Machine learning has arrived!

Aaron Lee,[1] Paul Taylor,[2] Jayashree Kalpathy-Cramer,[3] Adnan Tufail[4]

1 University of Washington, Department of Ophthalmology, Seattle, Washington.

2 Centre for Health Informatics and Multiprofessional Education, Institute of Health Informatics, University College London, London, United Kingdom.

3. Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

4. Moorfields Biomedical Research Centre, Moorfields Eye Hospital, London, United Kingdom

Corresponding Author

Adnan Tufail

Moorfields Eye Hospital NHS Trust

162 City Road

London SE1 4TT

Adnan.Tufail@moorfields. nhs.uk

Keywords

Machine learning, Deep learning, image analysis, automated, diabetic retinopathy, screening, retinopathy of prematurity, glaucoma, cataract, visual fields, age-related macular degeneration

Text word count 1554

Over the last quarter of a century machine learning has become one of the most important parts of the information technology revolution impacting our lives. Artificial intelligence, the field from which machine learning emerged, attempts to make machines that think the way humans think. Machine learning is the subfield of artificial intelligence that "gives computers the ability to learn without being explicitly programmed", a definition attributed to Arthur Samuel, who applied machine learning to play checkers, starting a tradition of using games as a testbed for machine learning algorithms.[1]

Initially research in artificial intelligence tackled problems that are intellectually difficult for human beings but relatively straightforward for computers, problems which can be addressed by an algorithm expressed in rules identified and articulated by the programmer. These could be very successful, for example IBM's Deep Blue defeated Gary Kasparov, the then world chess champion in 1996.[2] A greater challenge in artificial intelligence is solving the tasks that are easy for people to perform but hard for people to describe, i.e. the ones we solve intuitively, such as understanding language or recognising faces.

In machine learning the computer is programmed to optimize a performance criterion using data from past experience, typically a set of example cases with a known classification. We need such learning in cases where we cannot directly or easily express how a problem could be solved algorithmically but we can identify sets of examples that illustrate the solution. Such problems can be incredibly challenging, as in the example of speech recognition, something that humans perform seemingly without difficulty, but without being able to explain how it is done. People of different ages and genders or with different accents pronounce the same word differently, indeed the same person, having a cold or in a temper, will pronounce a word in a very different way. Huge resources were invested into research in speech recognition with relatively little impact until, in around 2009, scientists worked out how to take advantage of the huge amounts of processing power and data becoming available to build machine learning algorithms capable of tackling these more complex problems. This technology, known as deep learning, was rapidly incorporated into speech recognition software by Google and, for example, is the reason voice control is a practical way to use an android phone.

Although machine learning dates from the early days of artificial intelligence in the late 1950s it underwent a first resurgence when the concept of data mining began to take off about 20 years ago. Data mining algorithms look for patterns in information. Machine learning does the same thing, but then goes one step further – the program changes its behaviour based on what it learns.

There are different approaches to machine learning. It is worth noting that statistical approaches such as regression that derive an equation from a set of data points are forms of machine learning algorithms. Other, less traditional approaches include artificial neural networks which are based on an analogy with the operation of the human brain. These are built up from networks of simple computational units, neurons, arranged in layers. A particularly powerful approach to building neural networks – called deep learning – uses a form of artificial neural network to identify low level features in the data, features that can then be used as the basis for a classification learned at higher levels of the network. Deep learning is made possible, by cheap hardware to perform large numbers of operations in parallel in the form of graphical processing units - chips originally developed for the gaming industry, together with huge datasets and open-source software frameworks. It has now triggered a second resurgence of interest in machine learning.

There are many successful applications of machine learning in domains outside of medicine. There are commercially available systems for recognizing speech, handwriting and for machine translation. Retail companies analyse their past sales data to learn their customers' behaviour to improve customer relations, and many web-based companies use machine learning to power their recommendation engines. Face recognition software is used in airport passport control and the development of self-driving cars requires the recognition of a wide variety of possible objects in the complex and changing context of the immediate driving environment. But how can developments of machine learning impact medicine, and more specifically ophthalmology? Excitingly, machine learning is already here, and is likely to make an even greater impact in many areas of clinical ophthalmology and eye research in the near future.

The most obvious areas for the application of machine learning are in health settings where the data to be analysed are complex, but the outcomes are simple and well-defined and the number of patients to process is large. Image-based screening programmes are an obvious target.

Automated retinal image analysis systems (ARIAS) that detect diabetic retinopathy on digital retinal images are already in deployment. Although a number are commercially available[3] it is essential to have independent validation of how they perform in detecting retinopathy that needs ophthalmology opinion, without over-referring those that do not have sight-threatening retinopathy. Independent validation is very important for large scale deployment and this has recently been undertaken in a study in the UK on over 100,000 retinal images that have never been used to train any ARIAS. The study by Tufail and coworkers showed that two of the ARIAS tested are safe and cost-effective for use in a diabetic screening programme.[4,5] Competitions such as Kaggle,[6] where teams compete to try and develop an software that has the best sensitivity and specificity are accelerating the interest in the area. Recently scientists at Google Health created a dataset of 128,000 images and used it to train a deep learning network for diabetic retinopathy .[7-9] Although the results are very exciting the system is still to be given regulatory approval or have an independent assessment on a diverse test as per the UK study.

