Reinforcement learning in ophthalmology: potential applications and challenges to implementation

Siddharth Nath, Edward Korot, Dun Jack Fu, Gongyu Zhang, Kapil Mishra, Aaron Y Lee, Pearse A Keane

Reinforcement learning is a subtype of machine learning in which a virtual agent, functioning within a set of predefined rules, aims to maximise a specified outcome or reward. This agent can consider multiple variables and many parallel actions at once to optimise its reward, thereby solving complex, sequential problems. Clinical decision making requires physicians to optimise patient outcomes within a set practice framework and, thus, presents considerable opportunity for the implementation of reinforcement learning-driven solutions. We provide an overview of reinforcement learning, and focus on potential applications within ophthalmology. We also explore the challenges associated with development and implementation of reinforcement learning solutions and discuss possible approaches to address them.

Lancet Digit Health 2022

Published Online
July 26, 2022
https://doi.org/10.1016/S2589-7500(22)00128-5

Department of Ophthalmology and Visual Sciences, McGill University, Montreal, QC, Canada (S Nath PhD); **National Institute for Health Research, Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust, UCL Institute of Ophthalmology, London, UK** (S Nath, E Korot MD, D J Fu PhD, G Zhang MSc, Prof P A Keane MD); **Byers Eye Institute, Stanford University, Palo Alto, CA, USA** (E Korot, K Mishra MD); **Department of Ophthalmology, University of Washington, Seattle, WA, USA** (D J Fu, A Y Lee MD)

Correspondence to:
Prof Pearse A Keane, National Institute for Health Research, Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust, UCL Institute of Ophthalmology, London EC1V 2PD, UK
pearse.keane1@nhs.net

Introduction

Over the past decade, machine learning has ushered in a new era of medical research—one centred on big data and computational solutions to previously intractable health-care issues.[1,2] Machine learning has been applied to an array of health-care problems, from evaluation of diagnostic radiology images to identification of clinical trial participants.[3,4] Within ophthalmology specifically, machine learning has found extensive use in evaluating ophthalmic imaging—from the identification of glaucoma medications from photographs of eye drop bottles, to subspecialty retina practice, where deep learning has shown promise in evaluating optical coherence tomography scans to determine when to make referrals to specialists.[5,6]

All these implementations centre on a form of machine learning called supervised learning, in which an algorithm is given both a task and what are known as ground truth labels. For instance, in the application of deep learning to the identification of glaucoma medications from photographs of eye drop bottles, the algorithm is provided with a training set of data in which photographs of eye drop bottles are labelled with the associated medication. However, clinical decision making is not always as linear or univariate as in this example, and often, clinicians are required to make treatment decisions on scarce and incomplete information. To address such health-care problems, researchers have turned to reinforcement learning, a subset of machine learning in which a virtual intelligent agent carries out actions according to a predefined set of rules to optimise a specified reward.

Reinforcement learning agents make decisions using a policy, which is a dynamic strategy intended to optimise the reward outcome. As the reward outcome is maximised, the agent's decision making is reinforced until an optimal policy is identified, a process that often requires trial-and-error and many policy iterations. Computation of these iterations is typically accomplished by solving a Markov decision process, involving variables for the state, action, and reward function, as well as a discount factor, for reducing the value of increasingly temporal outcomes (figure 1, 2).[7] Reinforcement learning is particularly well suited to complex tasks analogous to clinical decision making because agents are capable of taking actions on incomplete training and can perform sequential steps to optimise the reward.[8–10] Although supervised learning pipelines might also be able to complete tasks from incomplete data by imputation, such substitutions are not always accurate and might lead to incorrect decision making.[11,12] Moreover, rewards for reinforcement learning agents can be sparsely defined and, given the sequential nature of their interaction with the preset environment, reinforcement learning algorithms can address the problem of correlating immediate action with delayed outcomes. For example, a reinforcement learning agent designed to optimise the reward of completing a video game quest would not necessarily know whether immediate actions

