associated with reduced disease activity and better control of the disease [54].

Therefore, accurately quantifying the volume and thickness of the different retinal layers is an important step to assess level of disease activity. Existing machine learning methods that automatically segment and measure retinal layers involve significant manual preprocessing for feature and parameter selection, a tedious and time-consuming process [55]. He and colleagues have developed a deep neural network that bypasses manual feature and parameter selection and can fully segment 3D retinal layers in only 10 s with accuracy levels comparable to existing segmentation methods [55].

Future Direction AI has a twofold role in MS research—not only can it be used for its diagnostic and monitoring capabilities, but more importantly, it can shed insight into MS pathophysiology and course progression, areas that remain poorly understood.

9.6.4 Other Neurodegenerative Diseases

As seen in its role in multiple sclerosis, the retina is being explored as a barometer for other neurodegenerative diseases such as Alzheimer's and Parkinson's diseases. Studies have explored various retinal biomarkers as tools to detect Alzheimer's and Parkinson's. RNFL thinning, macular volume and thickness, retinal vasculature parameters have all been linked to Alzheimer's, though results have been inconsistent [56].

Rather than focusing on retinal layers, one study built a machine learning-enabled classification model that differentiated between patients with Alzheimer's, Parkinson's, and who are healthy based on retinal "texture" biomarkers [57]. Texture analysis is an emerging technique that enables quantification of signal changes that are not visible among existing image pixels [58]. This method has been increasingly used in neurodegenerative research due to its potential to detect changes earlier. Their study yielded a median sensitivity of 88.7%, 79.5%, and 77.8% for healthy controls, Alzheimer's, and Parkinson's, respectively.

Specifically focusing on Alzheimer's disease, Optina Diagnostics explores the link between retinal vasculature and β-amyloid plaques. Using an AI-enabled hyperspectral retinal imaging platform, researchers discovered that retinal venules in patients with β-amyloid plaques in the brain have higher tortuosity than in patients who are β-amyloid negative [59]. Furthermore, retinal texture analysis also revealed significant differences between these two cohorts. Optina Diagnostics seeks to predict cerebral β-amyloid status in patients with Alzheimer's and further elucidate the relationship between the disease pathophysiology and clinical course [60]. Additionally, their platform recently received Breakthrough Device Designation from FDA, enabling an expedited pathway for device development, assessment, and review [61].

AlzEye is another initiative that has amassed over 2 million retinal photos and scans of 250,000 patients who developed Alzheimer's disease and other forms of dementia in its efforts to explore retinal changes in these neurodegenerative diseases [62]. As incoming data continues to accumulate, this provides researchers with a tremendous source of information from which to investigate and validate specific markers of clinical significance.

The link between neurodegenerative disease and retina is an active line of research. Existing studies are based on small cohorts and limited data. These discriminatory markers, while promising, remain vague and not specific—further exploration with larger datasets and clinical validation is needed.

9.7 AI in Telemedicine

Artificial intelligence applications in tele-ophthalmology have the power to transform eye care delivery. Increasing access to care, cost savings for patients, providers, and payers, and reducing provider burden are just a few of many benefits that tele-ophthalmology brings. Three delivery methods have been described for tele-ophthalmology services:

- *Asynchronous:* This "store and forward" technique captures clinical information (i.e., imaging scans) at one site and sends to another (i.e., ophthalmologist) for clinical assessment. This method is already commonly used for diabetic retinopathy [63].
- *Synchronous:* This method features a real-time telemedicine interaction between patients and ophthalmologists via various communication channels (i.e., video chat, telephone call, smartphone application). Synchronous tele-ophthalmology services have been successfully implemented in many urgent care centers and emergency rooms, where immediate remote ophthalmic triage services are provided to centers that lack robust eye care services.
- *Remote monitoring:* This form of delivery enables providers to monitor patients at home or at a distance. Intraocular pressure (IOP) contact lenses, IOP home monitoring devices, IOP sensors, visual field devices, and age-related macular degeneration (AMD) devices are examples of such devices that automatically transmit captured data to providers to augment clinical management [64].

While the applications of AI described in the above sections largely focus on diagnostic and monitoring capabilities, AI uses actually permeate the entire health-care continuum.

- *Administrative workload:* AI programs reduce administrative burden through smart scheduling, automatic invoice generators, patient follow-up, and insurance claim management [65]. A few German health insurers have implemented AI in claims management, and early estimates indicate that health insurers could save up EUR 500 million annually with this technology. Specifically, AI is able to systematically classify claims that are likely to be successful, and route those that may need to be manually reviewed, thereby optimizing staff capacity on only claims that require attention.
- *Robotics and procedures:* AI is implemented in many ophthalmic devices as guides for auto-alignment, focus, and data capture. Additionally, AI has enabled minimally invasive surgery and remote surgery and slit-lamp exams, options that will become increasingly commonplace as technology advances [63, 66]. Robotic tools such as co-manipulation devices, tele-manipulators, and steady hands are currently being developed [67]. For example, Preceyes surgical system achieved the first successful robot-assisted intraocular surgery [68]. Da Vinci surgical systems have also been used to perform pterygium repairs and corneal surgeries [67]. AI-guided robotics are particularly important in increasing access to care in underserved and remote communities.
- *Diagnostics and screening:* The interpretative and predictive capabilities of AI are well described. Integrating these skills for tele-diagnosis and detection greatly increases access to care, reduces unnecessary visits to ophthalmologists, and saves both money and time for patients, providers, and payors. Tele-ophthalmology and AI already have demonstrated its benefit for diabetic retinopathy and are being explored for glaucoma screening, AMD monitoring, abnormal cornea topography detection, and pediatric cataract screening, among others [63]. FDA-approved IDx-Dr the first provider-independent AI platform that reads and interprets fundus photos to assess for referable diabetic retinopathy [69]. IDx-Dr surpassed predefined sensitivity and specificity levels during trial by achieving 87.2% sensitivity and 90.7% specificity [70]. This system is the first FDA-approved AI diagnostic system in all of medicine.
- *Remote monitoring:* As previously mentioned, remote monitoring is the next frontier of tele-ophthalmology. AI programs can sort through large swathes of individual patient information and assist providers with clinical decision-making.

The potential for AI in tele-ophthalmology is clear. The next step for the field is constructing a robust enough tele-image dataset, on which ML algorithms can be trained. Simultaneously, consumer technology and tele-ophthalmology technology must continue to advance to enable mass adoption. Once accomplished, AI will have an important role in managing patients in and out of the clinic.

9.8 Limitations

Despite the tremendous potential that AI possesses in serving as powerful clinical tools, clinical implementation and widespread adoption still face tall barriers.

AI clinical applications are still in their nascence, with unstandardized processes and limited peer-reviewed publications. Many of the studies described in this section trained and tested algorithms on datasets that were not systematically evaluated for quality and bias or clinically validated. Thus, standardizing processes for clinical AI applications is the next critical step [71]. A systematic approach to building and assessing robust training, testing, and external validation datasets must be established—data quality, sources, and aggregation methods have to be evaluated according to an accepted standard to ensure unbiased, generalizable, and accurate outputs. Additionally, as of yet, there are no existing means of determining clinical readiness of AI applications. Acceptable levels of accuracy, precision, and/or AUC that algorithms must achieve as well as potential risks and liabilities must be agreed upon. Furthermore, the clinical validation process must be further elucidated.