Deep learning has been applied in software that analyses optical coherence tomography (OCT) to differentiate normal from AMD eyes. The recent publication by Lee and coworkers demonstrated high sensitivity and specificity to detect AMD using a big data approach of mining 2.6 million OCT images, linking it to EMR data, to finds the tens of thousands of patient's images required to train the algorithm. [10]

Another data rich subspecialty that may benefit from machine learning approaches to is glaucoma, where it has been applied to analyse visual fields,[11] and identify glaucomatous disc cupping.[12] This could help screening for glaucoma and to support ophthalmologists assessing the progression of the disease in visual fields and disc.

Machine learning has already been used in the diagnosis of cataract, and improving lens implant power selection before cataract surgery using radial based formula[13] or support vector machines [14] but could be applied to other anterior segment areas such as the analysis of corneal topography. It has also been used to develop diagnostic tools for retinopathy of prematurity (ROP).[15] Deploying these algorithms, for example integrated within camera systems, has great potential to impact screening especially in low resource settings. Given how visual the practice of ophthalmology is, and how in many areas diagnosis is made on an image, machine learning could have an impact across the specialty.

With the increasing use of EMR systems, large repositories of both EMR data and images will be collected that will allow not just the training of machine learning algorithms but also the development of predictive analytics that help model outcomes on an individual level. As the datasets and standardised and enlarge we may be able to apply these techniques to less common disorders such as uveitis.[16]

Machine learning algorithms, especially deep learning based methods require large amounts of well annotated data ("ground truth") where the annotations are typically generated by expert readers. The EMR is a common source of the annotations. However, as reported, the variability and bias of expert readings can be challenging for these algorithms and recent publications have argued for the need for obtaining readings from multiple raters per image.[8] Modelling the user to account for these biases and variability can be an important aspect of machine learning.

We argue that, in parallel with these technical developments, the thinking around the validation of emerging machine learning algorithms before clinical deployment, also needs to advance. Researchers developing machine learning algorithms generally split the available data into separate sets to be used for training, testing and validation. The data may however all be drawn from a single site or a narrow range of sites and the results may not generalise to other populations or to all imaging devices and protocols. We argue that before clinical deployment is considered, an independent study, or one with independent oversight (like clinical drug trials) should be required of sufficient size to be confident about detecting clinically important but less common events. Also should learning algorithms be allowed to continue to learn after independent approval when exposed to real life data which may alter its performance possibly for the better but in an unchecked way? Regulatory authorities will need to adapt to these new challenges.

The convergence of the availability of large ophthalmology datasets, improvement in machine learning algorithms and further improvement in computer processing power has resulted in an explosion of interest in machine learning applications for ophthalmology. But this is not the future, the future has already begun.

References

1. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev.* 1959;3(3):210-229.
2. Research DBI. Deep Blue. http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/. Accessed 08/04/2017, 2017.
3. Sim DA, Keane PA, Tufail A, Egan CA, Aiello LP, Silva PS. Automated retinal image analysis for diabetic retinopathy in telemedicine. *Current diabetes reports.* 2015;15(3):577.
4. Tufail A, Rudisill C, Egan C, et al. Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders. *Ophthalmology.* 2017;124(3):343-351.
5. Tufail A, Kapetanakis VV, Salas-Vega S, et al. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health technology assessment.* 2016;20(92):1-72.
6. Diabetic Retinopathy Detection. https://www.kaggle.com/c/diabetic-retinopathy-detection. Accessed 03/04/2017.
7. Abramoff MD, Lou Y, Erginay A, et al. Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning. *Investigative ophthalmology & visual science.* 2016;57(13):5200-5206.

8. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA.* 2016;316(22):2402-2410.
9. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology.* 2017.
10. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus Age-related Macular Degeneration. *bioRxiv.* 2016.
11. Asaoka R, Murata H, Iwase A, Araie M. Detecting Preperimetric Glaucoma with Standard Automated Perimetry Using a Deep Learning Classifier. *Ophthalmology.* 2016;123(9):1974-1980.
12. Zhu H, Poostchi A, Vernon SA, Crabb DP. Detecting abnormality in optic nerve head images using a feature extraction analysis. *Biomedical optics express.* 2014;5(7):2215-2230.
13. Hill-RBF Calculator- IOL Power Calculations for Cataract Surgery. http://rbfcalculator.com/. Accessed 20 March 2017.
14. Yarmahmoodi M, Arabalibeik H, Mokhtaran M, Mokhtaran A. Intraocular Lens Power Formula Selection Using Support Vector Machines. *Frontiers in Biomedical Technologies.* 2015;2(1):8.
15. Kalpathy-Cramer J, Campbell JP, Erdogmus D, et al. Plus Disease in Retinopathy of Prematurity: Improving Diagnosis by Ranking Disease Severity and Using Quantitative Image Analysis. *Ophthalmology.* 2016;123(11):2345-2351.
16. Lee CS, Lee AY, Holland GN, Van Gelder RN, Tufail A. Big Data and Uveitis. *Ophthalmology.* 2016;123(11):2273-2275.