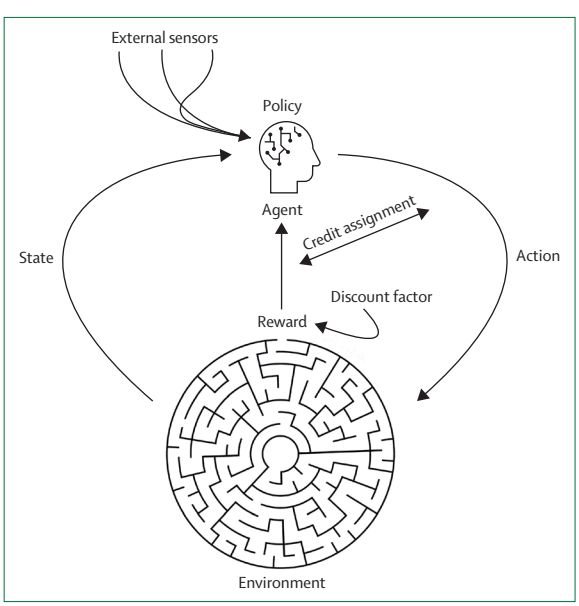

Figure 1: **Schematic representation of the reinforcement learning process**
Reinforcement learning involves actions taken by an agent on the basis of a predefined policy, continuously adjusted by input from the environment state in an effort to maximise reward. Figure created using BioRender.com.

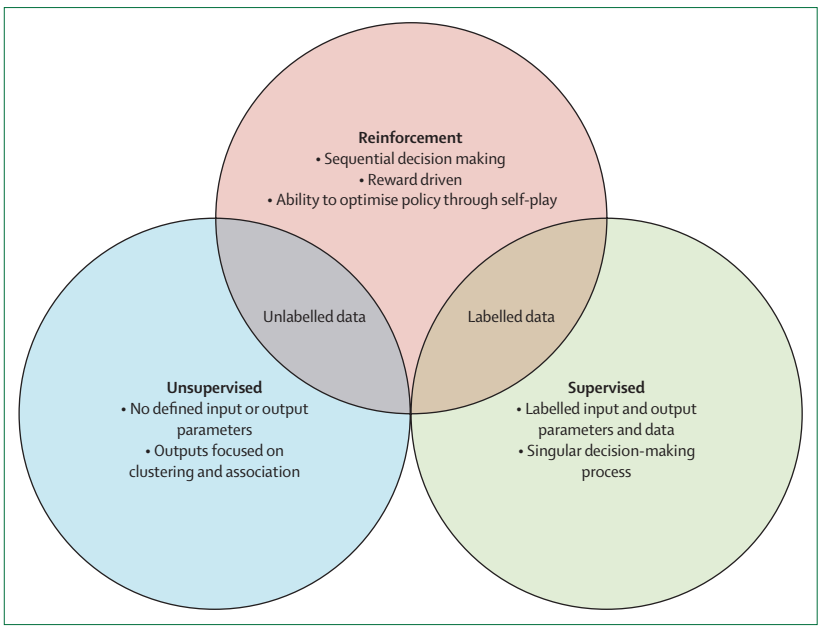

***Figure 2*: Comparison of reinforcement learning, supervised learning, and unsupervised learning**
A Venn diagram outlining the key features of the three major subtypes of machine learning.

bring it closer to this sparsely defined reward. However, sequential actions and self-play would allow it to optimise its completion of the required task.

Outside of medicine, reinforcement learning has been used in tasks stereotypically considered to be within the realm of artificial intelligence in its mainstream definition. For instance, reinforcement learning algorithms have been trained to play video games (and beat their human competitors), develop frameworks for autonomous vehicles, and create lifelike robots.[13–15] In fact, researchers in machine learning at DeepMind (London, UK) have recently posited that reinforcement learning itself might be sufficient for creating an artificial general intelligence.[16] In some ways then, reinforcement learning might be considered to be classical conditioning of an artificial general intelligence, whereby optimisation of a reward drives the completion of a complex processing task.

Within medicine, reinforcement learning has been used in areas where clinical decisions are on a long-temporal scale, or where supervision is not possible. For instance, reinforcement learning algorithms have been developed to manage patients with sepsis in the intensive care unit, titrate propofol infusion during anaesthesia, optimise breast cancer screening, treat patients with epilepsy, manage patients with diabetes, and monitor and treat patients with HIV infection.[17–22] Notably, applications of reinforcement learning to ophthalmology are unexplored. Here, we provide an overview of potential uses of reinforcement learning in contemporary ophthalmology and discuss the challenges associated with implementation and actionable solutions to move the field forward.