Only with standardized processes, and perhaps regulatory oversight, that promote development of safe, inclusive, and accurate platforms, can AI gain a foothold in clinical management.

9.9 Conclusion

The utility and many capabilities of AI in ophthalmology are clear. The power to simplify and automate screening, diagnosis, and monitoring with comparable, if not improved, accuracy cannot be understated. AI is already reweaving the fabric of clinical practice in ocular oncology, pathology, pediatric ophthalmology, and teleophthalmology. Researchers are discovering and confirming new links between the eye and the rest of the body, introducing a whole host of noninvasive techniques to assess a patient's overall health.

Not only does AI revolutionize clinical practice, but it has also pushed research forward in areas that we otherwise would not have been able to reach with existing technology. Our understanding of disease processes, clinical manifestations, and treatment responses has accelerated due to AI-enabled processes.

Still, widespread AI implementation faces significant challenges in the future. To create programs that yield unbiased, reproducible, and accurate results, training and testing sets must be large, diverse, and clinically verified. Building these databases require time, collaboration, and patience. Additionally, provider hesitation due to the "black box" nature of AI algorithms remains a real threat to widespread adoption and implementation. Additional research and provider education are needed.

References

1. Kann BH, Thompson R, Thomas CR Jr, Dicker A, Aneja S. Artificial intelligence in oncology: current applications and future directions. Oncology. 2019;33:46–53.
2. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast Cancer. JAMA. 2017;318(22):2199–210.
3. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24:1559–67.
4. Kann BH, Aneja S, Loganadane GV, Kelly JR, Smith SM, Decker RH, Yu JB, Park HS, Yarbrough WG, Malhotra A, Burtness BA, Husain ZA. Pretreatment identification of head and neck Cancer nodal metastasis and Extranodal extension using deep learning neural networks. Sci Rep. 2018 Sep 19;8(1):14036.
5. Pella A, Cambria R, Riboldi M, et al. Use of machine learning methods for prediction of acute toxicity in organs at risk following prostate radiotherapy. Med Phys. 2011;38:2859–67.
6. Carrara M, Massari E, Cicchetti A, Giandini T, Avuzzi B, Palorini F, et al. Development of a ready-to-use graphical tool based on artificial neural network classification: application for the prediction of late fecal incontinence after prostate cancer radiation therapy. Int J Radiat Oncol. 2018;102(5):1533–42.
7. Lee S, Kerns S, Ostrer H, et al. Machine learning on a genome-wide association study to predict late genitourinary toxicity after prostate radiation therapy. Int J Radiat Oncol Biol Phys. 2018;101:128–35.
8. Ibragimov B, Toesca D, Chang D, et al. Development of deep neural network for individualized hepatobiliary toxicity prediction after liver SBRT. Med Phys. 2018;45:4763–74.
9. Zhen X, Chen J, Zhong Z, et al. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study. Phys Med Biol. 2017;62:8246–63.
10. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep. 2016;6:26094.
11. Kaiserman I, Rosner M, Pe'er J. forecasting the prognosis of choroidal melanoma with an artificial neural network. Ophthalmology. 2005;112:1608.
12. Damato B, Eleuteri A, Fisher AC, Coupland SE, Taktak AF. Artificial neural networks estimating survival probability after treatment of choroidal melanoma. Ophthalmology. 2008;115(9):1598–607.
13. Habibalahi A, Bala C, Allende A, Anwer AG, Goldys EM. Novel automated non-invasive detection of ocular surface squamous neoplasia using multispectral autofluorescence imaging. Ocul Surf. 2019;17(3):540–50.
14. Tan E, Lin F, Sheck L, et al. A practical decision-tree model to predict complexity of reconstructive surgery after periocular basal cell carcinoma excision. J Eur Acad Dermatol Venereol. 2017;31:717–23.
15. Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA, editors. SEER cancer statistics review, 1975–2012. Bethesda, MD: National Cancer Institute. Available at: http://seer.cancer.gov/csr/1975_2012.
16. Abramson DH, Beaverson K, Sangani P, Vora RA, Lee TC, Hochberg HM, Kirszrot J, Ranjithan M. Screening for retinoblastoma: presenting signs as prognosticators of patient and ocular survival. Pediatrics. 2003;112(6 Pt 1):1248–55.
17. Munson MC, Plewman DL, Baumer KM, et al. Autonomous early detection of eye disease in childhood photographs. Sci Adv. 2019;5:eaax6363.

18. Rivas-Perea P, Baker E, Hamerly G, Shaw BF. Detection of leukocoria using a soft fusion of expert classifiers under non-clinical settings. BMC Ophthalmol. 2014;14:110.
19. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Tan GSW, Schmetterer L, Keane PA, Wong TY. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019;103(2):167–75.
20. Chang P, Grinband J, Weinberg BD, Bardis M, Khy M, Cadena G, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. AJNR Am J Neuroradiol. 2018;39(7):1201–7.
21. Fujinami-Yokokawa Y, Pontikos N, Yang L, et al. Prediction of causative genes in inherited retinal disorders from spectral-domain optical coherence tomography utilizing deep learning techniques. J Ophthalmol. 2019;2019:1691064.
22. Reid JE, Eaton E. Artificial intelligence for pediatric ophthalmology. Curr Opin Ophthalmol. 2019;30:337–46.
23. Early Treatment for Retinopathy of Prematurity Cooperative Group, Good WV, Hardy RJ, Dobson V, Palmer EA, Phelps DL, Tung B, Redford M. Final visual acuity results in the early treatment for retinopathy of prematurity study. Arch Ophthalmol. 2010;128(6):663–71.
24. Classification of Retinopathy of Prematurity*. The international classification of retinopathy of prematurity revisited. Arch Ophthalmol. 2005;123(7):991–9.
25. Solarte CE, Awad AH, Wilson CM, Ells A. Plus disease: why is it important in retinopathy of prematurity? Middle East Afr J Ophthalmol. 2010;17(2):148–55.
26. Redd TK, Campbell JD, Brown JC. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. Br J Ophthalmol. 2019;103:580–4.
27. Chiang MF, Jiang L, Gelman R, Du YE, Flynn JT. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. Arch Ophthalmol. 2007;125(7):875–80.
28. Wang J, Ju R, Chen Y, et al. Automated retinopathy of prematurity screening using deep neural networks. EBioMedicine. 2018;35:361–8.
29. Redd T, Campbell JP, Brown JM, Shah P, Kim SJ, Ostmo S, et al. Utilization of a deep learning image assessment tool for epidemiologic surveillance of retinopathy of prematurity. Invest Ophthalmol Vis Sci. 2019;60(9):1523.
30. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol. 2018;136(7):803–10.
31. Coyner AS, Swan R, Brown JM, et al. Deep learning for image quality assessment of fundus images in retinopathy of prematurity. AMIA Annu Symp Proc. 2018;2018:1224–32.
32. Medsinge A, Nischal KK. Pediatric cataract: challenges and future directions. Clin Ophthal. 2015;9:77–90.
33. Lin H, Li R, Liu Z, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. EClin Med. 2019;9:52–9.
34. Zhang K, Liu X, Jiang J, et al. Prediction of postoperative complications of pediatric cataract patients using data mining. J Transl Med. 2019;17:2.
35. Jiang J, Liu X, Zhang K, et al. Automatic diagnosis of imbalanced ophthalmic images using a cost-sensitive deep convolutional neural network. Biomed Eng Online. 2017;16:132.
36. Lu J, Feng J, Fan Z, Houng L. Automated strabismus detection based on deep neural networks for telemedicine application. IEEE; 2018. p. 8–10. https://arxiv.org/abs/1809.02940.
37. Chen Z, Fu H, Lo WL, Chi Z. Strabismus recognition using eye-tracking data and convolutional neural networks. J Healthcare Eng. 2018;2018:7692198.
38. Gramatikov BI. Detecting central fixation by means of artificial neural networks in a pediatric vision screener using retinal birefringence scanning. Biomed Eng Online. 2017;16:52.
39. Van Eenwyk J, Agah A, Giangiacomo J, Cibis G. Artificial intelligence techniques for automatic screening of amblyogenic factors. Trans Am Ophthalmol Soc. 2008;106:64–73. discussion 73–4
40. de Figueiredo LA, Debert I, Dias JVP, Polati M. An artificial intelligence app for strabismus. Invest Ophthalmol Vis Sci. 2020;61(7):2129.

41. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng. 2018;2(3):158–64.
42. Ting DSW, Wong TY. Eyeing cardiovascular risk factors. Nat Biomed Eng. 2018;2(3):140–1.
43. Cheung CY, Tay WT, Ikram MK, Ong YT, De Silva DA, Chow KY, Wong TY. Retinal microvascular changes and risk of stroke: the Singapore Malay eye study. Stroke. 2013;44(9):2402–8.
44. Akkara JD, Kuriakose A. Role of artificial intelligence and machine learning in ophthalmology. Kerala J Ophthalmol. 2019;31:150–60.
45. Mitani A, Huang A, Venugopalan S, Corrado GS, Peng L, Webster DR, Hammel N, Liu Y, Varadarajan AV. Detection of anaemia from retinal fundus images via deep learning. Nat Biomed Eng. 2020;4(1):18–27.
46. Tham YC, Cheng CY, Wong TY. Detection of anaemia from retinal images. Nat Biomed Eng. 2020;4(1):2–3.
47. Pérez Del Palomar A, Cegoñino J, Montolío A, et al. Swept source optical coherence tomography to early detect multiple sclerosis disease. The use of machine learning techniques. PLoS One. 2019;14(5):e0216410.
48. Fjeldstad C, Bemben M, Pardo G. Reduced retinal nerve fiber layer and macular thickness in patients with multiple sclerosis with no history of optic neuritis identified by the use of spectral domain high-definition optical coherence tomography. J Clin Neurosci. 2011;18(11):1469–72.
49. Cavaliere C, Vilades E, Alonso-Rodríguez MC, et al. Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. Sensors (Basel). 2019;19(23):5323.
50. Martinez-Lapiscina EH, Arnow S, Wilson JA, et al. Retinal thickness measured with optical coherence tomography and risk of disability worsening in multiple sclerosis: a cohort study. Lancet Neurol. 2016;15(6):574–84.
51. Sheehy CK, Bensinger ES, Romeo A, et al. Fixational microsaccades: a quantitative and objective measure of disability in multiple sclerosis. Mult Scler. 2020;26(3):343–53.
52. C. Light Technologies. Technology. https://www.clighttechnologies.com/technology. Accessed 7 July 2020.
53. Saidha S, Sotirchos ES, Ibrahim MA, Crainiceanu CM, Gelfand JM, Sepah YJ, 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;11:963–72.
54. Knier B, Schmidt P, Aly L, et al. Retinal inner nuclear layer volume reflects response to immunotherapy in multiple sclerosis. Brain. 2016;139(11):2855–63.
55. He Y, Carass A, Liu Y, et al. Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT. Biomed Opt Express. 2019;10(10):5042–58. https://doi.org/10.1364/BOE.10.005042.
56. Lee CS, Apte RS. Retinal biomarkers of Alzheimer's disease. Am J Ophthalmol. 2020; S0002-9394(20)30226-9.
57. Nunes A, Silva G, Duque C, et al. Retinal texture biomarkers may help to discriminate between Alzheimer's, Parkinson's, and healthy controls. PLoS One. 2019;14(6):e0218826.
58. Cai JH, He Y, Zhong XL, et al. Magnetic resonance texture analysis in Alzheimer's disease. Acad Radiol. 2020; S1076-6332(20)30033-7.
59. Sharafi SM, Sylvestre JP, Chevrefils C, et al. Vascular retinal biomarkers improves the detection of the likely cerebral amyloid status from hyperspectral retinal images. Alzheimers Dement (N Y). 2019;5:610–7.
60. https://www.prnewswire.com/news-releases/optina-diagnostics-receives-breakthrough-device-designation-from-us-fda-for-a-retinal-imaging-platform-to-aid-in-the-diagnosis-of-alzheimers-disease-300846450.html. Accessed 7 July 2020.
61. https://www.fda.gov/medical-devices/how-study-and-market-your-device/breakthrough-devices-program. Accessed 7 July 2020.
62. https://readingcentre.org/workstreams/artificial_intelligence_hub/alzeye/. Accessed 7 July 2020.
63. ASCRS. Guide to Teleophthalmology. https://ascrs.org/advocacy/regulatory/telemedicine
64. https://www.foreseehome.com/clinical-information/. Accessed 7 July 2020.

65. https://healthcare.mckinsey.com/artificial-intelligence-health-insurance-smart-claims-management-self-learning-software/. Accessed 7 July 2020.
66. Nankivil D, Gonzalez A, Rowaan C, Lee W, Aguilar MC, Parel J-MA. Robotic remote controlled stereo slit lamp. Trans Vis Sci Tech. 2018;7(4):1. https://doi.org/10.1167/tvst.7.4.1.
67. de Smet MD, Naus GJL, Faridpooya K, Mura M. Robotic-assisted surgery in ophthalmology. Curr Opin Ophthalmol. 2018;29(3):248–53.
68. https://www.bbc.com/news/health-37246995. Accessed 7 July 2020.
69. https://www.modernretina.com/view/how-autonomous-diagnostics-dr-gave-jump-start-ai-ophthalmology. Accessed 7 July 2020.
70. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018;1:39.
71. Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. Surv Ophthalmol. 2019;64:233–40.

The Economics of Big Data

10

John Davis Akkara and Anju Kuriakose

10.1 Definitions

Most definitions related to artificial intelligence (AI) and machine learning (ML) are covered in the first chapter "The Terminology of AI." Here we cover a few more definitions related to the economics of big data [1].

Big Data Big data is a term, popularized by John Mashey, used to describe large datasets that are difficult to process using traditional database and software techniques [2].