Applications of reinforcement learning in ophthalmology

Chronic disease monitoring and treatment

Perhaps the most apparent and readily scalable implementation of reinforcement learning in ophthalmology is in the management of chronic ocular diseases. A substantial portion of patient volume in modern ophthalmology clinics is devoted to management of chronic illnesses, such as diabetic retinopathy, age-related macular degeneration, uveitis, and glaucoma.[23] These diseases also carry a substantial economic burden and require intensive cognitive and time effort from clinicians to manage.[24] For these reasons, reinforcement learning algorithms offer a potential solution to streamline patient care and improve the equity of valuable clinical resources. For example, in the care of age-related macular degeneration and diabetic retinopathy, a reinforcement learning algorithm could be developed with a reward of a decrease in subretinal or intraretinal fluid on optical coherence tomography, or improvement in functional outcomes such as vision. In these instances there could be three possible decisions as outcomes: inject or do not inject with anti-vascular endothelial growth factor (VEGF); choice of medication if injecting; and when the clinician should see the patient in follow-up (figure 3). Such an algorithm might not only optimise treatment of patients beyond conventional treat-and-extend protocols, but it might also tailor treatment to each individual patient's disease process. Similarly, in the setting of glaucoma, a decrease in intraocular pressure or, more specifically, a set intraocular pressure value, could be outlined as a reward for a reinforcement learning algorithm, with actions including the addition of different topical therapies and, possibly, considerations for micro-invasive glaucoma surgery or traditional filtering procedures.

Analysis of ophthalmic imaging

Contemporary ophthalmology involves the use of multiple imaging modalities, such as optical coherence tomography, fundus photography, fundus autofluorescence, fluorescein angiography, and visual fields, and the breadth of data captured continues to expand. Reinforcement learning solutions might be implemented to assist ophthalmologists with image segmentation to improve efficiency in the clinic by highlighting relevant pathology and disease progression.[25] Although supervised learning models have been developed to assist with ophthalmic image segmentation, such implementations require large volumes of training data with pixel-level ground truth labels. Moreover, accurate labelling of training datasets requires substantial human clinician hours and might not be scalable in real-world settings. However, reinforcement learning models have shown potential ability to conduct image segmentation with far fewer pixel-level labels than their supervised learning counterparts. In fact, in an autonomous driving