Data Analytics Data analytics is defined as the science of analyzing raw data to make conclusions about that information. These techniques can reveal trends and metrics that would otherwise be lost in a sea of information [3].

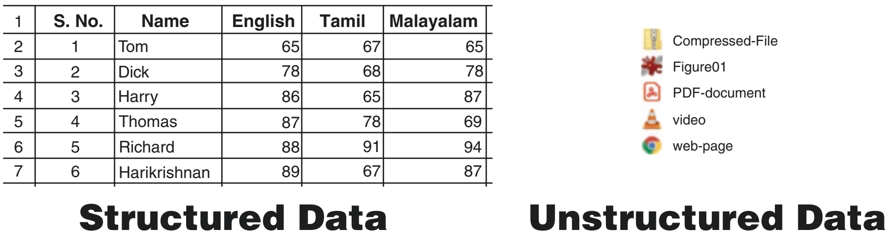
Data Mining Data mining is defined as the process of finding patterns in large datasets involving methods at the intersection of ML, statistics, and database systems [4].

Blockchain Blockchain refers to a growing list of records linked using cryptography via a peer-to-peer network. It works as an open, distributed ledger that can record transactions between two parties efficiently in a verifiable and permanent manner [5].

J. D. Akkara (✉)
Glaucoma Department, Westend Eye Hospital, Cochin, Kerala, India
e-mail: DrJohnDavis@westendeyehospital.com

A. Kuriakose
Retina Department, Westend Eye Hospital, Cochin, Kerala, India
e-mail: DrAnjuKuriakose@westendeyehospital.com

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_10



1	S. No.	Name	English	Tamil	Malayalam
2	1	Tom	65	67	65
3	2	Dick	78	68	78
4	3	Harry	86	65	87
5	4	Thomas	87	78	69
6	5	Richard	88	91	94
7	6	Harikrishnan	89	67	87

Fig. 10.1 Structured vs unstructured data

Structured and Unstructured Data Structured data consists of clearly defined data types, usually in a relational database, which is easily searchable. For example, a patient detail's database or sales transactions (Fig. 10.1). All other data is essentially unstructured data. This may include text files, email, audio, images, video, social media, or websites. Analysis of structured data is much easier than unstructured data [6].

10.2 Introduction to Big Data

We are now in the age of information, the 4th Industrial Revolution (4IR). As the years progressed, we moved from a bottleneck at availability of information to a bottleneck at analyzing information. With globalization and digitization, the amount of easily available information increased, that it greatly outstripped our capacity to analyze by conventional methods. While big data usually refers to huge volumes of data, that is not the complete picture, often it refers to the advanced technologies used to analyze these massive datasets. Although big data and ML seem complicated, they are closely related to the traditional statistical models used for all data analysis [7].

10.2.1 Seven Dimensions of Big Data

Data scientists describe big data in three, four, or sometimes seven dimensions: Volume, Velocity, Variety, Veracity, Variability, Visualization, and Value (Fig. 10.2) [8]. Volume, Variety, and Velocity were the original three dimensions suggested by Laney, while the others were added by IBM, SAS, and Oracle [9]. Volume refers to the amount of data, measured in Petabytes, Zettabytes, or even Yottabytes. Velocity is the speed at which data is accessible. Variety refers to the unstructured nature of the various types of data. Variability refers to the changing nature of the data. Veracity is about the accuracy of the data. Visualization refers to charts and graphs to help visualize the complex data. Value is the benefit derived from the analysis of the data.

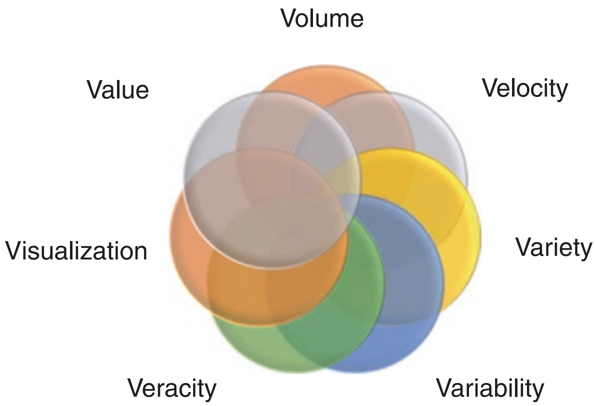


Fig. 10.2 Seven dimensions of big data

10.2.2 How Big Is Big Data?

There is no universal definition on a minimum size for big data in terms of numbers or storage space. Keep in mind that today's big data may be tomorrow's small data. One way to define big data is that it consists of large amount of data distributed over multiple systems. Another way to look at it is, data that requires advanced analytical tools for processing in a reasonable amount of time. The seven dimensions of big data provide a good guide to recognizing it.

10.2.3 Big Data, Small Metadata

Metadata is data that describes other data. It can be keywords that helps to organize big data. There can be descriptive, structural, and administrative metadata. Descriptive metadata gives information about who created a resource, what it is about, and what it includes. Structural metadata gives information about how the data is organized and the structure they are in. Administrative metadata gives information about the origin of resources, their type and access rights.

10.2.4 Information Economics and Economics of Big Data

Information economics is the study of how information and information systems affect an economy and economic decisions [10]. Big data extends our understanding of the value of information. Thus, the economics of big data is a natural extension of information economics. As the economics of information suggests, the monetary value of information must be presented in such a way as to create an opportunity. As such, big data is not free since much investment is needed to search and analyze it for risk assessment. Thus, it is important to consider the different qualities and costs in collecting, analyzing, and applying the data [11, 12].

10.3 Getting Started with Big Data

Big data is not the same for everyone. Especially in healthcare, the data, the required analytics, and the entire strategy would be different. To understand the need for big data in a particular business, there are a few questions to help guide the strategy.

10.3.1 What Data Is Needed?

Medical records are valuable treasure troves of information. Numerical data of parameters like visual acuity, refraction, intraocular pressure, corneal thickness, and ocular biometry are much easier to analyze. This consists of mostly structured data. Big data techniques and machine learning will be useful to analyze unstructured data from medical records and from investigations such as fundus photographs, optical coherence tomography, and fluorescein angiography. Techniques such as natural language processing can derive meaning from unstructured data [13].

10.3.2 Where Do You Get the Data?

Most major hospitals and research institutions now use streamlined Electronic Medical Record (EMR) systems. In most EMRs, the relevant information can be exported from the system in a usable format. With the proper permissions, this data, along with the investigations, can be used for big data analysis or ML [14].

10.3.3 What Can You Do with the Data?

Using big data analytics and ML, one can get insights into various diseases, their diagnostic criteria, and the efficacy of their treatment. One may also gain insights into the relation between several parameters which may otherwise not seem related. Studies that require large number of subjects can easily be conducted using the power of big data.

10.3.4 Who Maintains Ownership of the Data?

Depending on the laws, in most places the institute maintains ownership of the data while respecting the privacy of the patients. Only with a clear consent can the data be used by another agency for data analysis and anonymity of data has to be maintained.

10.3.5 Can you Trust the Data and its Source?

The veracity of the data depends on the source. In many cases, it is noted that data is often incomplete or entered incorrectly, sometimes duplicated or even conflicting. Errors can also be due to incorrect technique, outliers, or mistakes in the investigations. Care has to be taken from the point of data entry to make sure the data is trustworthy and that incorrect information is removed during data cleaning.

10.4 Best Use Practices

Big data generates very powerful information and has to be handled with care. It is essential to follow appropriate best practices for the collection, storage, and use of this data. In addition to the validity and veracity of the information, the unauthorized use of big data can lead to false conclusions and breach of privacy. Medical data is very private and improper handling can lead to disastrous consequences [15].

10.4.1 Electronic Health Records (EHR)

Patient medical records were on paper charts in the past and that is still the case in many places. The rapid progression of technology has enabled the use of electronic medical records which have been widely adopted in most of the developed countries and adopted by some in the developing countries.

Paper records were very limited in accessibility, available to only one person at a time for updating or viewing. Manual updating of these records took time, in addition to the delays in searching for and finding the appropriate record. Storing the paper medical records was often done in large dusty basements of the hospitals due to the large volume of space required and the combined weight of the records. It was very difficult to search for a particular record, and losing old records was common. Security was also limited to locks, doors, and personnel in charge. There was no automatic alert for unauthorized viewing or editing of information, or even documentation of what information was viewed or altered by whom.

A good EHR or EMR system obviates most of these shortcoming of paper medical records. In addition, it allows the creation of structured data by design. This helps in easier analytics of the data to obtain useful information. The four major ethical priorities for EMRs are: privacy/confidentiality, security, data integrity, and availability [16, 17].

10.4.2 Telemedicine

Telemedicine has been playing a minor role in medicine for over 40 years, but with the rapid progress and availability of technology, it has come into the mainstream.

It refers to remote delivery of diagnostics and clinical advice using technology. With excellent cameras and fast internet connections, the doctor–patient interaction has become much smoother. The integration of a good EMR allows quick and accurate documentation. A lot of data will be generated with telemedicine interactions which can be analyzed for useful results [18]. In addition to a videoconferencing camera, various devices in the medical Internet of Things (mIoT) can be used to measure patient vitals or image things like the cornea or retina [19].

10.4.3 Data Collection

Data collection plays a very important role in the big data cycle. Traditionally, structured data is considered more valuable than unstructured data. During the process of data collection, it is prudent to ensure that useful information is correctly tagged and collected in a relational database with data validation measures in place. This would reduce errors and improve data quality, making it much easier for data cleaning processes.

10.4.4 Big Data Storage

The amount of data being gathered is rapidly increasing, and it is coming from heterogeneous devices in various formats. Complex data such as those generated by web resources are unstructured by nature. The increase in the volume of unstructured data makes traditional relational databases unsuitable for storage. Big data requires a scalable, reliable, and efficient storage system [20]. To meet the needs of big data, NoSQL (not only Structured Query Language) databases have been adopted as the technology solution. Unlike traditional relational databases, NoSQL databases support large number of users, consistency, fault tolerance, scalability, availability, and secondary indexing [21].

10.4.5 Cloud Computing

Cloud computing is a powerful technology to perform massive-scale and complex computing. Big data analysis requires a lot of computing power, which can be provided by cloud computing at the best cost. There is no need to maintain expensive computing hardware, software, and space in cloud computing [22]. Some of the cloud computing programming frameworks are Apache Hadoop, Spark, Storm, Ceph, Hydra, and Google BigQuery. One of the most popular is Apache Hadoop, which provides a framework of open-source software for distributed storage and processing of big data using the MapReduce programming model. They were designed for computer clusters built from commodity hardware and with an assumption that hardware failures are common and should be automatically handled by the network [23]. Cloud computing allows practically unlimited and on-demand processing power [24].

10.5 Big Data in Today's World

With the amount of technology all around us, we are constantly generating a lot of data which can be analyzed. Traditionally, smaller amounts of data could be easily analyzed using statistical methods to derive useful information. Monetization or extracting the value from data can be done with the right technique so that the benefit is clearly more than the cost of analyzing the data. In the past, there was always a shortage of data, so most data analysis techniques would work well for most applications. But with the information overload of the present times, several newer data analysis techniques were developed. Large and complex datasets would require modern analytical methods including machine learning. Slowly, but surely, medical record data is going online. Multiple e-health initiatives including Google Health store data in the cloud, which makes application of big data much easier [25].

10.5.1 Big Data in Machine Learning

Machine learning is a field of AI in which computers learn without being explicitly programmed. Big data and ML are independent concepts with a weak link between them. We can apply ML on big data, as ML requires large datasets to learn better (Fig. 10.3). Many big data analytics techniques can be used in machine learning.

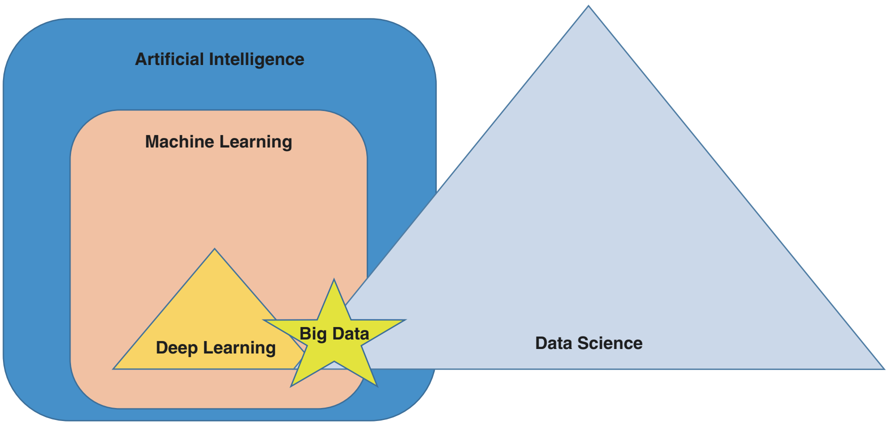


Fig. 10.3 Relation between artificial intelligence, machine learning, data science, and big data

Their common goal is learning from data, where big data focuses on handling the large volumes of data, and machine learning focuses on developing algorithms from the data. The availability of big data tools has helped rapid growth of ML and AI.

10.5.2 Big Data and Machine Learning in Healthcare

Healthcare has always been generating a lot of useful data and this has been the basis for symptoms, signs, diagnostics, and treatment. With advancement in technology, it has become easier to collect, store, and analyze this information leading to a much faster pace of medical development. This is in addition to the data about patient flow and appointments which are useful in their own way [26].

Several branches of the medical field have started using ML in a big way. Ophthalmology, radiology, dermatology, pathology, pediatrics, gynecology, oncology, endocrinology, and cardiology are some of the major ones [27].

10.5.3 Big Data in Ophthalmology

Ophthalmology uses a lot of technology and this generates a lot of data. A lot of information gathered from the EMRs starting from the patient information and history can be useful for analysis. Visual acuity, refraction, intraocular pressure, corneal thickness, and similar numerical data can be easily analyzed. There is a lot of imaging including slit-lamp imaging, fundus imaging, fluorescein angiography, optical coherence tomography which are well suited for image analysis and machine learning algorithms [28].