 Published online July 26, 2022 https://doi.org/10.1016/S2589-7500(22)00128-5

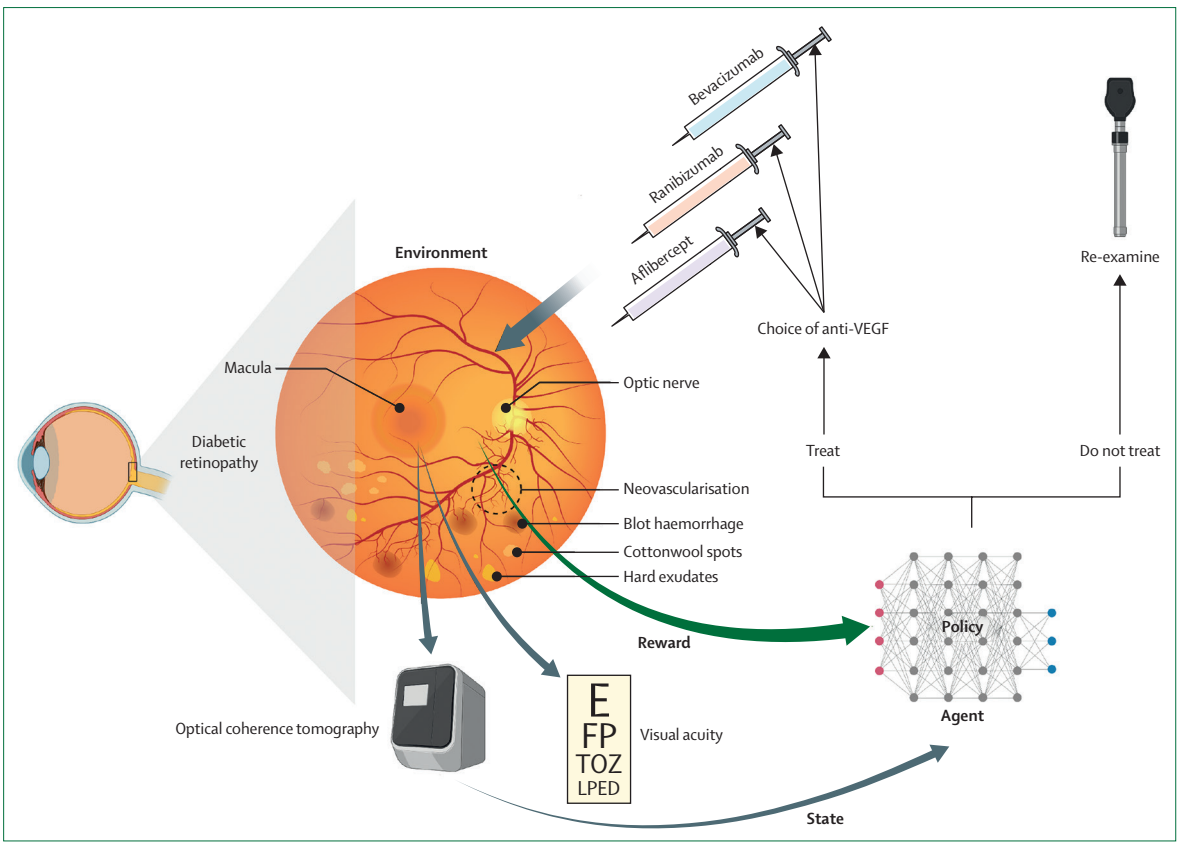

***Figure 3*: Schematic representation of a reinforcement learning policy for diabetic retinopathy**
A reinforcement learning policy for treatment of diabetic retinopathy incorporating patient optical coherence tomography data and additional variables in determining when to treat or extend the treatment time. Figure created using BioRender.com. VEGF=vascular endothelial growth factor.

implementation, a reinforcement learning model required 30% fewer labels than a conventional supervised learning model.[26] Such leaps might accelerate the development of intelligent image-based diagnostic and monitoring systems for ophthalmology and expedite their real-world usage.

Intelligent ophthalmic surgery systems

Reinforcement learning also offers ophthalmologists potential for enhancements in all stages of ophthalmic surgery, from preoperative planning to postoperative management. Similar to reinforcement learning algorithms managing chronic ocular conditions, an automated model could aid in identifying the optimal time to pursue surgery on the basis of previous patient outcomes. Intraoperatively, surgical decision making is complex, multifactorial, and not always amenable to supervision. Reinforcement learning models might have a role in automating ophthalmic surgery, from simple warning mechanisms to avoid intraoperative complications to completely autonomous robotic surgery.

Intraocular surgery, usually taking place under a surgical microscope, affords the ability to collect large volumes of surgical video footage from a relatively stable viewpoint. Moreover, newer microscope systems incorporate three-dimensional (3D) optics, allowing for collection of 3D video, with excellent delineation of tissue depth intraoperatively. Reinforcement learning models could be trained by use of such 3D surgical video databases and the movement of instruments encoded similar to a video game, with specific stages (eg, capsulorhexis or epiretinal membrane peeling) encoded as rewards. Such systems could be developed to optimise policies that result in favourable surgical outcomes and then built into modern operating microscopes to show an indicator light or audible warning when a surgeon moves an instrument in a manner that differs from the optimal policy. Additional variables, such as data taken from intraoperative real-time optical coherence tomography, could also be incorporated into these algorithms to provide high-fidelity policies, based not only on movement of instruments and surgical steps, but also on patient physiology. Unlike supervised learning implementations, such systems would require far fewer labelled data and be able to complete sequential steps within one algorithm instead of multiple algorithms. Moreover, such systems offer the potential for improving training of junior doctors and fellows to reduce early surgical complications, as well as of novice surgeons in

For more on **Orbis** see https://www.orbis.org/

low-resource settings, and would be a boon for global health organisations such as Orbis to greatly expand their reach and mission. Not with the intent of replacing surgeons, reinforcement learning systems would be used to augment the surgeon's experience and procedural safety profile. However, further in the future, with concomitant advancements in robotics technology, such reinforcement learning systems could facilitate the use of intelligent surgery systems capable of highly delicate operations independent of human input, with reproducible, consistent outcomes.

Challenges to implementation of reinforcement learning in ophthalmology

Although reinforcement learning offers powerful capabilities for ophthalmologists, both in the clinic and the operating room, implementation of solutions is not without challenges. The potential hurdles of integration of reinforcement learning solutions into clinical practice have been explored extensively in the literature and carry important lessons for both ophthalmologists and for researchers who are working to develop use cases.[7,10,27]

The absence of separate training and evaluation environments in the real world

One of the most substantial limitations of reinforcement learning for applications in health care is the inability to explore potential policies in the real environment. It is unethical to pilot untested reinforcement policies that could lead to patient harm in the clinical realm, and thus, researchers are limited by training of reinforcement learning algorithms using previous observational data only. To minimise the potential for harm to patients, reinforcement learning models could be trained in a graded fashion, similar to modern medical training. In such a scenario, a surgical model, for example, would first begin operating on simulated eyes, either in the operating room or in virtual reality. Next, the model could advance to cadaveric eyes and then to completing specific steps of a surgery on a live patient under supervision. Completing all this testing could be requisite for advancing to finally completing full procedures on live patients. Such training could be coupled with off-policy evaluation on a validation dataset and would offer the safety of a largely patient-free environment, but with the caveat that not all policy actions might be completely tested. Importantly, it should be considered that such approaches might incorporate bias into algorithm policies that are inherent within the training and validation datasets, and thus limit the creativity of the algorithm in optimising its reward.[28] Conceivably then, although a reinforcement learning algorithm could optimise treatment of patients with age-related macular degeneration or movement of surgical instruments, it would be unable to develop an entirely new treatment strategy or surgical manoeuvre. Careful consideration of biases within datasets is also essential to ensuring algorithms produce effective and equitable policies.

Inherent delays in real-world systems

Reinforcement learning algorithms are also susceptible to system delays, so-called because most real systems involve delays in sensing, actuation, or reward feedback because the manifestation of an action's effect might be on a long temporal scale. For example, within ophthalmology, this is most easily conceptualised in the management of patients with age-related macular degeneration, for which a reinforcement learning algorithm might make a decision to treat a patient with anti-VEGF but does not realise its reward until the next optical coherence tomography image is viewed at the following appointment. To address this problem, developers of the algorithm could consider assigning weighted credit to past actions undertaken by the model that are determined to be useful for predicting the future, such as previous optical coherence tomography images showing a reduction in subretinal or intraretinal fluid, associated with disease improvement.[29] Reinforcement learning algorithms are also limited by the partial observability and non-stationarity of the clinical environment. Within ophthalmology, clinicians are limited by the patient data immediately available. In patients with age-related macular degeneration, those data might be only an optical coherence tomography image and fundus photo; in patients with glaucoma, those data might be an optical coherence tomography image and intraocular pressure; and intraoperatively, those data might be only the surgical field. All these environments are also dynamic; that is, disease processes are continuous and multifactorial, and surgical procedures are non-stationary. To address this problem, reinforcement learning algorithms could be given access to additional sensors with which to guide decision making. For instance, in addition to a patient's intraocular pressure in glaucoma management, the algorithm might have access to all intraocular pressure values in that specific patient's demographic, or the algorithm might be trained with all surgical video footage beyond the specific surgery the algorithm has been designed for, both of which could be used as sensors to inform iterative play within the confines of the specific disease or procedural environment. Crucially, this access to and incorporation of additional sensors might add to another problem with the implementation of reinforcement learning—one of appropriate credit and reward assignment. Although in many environments in ophthalmology a reward is easily defined as a quantifiable medical outcome (eg, lower intraocular pressure in patients with glaucoma), this is not always the case. Within surgical reinforcement learning algorithms, careful evaluation of all intraoperative movements, with appropriate credit assignment given with successful completion of specific steps, requires substantial surgeon investment of time, and might be heterogeneous across

surgeons. Assignment of credit in association with memory, as described above, is also an important consideration because the prognostic factors of many chronic ocular diseases are poorly defined and difficult to implement, even in routine clinical practice.