ML has been successfully used in detection of diabetic retinopathy, glaucoma, age-related macular degeneration, retinopathy of prematurity, retinal vascular occlusions, keratoconus, cataract, refractive errors, retinal detachment, squint, and ocular cancers. It is also useful for intraocular lens power calculation, planning squint surgeries, and planning intravitreal anti-vascular endothelial growth factor (anti-VEGF) injections. Surprisingly, analysis of fundus photographs and optical coherence tomography of the eye can even detect cognitive impairment, dementia, Alzheimer's disease, and stroke risk [27]. The use of big data and ML in ophthalmology is extensively covered in the other chapters of this book.

With major health institutions adopting EHRs, hundreds of thousands of medical records can easily be analyzed to study the relationship between risk factors and disease, refractive error and glaucoma, for example [14]. This type of analysis allows study of obscure associations which would have been missed in smaller studies. The scale of these studies is unimaginable to someone without access to big data analytics. Another study used the American Academy of Ophthalmology Intelligent Research in Sight (IRIS) Registry and the Medicare claims files to study the risk of rare events such as endophthalmitis after cataract surgery [29]. Similarly, the rarity of neuro-ophthalmic diseases makes it especially amenable to big data analysis

[30]. In addition to the rarity, the complex interconnected associations of various systemic factors make uveitis another area of ophthalmology that greatly benefits from big data analytics [31].

10.6 Big Data Security

With the large amount of data spread across many servers and the power of information, security of the data is a very important issue. Unlike tangible assets, big data cannot be physically locked up in one place. Modern data security measures have to be used to encrypt and secure the data to prevent misuse of data from data leaks or hacking, while still being accessible to analytics. Health records are even more important with regard to privacy issues. There are strict rules regarding the handling of medical records and other patient information.

10.6.1 Misuse of Data

There are rules regarding the proper protocols and permissions in handling the data, and this has to be followed. It is necessary to ensure that the proper consent for big data analysis is included in the informed consent from the patients and anonymization of data may be required before analysis.

10.6.2 Data Leaks

If the data is not secured properly, there is the possibility of it reaching the hands of unauthorized users. This would violate the rules of data security. Carelessness in handling of data and passwords leads to unwanted data leaks.

10.6.3 Hacking of Data

In spite of adequate data security, hackers may exploit some weakness in the security, or often use social engineering, to gain illegal access to data. If the data is valuable, there is a higher risk of unwanted attempts to infiltrate the data. This has to be constantly monitored and guarded against.

10.6.4 Encryption

Data has to be encrypted with the optimal level of security. It should not be unencrypted except during analysis. Passwords and encryption keys should be handled safely and used only by trusted personnel.

10.6.5 Blockchain

Blockchain is a cryptographically secure distributed database technology for storing and transmitting information. It focuses on data integrity and validation. Blockchain and big data complement each other, especially to improve the security and quality of the data. Blockchain-generated big data is secure, as it cannot be forged due to the network architecture. Blockchain-based big data is valuable, meaning it is structured, abundant, and complete, making it a perfect source for further analysis. For example, if hospital records are stored on a blockchain over a network, it eliminates the risk of data being corrupt or lost.

10.6.6 Privacy

Medical records consist of private information from patients and it is very essential to maintain that privacy. There are several laws that govern the privacy of this data and that have to be heeded. There are costs and benefits to personal information—for the subject, for the data holder, and for society as a whole [32].

10.6.7 False Conclusions

Data analytics may not always provide the right answer. When the information derived is used in healthcare, there are more repercussions in case of false conclusions. We should use big data and ML with caution to augment our judgment but not to replace it [33].

10.7 Big Data in India Versus the West

In a vast country like India, there is a wide disparity in the access and importance given to technology. The wide adoption of EMRs in the west allows medical record data to be more easily collected for analysis in spite of interoperability issue between the various EMR systems. The easy access to technology allows easier collection, storage, analysis, and dissemination of data. In India, much of the medical record data is in paper records while only a small percentage is electronic and an even smaller percentage is on the cloud. There is definitely the scope for a national-level EMR system at an affordable cost and assured security that can allow the implementation of big data and ML algorithms.

10.8 Future Trends

Data science is definitely an essential part of the future. Large amounts of data that are currently not utilized will be analyzed to yield useful information. As technology advances, the hardware required to process big data will become less expensive and accessible to all levels of business including small private healthcare. The science of big data would also be understood better, and newer algorithms and methods would be easily available in an open-source format. With a large number of smart young engineers working on these technologies, it is inevitable that big data and ML become a routine part of all business strategies.

10.9 Conclusion

There is no doubt that information is power in the new world. The volume of data generated will keep increasing every day with modern technology. By leveraging this data using modern information tools, we can derive useful insights and utilize this for progress.

References

1. The 7 V's of big data. In: Impact. https://impact.com/marketing-intelligence/7-vs-big-data/. Accessed 1 July 2020.
2. Definition of BIG DATA. https://www.merriam-webster.com/dictionary/big+data. Accessed 1 July 2020.
3. Frankenfield J. How data analytics work. In: Investopedia https://www.investopedia.com/terms/d/data-analytics.asp. Accessed 1 July 2020.
4. Data mining. Wikipedia. https://en.wikipedia.org/wiki/Data_mining. Accessed 1 July 2020.
5. Blockchain. Wikipedia. https://en.wikipedia.org/wiki/Blockchain. Accessed 1 July 2020.
6. Structured vs. unstructured data. https://www.datamation.com/big-data/structured-vs-unstructured-data.html. Accessed 1 July 2020.
7. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319:1317–8.
8. Lee I. Big data: dimensions, evolution, impacts, and challenges. Bus Horiz. 2017;60:293–303.
9. Manoj KS, Dileep KG. Effective big data management and opportunities for implementation. IGI Global; 2016.
10. Information economics. Wikipedia. https://en.wikipedia.org/wiki/Information_economics. Accessed 1 July 2020.
11. Yan J, Yu W, Zhao JL. How signaling and search costs affect information asymmetry in P2P lending: the economics of big data. Financ Innov. 2015;1:19.
12. Feijóo C, Gómez-Barroso J-L, Aggarwal S. Economics of big data. In: Bauer JM, Latzer M, editors. Handbook on the economics of the internet: Edward Elgar Publishing; 2016. p. 510–26.
13. Shah ND, Steyerberg EW, Kent DM. Big data and predictive analytics: recalibrating expectations. JAMA. 2018;320:27–8.
14. Boland MV. Big data, big challenges. Ophthalmology. 2016;123:7–8.
15. Collins B. Big data and health economics: strengths, weaknesses, opportunities and threats. PharmacoEconomics. 2016;34:101–6.
16. Harman LB, Flite CA, Bond K. Electronic health records: privacy, confidentiality, and security. AMA J Ethics. 2012;14:712–9.
17. Ozair FF, Jamshed N, Sharma A, Aggarwal P. Ethical issues in electronic health records: a general overview. Perspect Clin Res. 2015;6:73–6.
18. Akkara JD, Kuriakose A. Commentary: teleophthalmology and electronic medical records: weighing the pros and cons of unavoidable progress. Indian J Ophthalmol. 2020;68:367.
19. Dimitrov DV. Medical internet of things and big data in healthcare. Healthc Inform Res. 2016;22:156–63.

20. Siddiqa A, Karim A, Gani A. Big data storage technologies: a survey. Front Inf Technol Electron Eng. 2017;18:1040–70.
21. Cattell R. Scalable SQL and NoSQL data stores. ACM SIGMOD Rec. 2011;39:12–27.
22. Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Ullah Khan S. The rise of "big data" on cloud computing: review and open research issues. Inf Syst. 2015;47:98–115.
23. Apache Hadoop. Wikipedia. https://en.wikipedia.org/wiki/Apache_Hadoop. Accessed 1 July 2020.
24. Yang C, Huang Q, Li Z, Liu K, Hu F. Big Data and cloud computing: innovation opportunities and challenges. Int J Digit Earth. 2017;10:13–53.
25. Clark A, Ng JQ, Morlet N, Semmens JB. Big data and ophthalmic research. Surv Ophthalmol. 2016;61:443–65.
26. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA. 2013;309:1351–2.
27. Akkara JD, Kuriakose A. Role of artificial intelligence and machine learning in ophthalmology. Kerala J Ophthalmol. 2019;31:150.
28. Akkara JD, Kuriakose A. Commentary: rise of machine learning and artificial intelligence in ophthalmology. Indian J Ophthalmol. 2019;67:1009.
29. Coleman AL. How big data informs us about cataract surgery: the LXXII Edward Jackson memorial lecture. Am J Ophthalmol. 2015;160:1091–1103.e3.
30. Moss HE, Joslin CE, Rubin DS, Roth S. Big data research in neuro-ophthalmology: promises and pitfalls. J Neuroophthalmol. 2019;39:480–6.
31. Lee CS, Lee AY, Holland GN, Gelder RNV, Tufail A. Big data and uveitis. Ophthalmology. 2016;123:2273–5.
32. Acquisti A. From the economics of privacy to the economics of big data. 2014. doi: https://doi.org/10.1184/R1/6471425.v1. Accessed 2 July 2020.
33. Akkara JD, Kuriakose A. Commentary: artificial intelligence for everything: can we trust it? Indian J Ophthalmol. 2020;68:1346.

Ethics and Artificial Intelligence: The Pandora's Box

11

Parul Ichhpujani and Sahil Thakur

11.1 Introduction

Artificial intelligence (AI) has made tremendous contributions to delivery of healthcare, biomedical research and even medical education. AI applications extend to physical task support systems, robotic prostheses and mobile manipulators assisting in telemedicine. AI can learn and integrate information from large clinical datasets, and therefore can help in diagnosis and decision-making.

Although AI has proven to be very useful, it has several associated ethical challenges that must be identified and addressed. AI technology runs the risk of compromising patient preference, safety and privacy. Current policy and ethical guidelines for AI technology are still a grey zone [1]. There are no clear-cut guidelines about what constitutes 'ethical AI' and which ethical requirements and technical standards are needed.

AI applications and algorithms are programmed/coded by the computer engineers building the systems and this human element can introduce errors that may result in unforeseen outcomes.

AI and big data threaten the basic human right of privacy, as machine learning gathers and stores data from several sources. In order to establish trust between systems and the end-users, private data must be protected and any potential for biases eliminated. To ensure safe implementation of these systems, algorithmic procedures would be needed.

P. Ichhpujani (✉)
Department of Ophthalmology, Government Medical College and Hospital, Chandigarh, India

S. Thakur
Ocular Epidemiology Research Group, Singapore Eye Research Institute, Singapore

P. Ichhpujani, S. Thakur (eds.), *Artificial Intelligence and Ophthalmology*, Current Practices in Ophthalmology,
https://doi.org/10.1007/978-981-16-0634-2_11

When we talk of ethics in relation to AI, we are actually looking at two concerns: first is the concern with the moral behaviour of humans when they design, manufacture and use an artificially intelligent systems, and second one is the behaviour of machines [2].

11.2 Key Ethical Principles

Ethical principles that concern the patient care and treatment comprise non-maleficence, beneficence, respect for patient autonomy and justice [3].

11.2.1 Beneficence and Non-maleficence

Primum non nocere (First, do no harm) is the primary moral obligation of a doctor. However, this not only includes the medical aspect but also the overall quality of life of the patients. Doctors have to maintain the well-being of their patients while respecting the desires and values of every patient.

11.2.2 Respect for Autonomy

Merriam Webster dictionary defines Autonomy as “quality or state of being self-governing”.

In relation to AI, the concept of autonomy applies majorly to ‘brain–computer interfaces’ (BCI) or a ‘neural control interface’ (NCI), which involve direct communication between enhanced or wired brain and external devices, wherein brain signals are converted into commands for output devices. BCIs are commonly used as assistive devices by patients disabled due to neuromuscular disorders such as stroke, cerebral palsy or a spinal injury.

The ultimate goal is to have practical and effective BCI models. Long-term studies are needed for validating real-world use of BCIs by people with severe disabilities. Day-to-day reliability of BCI performance must be monitored and fallacies corrected so as to approach similar reliability to natural muscle-based function.

11.2.3 Justice

Both the development and distribution processes of AI are associated with justice-related issues. According to the justice principle, doctors can suggest therapeutic options that benefit their patients’ interests without subjecting them to unwarranted risk. As more and more BCIs are now available for end-users, it is imperative for the intended individual/patient to give an informed consent as regards knowledge of the design process of the device and the conflicting requirements. The ethical concern here stems from the fact that the available literature on BCI treats 'disability' as a medical issue rather than a social one, and therefore the perspective of individuals with disability is likely to be ignored.

AI systems are likely to replace human beings in some sectors and positions, which can adversely affect the human dignity especially jobs where ethical practice is pivotal such as doctors, nurses, judges and police officers. Therefore, extreme caution needs to be exercised when formulating an AI-based system for such streams. Self-improving AI systems can become so assertive than humans may encounter barriers in realising their intentions, which may lead to undesirable consequences.

11.3 The 'Black-Box' Problem: AI's Conundrum

The unknowable reasoning of 'black-box' AI, also known as opacity, stems from deep neural networks. When provided with input data, for example, such as a fundus image, a neural network trained on a large dataset can find a complex underlying pattern in the data and produce an output, such as a retinopathy classification, but is incapable of explaining how it led to the conclusive output. Additionally, since the neural network learns in ways similar to the human brain through self-teaching, when given additional data, the neural network modifies its decision-making process for a more accurate output, again without any explanation of how it accomplished the end. As the deep neural networks become increasingly autonomous with each update, the algorithms by which the technology operates become less intelligible to both the primary developers and end-users.

In case of any possible medical malpractice resulting from such technology, this opacity can lead to possible legal issues.

Legal doctrines of tort liability are not sufficient to handle medical malpractice resulting from the use of black-box AI. Modifications are needed to traditional tort law to address AI systems involved in medical malpractice.