Exploration versus exploitation

In addition to system delays, reinforcement learning agents are also affected temporally by their inherent focus on optimising reward. In this pursuit, agents might often engage in actions that obtain reward in the short term, without consideration of long-term implications to the environment. For instance, a reinforcement learning policy implemented for the management of diabetic retinopathy might choose to always inject patients with an anti-VEGF medication, to optimise a short-term reward of reduction in subretinal fluid on optical coherence tomography as learned from previous iterations. However, such policies might not recognise when disease control has been achieved and thus continuously expose patients to the potential for iatrogenic injury and achieve a state of diminishing returns. To minimise exploitation-driven actions by a reinforcement learning agent, rewards could be more specifically defined to capture the multifactorial nature of disease amelioration. In the instance of diabetic retinopathy, such rewards could encapsulate not only reduction of subretinal fluid on imaging, but also improvement or stability in visual acuity and minimisation of complications associated with therapy. Assigning corresponding weights to these factors would promote decision making that favours exploration to a long-term reward versus exploitation of short-term variables.

Explainability

Implementation of reinforcement learning algorithms is also limited by the ability to provide system operators with explainable policies. Because reinforcement learning algorithms can perform sequential decision making, there is an element of a black box encountered with their use. Ophthalmologists implementing reinforcement learning algorithms in routine patient care will need to be well informed of the nature of specific policies used by the algorithm, not only to improve the integration of the algorithms into clinical practice, but also to allow clinicians to act as a safeguard should an algorithm suggest a potentially dangerous action in its endeavour to maximise reward. Finally, and perhaps most importantly, with explainability challenges come difficulties in evaluating the safety and efficacy of reinforcement learning policies. Because there is no equivalent or standard-of-care algorithm to compare reinforcement learning algorithms with, and observers are not aware of real-time inputs, it is difficult to evaluate reinforcement learning algorithms without randomised trials. Thus, implementation of reinforcement learning solutions into clinical practice requires rigorous evaluation studies, and algorithms should be piloted extensively against existing human physicians to appropriately assess clinical performance.

Search strategy and selection criteria

We identified relevant articles by searching the MEDLINE database using PubMed on Dec 2, 2021, for articles published between Jan 1, 2016, and Dec 1, 2021, with the following search terms: ("Reinforcement learning" [all fields]) AND ("Ophthalmology" [all fields]). We searched the Google Scholar database on Dec 2, 2021, using the keywords "Reinforcement learning" in the all fields setting. The search was re-run and updated with the same parameters on May 1, 2022. Database searching was complemented by review of the references of relevant articles. No language restrictions were used in any of our searches.

Conclusions

As machine learning solutions move from the research realm into routine clinical practice, ophthalmology stands to experience another revolution.[30] Reinforcement learning offers ophthalmologists the potential to optimise chronic disease management, increase the equity of valuable clinical resources, improve surgical outcomes, and facilitate the training of the next generation of surgeons. Nonetheless, reinforcement learning is not without its limitations, and careful consideration of the challenges associated with its implementation will ensure that we are able to reinvent not only the diagnosis of ocular diseases, but also their medical and surgical management.

Contributors

SN, EK, and PAK conceptualised the research. DJF, GZ, KM, and AYL provided subject matter expertise and assisted in writing the manuscript. All authors reviewed and approved the final version of the manuscript before submission.

Declaration of interests

EK has acted as a consultant for Genentech and Google Health, and is an equity owner in Reti Health. PAK has acted as a consultant for DeepMind, Roche, Novartis, Apellis, and BitFount; is an equity owner in Big Picture Medical; has received speaker fees from Heidelberg Engineering, Topcon, Allergan, and Bayer; and is supported by a Moorfields Eye Charity Career Development Award (R190028A) and a UK Research & Innovation Future Leaders Fellowship (MR/T019050/1). AYL is a special government US Food and Drug Administration employee, and has received grants from Santen, Carl Zeiss Meditec, and Novartis, as well as personal fees from Genentech, Topcon, and Verana Health, outside of the submitted work. All other authors declare no competing interests.

References

1 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019; **380:** 1347–58.

2 No authors listed. Ascent of machine learning in medicine. *Nat Mater* 2019; **18:** 407.

3 Tang Y-X, Tang Y-B, Peng Y, et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Digit Med* 2020; **3:** 70.

4 Beck JT, Vinegra M, Dankwa-Mullan I, et al. Cognitive technology addressing optimal cancer clinical trial matching and protocol feasibility in a community cancer practice. *J Clin Oncol* 2017; **35** (suppl)**:** 6501.

5 Tran TT, Richardson AJW, Chen VM, Lin KY. Fast and accurate ophthalmic medication bottle identification using deep learning on a smartphone device. *Ophthalmol Glaucoma* 2022; **5:** 188–94.

6 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:** 1342–50.

7 Riachi E, Mamdani M, Fralick M, Rudzicz F. Challenges for reinforcement learning in healthcare. *arXiv* 2021; published online March 9. https://doi.org/10.48550/arXiv.2103.05612 (preprint).

8 Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature* 2015; **518:** 529–33.

9 Littman ML. Reinforcement learning improves behaviour from evaluative feedback. *Nature* 2015; **521:** 445–51.

10 Yu C, Liu J, Nemati S. Reinforcement learning in healthcare: a survey. *arXiv* 2019; published online Aug 22. https://doi.org/10.48550/arXiv.1908.08796 (preprint).

11 van Buuren S. Flexible imputation of missing data, 2nd edn. New York, NY: CRC Press, 2018.

12 Josse J, Prost N, Scornet E, Varoquaux G. On the consistency of supervised learning with missing values. *arXiv* 2019; published online Feb 19. https://doi.org/10.48550/arXiv.1902.06931 (preprint).

13 Mnih V, Kavukcuoglu K, Silver D, et al. Playing Atari with deep reinforcement learning. *arXiv* 2013; published online Dec 19. https://doi.org/10.485550/arXiv.1312.5602 (preprint).

14 Kiran BR, Sobh I, Talpaert V, et al. Deep reinforcement learning for autonomous driving: a survey. *IEEE Trans Intell Transp Syst* 2021; **23:** 4909–26.

15 Singh B, Kumar R, Singh VP. Reinforcement learning in robotic applications: a comprehensive survey. *Artif Intell Rev* 2021; **55:** 945–90.

16 Silver D, Singh S, Precup D, Sutton RS. Reward is enough. *Artif Intell* 2021; **299:** 103535.

17 Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* 2018; **24:** 1716–20.

18 Lowery C, Faisal AA. Towards efficient, personalized anesthesia using continuous reinforcement learning for propofol infusion control. 6th International IEEE/EMBS Conference on Neural Engineering (NER); Nov 6–8, 2013.

19 Guez A, Vincent RD, Avoli M, Pineau J. Adaptive treatment of epilepsy via batch-mode reinforcement learning. https://www.aaai.org/Papers/IAAI/2008/IAAI08-008.pdf (accessed May 1, 2022).

20 Bothe MK, Dickens L, Reichel K, et al. The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. *Expert Rev Med Devices* 2013; **10:** 661–73.

21 Yu C, Dong Y, Liu J, Ren G. Incorporating causal factors into reinforcement learning for dynamic treatment regimes in HIV. *BMC Med Inform Decis Mak* 2019; **19** (suppl 2)**:** 60.

22 Yala A, Mikhael PG, Lehman C, et al. Optimizing risk-based breast cancer screening policies with reinforcement learning. *Nat Med* 2022; **28:** 136–43.

23 Lee PP, Feldman ZW, Ostermann J, Brown DS, Sloan FA. Longitudinal prevalence of major eye diseases. *Arch Ophthalmol* 2003; **121:** 1303–10.

24 Marques AP, Ramke J, Cairns J, et al. Global economic productivity losses from vision impairment and blindness. *EClinicalMedicine* 2021; **35:** 100852.

25 Coronato A, Naeem M, De Pietro G, Paragliola G. Reinforcement learning for intelligent healthcare applications: a survey. *Artif Intell Med* 2020; **109:** 101964.

26 Casanova A, Pinheiro PO, Rostamzadeh N, Pal CJ. Reinforced active learning for image segmentation. *arXiv* 2020; published online Feb 16. https://doi.org/10.48550/arXiv.2002.06583 (preprint).

27 Dulac-Arnold G, Levine N, Mankowitz DJ, Li J. An empirical investigation of the challenges of real-world reinforcement learning. *arXiv* 2020; published online March 24. https://doi.org/10.48550/arXiv.2003.11881 (preprint).

28 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019; **366:** 447–53.

29 Hung C-C, Lillicrap T, Abramson J, et al. Optimizing agent behavior over long time scales by transporting value. *Nat Commun* 2019; **10:** 5223.

30 Keane P, Topol E. Reinventing the eye exam. *Lancet* 2019; **394:** 2141.