11.3.1 Possible Legal Solution: AI Personhood

If an artificially intelligent machine is conferred 'personhood' and the machine is considered as an independent 'person' under the law then that resolves the questions that are important for analysis of vicarious liability, as the machine will be viewed as the 'principal' [4]. The AI machine will have duties and responsibilities of its own and will then be liable to be sued directly for any negligence claims.

For this, the AI system needs to be insured akin to a doctor's medical malpractice insurance so the claims will be paid out from the insurance. In that scenario, the AI system will be considered a quasi-juridical person and treated like any other physician. The users of the AI technology will be encouraged to bear some cost of the AI technology being used and fund such insurance; leading to a different form of cost-spreading that promotes fairness, as it extends beyond the technology's creators.

11.3.2 Common Enterprise Liability

The common enterprise theory of liability suggests that instead of assigning fault to an individual or entity such as an AI system, all individuals/groups involved in the use and implementation of the AI system should jointly bear some responsibility [4]. Since black-box nature of AI will make it impossible to find fault in an AI system, inference of liability shared among all relevant parties is a good way to resolve legal issues.

11.3.3 Modify the Standard of Care

Modification of the standard of care and duties of healthcare professionals using black-box AI is another solution. Healthcare professionals will need to exercise due care in procedural evaluation, implementation of black-box algorithms and validation of the algorithmic results [5]. Under this model, the onus will be on healthcare professionals for harm if they fail in adequate measures in properly evaluating the black-box AI technologies used for patient care.

Medical staff should be trained to supervise and critically evaluate AI systems and discuss the characteristics of AI such as potential errors with patients. An AI system needs to be built on public trust to achieve a desirable societal goal that AI benefits everyone.

Therefore, all individuals must also be educated regarding the basic expectations from AI to better understand the merits and demerits of AI-based health guidance.

11.4 Automation Bias

AI-based clinical decision support systems (CDSS) benefit the clinical decision-making process but possibly suffer from automation bias (AB). AB refers to complacency that sets in when a job once done by a healthcare professional is transferred to an AI program [6]. An automated AI program with virtually 100% success rate is ethically as well as clinically acceptable. But if the success rate of an AI program is lower than 100%, then it is important that the program has quality input.

For example, if we compare spotting of breast cancer cells by a pathologist versus an AI program, and if a female patient with cancer is missed by the AI program, then the point would be whether the AI program ‘learned’ from a cross section of female patients of diverse ages and races. Therefore, most ethically and clinically accurate diagnosis can be achieved when human diagnostician’s knowledge and skill is augmented by AI program with diversity of input. One must consider the fact that the AI program is likely to make different mistakes than humans; thus, combining both methods would yield the most precise diagnosis.

Automation bias can be reduced by the design of an automated system, such as reducing the prominence of the display, simplifying displayed information or couching automated assistance as supportive information rather than as directives.

11.5 Data Privacy

The age-old adage, 'Garbage in, Garbage Out', befits the data for AI too. But the question arises who is the rightful owner of the data; the patient (the source), the system (the aggregator) or the developer (the analyst of the raw data)?

Patients are not aware of how their demographic and disease-related information has been diced and spliced. When patients are informed about the option that their data would be used for research, very few opt out. Therefore, for maintaining the doctor–patient confidentiality when an AI system is going to use such data, it is the moral duty of the doctor to tell the patient about all possible ramifications of the data.

Fully informed consent and anonymity may be challenging to achieve. The vast size of the training datasets for AI algorithms for medical use makes consent an impractical concept to apply—it may be impossible to get specific informed consent from each and every patient whose data is in a particular training dataset. Although retrospective, historical data can be used in an anonymised manner for research without seeking specific consent from the individuals concerned, but there are no ethical guidelines for data that is collected prospectively for AI development. In the world of big data and AI, it is important that there are data protection laws in place that adequately protects the privacy of individuals. Data protection regulations in many countries suggest that companies should be able to alter or delete personal data on request, but no guidelines suggest what to do if the request is made after the data has been incorporated into an AI algorithm.

Another issue arises when an AI algorithm is only analysing data of small cohorts, for example when rare or orphan diseases are being considered, then the risk of identifying individual patients increases. No clear consensus exists about the fact whether the value of data about an individual's health is more important than an individual's right to withhold consent for its use for the good of all.

Non-medical researchers need to be apprised of the key principles of ethical medical research, including transparency, maintenance of confidentiality and minimising adverse effects.

11.6 Bypassing the Physician

A balance is needed between innovation and effective regulation. There is a plethora of AI health apps and chatbots, ranging from diet and exercise guides, health assessment tools and those that help to improve medication compliance and analysis of an individual's bodily parameters collected via wearable sensors. Such products that provide autonomous diagnosis and possible management options do not have a 'licence to practice' or are approved by any regulatory body [7]. Since a 'user agreement' is not the same as 'informed consent', there is a valid concern as regards indemnity. Some patients are likely to be at particular risk from 'bad advice' from digitised systems such as the psychiatric patients, and younger and elderly individuals. Eventually, it would be the clinicians who will have to deal with the aftermath of bad advice from an AI system.

Who decides which AI systems to watch closely? AI programs are dynamic but the current regulatory environment approves or disapproves people, medicines, devices, procedures or institutions in a static context [8]. Moreover, frequent updates of the software make it even more difficult for individuals to follow what terms of service they have agreed to.

11.7 Cybersecurity

Cybersecurity must also be considered when addressing legal challenges to the use of AI in healthcare. Most of the future healthcare-related services and products will operate using the Internet of Things (IoT), but the underlying infrastructure of IoT is vulnerable to cyberthreats. Health sector has several possible targets such as wearables, wireless smart pills, hospital servers, diagnostic tools and medical devices. These can be infected with software viruses, Trojan horses or worms that can compromise patients' privacy. Such infected algorithms or corrupted data can result in incorrect treatment recommendations. Therefore, an internationally enforceable, large-scale regulatory body for cybersecurity that ensures a high level of cybersecurity and resilience across nations is needed.

Currently, many new guidelines on appropriate use of AI are being issued by both governmental and professional institutions, but these need to be unified under universally accepted rules and limitations.

References

1. Luxton DD. Recommendations for the ethical use and design of artificial intelligent care providers. Artif Intell Med. 2014;62(1):1–10.
2. Johnson SLJ. AI, machine learning, and ethics in health care. J Leg Med. 2019;39(4):427–41.
3. Beil M, Proft I, van Heerden D, Sviri S, van Heerden PV. Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Med Exp. 2019;7(1):70.
4. Vladeck DC. Machines without principles: liability rules and artificial intelligence. Wash Law Rev. 2014;89(1):117–50.
5. Price WN. Medical malpractice and black-box medicine. Big data, health law, and bioethics. Cambridge, UK: Cambridge University Press; 2018.
6. Stumpe M, Peng L. Assisting pathologists in detecting cancer with deep learning. Google AI Blog. March 3, 2017. http://ai.googleblog.com/2017/03/assisting-pathologists-in-detecting.html. Accessed 3 Nov 2020.
7. Gerke S, Minssen T, Yu H, Cohen IG. Ethical and legal issues of ingestible electronic sensors. Nat Electron. 2019;2:329–34.
8. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. Big Data Soc. 2016;2:1–21.