



Feasibility of simple machine learning approaches to support detection of non-glaucomatous visual fields in future automated glaucoma clinics

Peter B. M. Thomas¹  · Thomas Chan² · Thomas Nixon³ · Brinda Muthusamy⁴ · Andrew White^{2,5}

Received: 16 July 2018 / Accepted: 13 February 2019 / Published online: 4 March 2019
© The Royal College of Ophthalmologists 2019

Abstract

Objectives To assess the performance of feed-forward back-propagation artificial neural networks (ANNs) in detecting field defects caused by pituitary disease from among a glaucomatous population.

Methods 24-2 Humphrey Visual Field reports were gathered from 121 pituitary patients and 907 glaucomatous patients. Optical character recognition was used to extract the threshold values from PDF reports. Left and right eye visual fields were coupled for each patient in an array to create bilateral field representations. ANNs were created to detect chiasmal field defects. We also assessed the ability of ANNs to identify a single pituitary field among 907 glaucomatous distractors.

Results Mean field thresholds across all locations were lower for pituitary patients (20.3 dB, SD = 5.2 dB) than for glaucoma patients (24.4 dB, SD = 5.0 dB) indicating a greater degree of field loss ($p < 0.0001$) in the pituitary group. However, substantial overlap between the groups meant that mean bilateral field loss was not a reliable indicator of aetiology. Representative ANNs showed good performance in the discrimination task with sensitivity and specificity routinely above 95%. Where a single pituitary field was hidden among 907 glaucomatous fields, it had one of the five highest indexes of suspicion on 91% of 2420 ANNs.

Conclusions Traditional artificial neural networks perform well at detecting chiasmal field defects among a glaucoma cohort by inspecting bilateral field representations. Increasing automation of care means we will need robust methods of automatically diagnosing and managing disease. This work shows that machine learning can perform a useful role in diagnostic oversight in highly automated glaucoma clinics, enhancing patient safety.

Introduction

There is a relentlessly growing demand for glaucoma services worldwide driven by an increasing population, increasing life expectancy, earlier diagnosis, and a proliferation of treatment modalities. Innovation in models of care delivery have sought to redress the imbalance between supply and demand: devolution of routine care to non-ophthalmologists and the introduction of virtual clinics, for example, have allowed high quality care to be delivered to more patients at lower cost. The next logical step in this trend is greater reliance on computers to perform clinical data analysis and to make automated care decisions for patients. Recent advances in machine learning, a field of artificial intelligence, substantially increase the capability of computers to perform human-like analysis of complex clinical data. This technology will likely drive significant automation of care in the future. In an automated clinic, it would be hoped that the ophthalmologist can take on the role of supervisor,

✉ Peter B. M. Thomas
pbmthomas@gmail.com

¹ NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London EC1V 9EL, UK

² Discipline of Ophthalmology, University of Sydney, Sydney, Australia

³ Department of Ophthalmology, Faculty of Clinical Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK

⁴ Department of Ophthalmology, Addenbrooke's Hospital, Cambridge, UK

⁵ PersonalEYES, Sydney, NSW, Australia

checking global metrics of clinic performance, reviewing cases flagged as high risk, and assessing a proportion of patients cared for to ensure appropriate management.

There are reasonable concerns, however, about changing models of care provision. Though there is evidence that glaucoma is well managed in virtual clinics [1, 2], there is still a reliance on the training and expertise of the supervising ophthalmologist who is always at risk of being overwhelmed by the deluge of assessments. The system may therefore miss important diagnoses when a non-expert assessor sees a patient—it is simply not feasible for the supervisor to review all work undertaken in the virtual clinic or in future automated clinics. A classic pitfall in a glaucoma clinic is the unrecognised chiasmal-compression-associated visual field loss from a pituitary mass [3, 4]. Machine learning tools may assist in flagging those cases in need of a second look, reducing this risk.

Machine learning is a branch of artificial intelligence. It allows computers to perform tasks by learning from examples, rather than requiring explicit, task-specific programming. When there is high quality data with which to train the computer, and the machine learning algorithm is appropriately chosen, this approach allows extremely complex tasks to be performed far more successfully than conventional computing approaches. The modern history of machine learning can be traced to the middle of the 20th century [5], and it has been applied in glaucoma for at least 20 years [6, 7].

Various paradigms in machine learning have enjoyed success, but the most flexible has proven to be the artificial neural network. These consist of nodes and links which bear a superficial resemblance to neurones and dendrites, respectively. The classic structure is a feed forward back-propagation neural network. These consist of at least three layers of nodes: an input layer of nodes which allows information to be fed into the network; one or more “hidden” layers of nodes where the inputs are combined; and an output layer of nodes where the response of the network is reported. Each layer is connected to the next by links which vary in strength as the neural network learns. Errors during the learning process are “propagated” backwards to alter link strengths and improve performance.

Recently, advances in hardware processing power and machine learning understanding have allowed successful deep neural networks to be developed. These show great promise in performing very complex tasks which previously confounded artificial neural networks, for example beating the world Go champion [8] or winning the game show Jeopardy [9]. Simply put, deep neural networks consist of many layers of nodes, with more complex structures and more complex rules determining learning behaviour. However, deep learning requires very large amounts of data to learn, and much more processing power, and is not always required when the learning task is simple.

Multiple paradigms in machine learning (among them artificial neural networks) have achieved notable successes in glaucoma including, but not limited to, detecting early visual field change [10], detecting progression [11], and analysing fundal and OCT images [12]. There has been much less interest in applying the technique to differentiate glaucoma from its masquerades. More traditional approaches to data analysis have had some success in this. The Neurological Hemifield Test [13, 14], for example, has been shown to perform, as well as sub-specialist clinicians in detecting neurological field defects from among a population of glaucoma patients.

In a theoretical automated clinic, it is vital that some of the general abilities of a human clinician are simulated. An exclusive focus on managing glaucoma could be dangerous if it ignores the existence of co-pathologies or masquerading conditions. In this paper, we assess the ability of artificial neural networks, a well-established paradigm in machine learning, to detect visual field defects due to pituitary disease from among a large population of glaucomatous visual fields. We also hope this paper will serve as an accessible example of how machine learning (a complex field for a clinician) could begin to help our clinical practice in the near future.

Materials and methods

The project was covered by UK Research Ethics Committee and Health Research Authority approval (IRAS project ID 232104). The work adheres to the Declaration of Helsinki.

Identification of patients

Patients with pituitary lesions were identified from the database of the pituitary multi-disciplinary team at Addenbrooke’s Hospital. Visual fields for these patients were identified from the Humphrey’s Visual Field machines. The visual fields were screened by an experienced clinician. Those with no significant field loss were excluded, as were noisy fields with no discernible pattern of loss. Fields with a pattern that was not classic for chiasmal compression were allowed in to the training dataset.

Glaucomatous visual field data ($n = 907$ bilateral fields) were obtained from a database of Humphrey’s Visual Field data kept by private practice group based in Sydney (Australia). This dataset included a range of severities (from mild to severe) and patients with monocular and binocular field loss.

Extraction and handling of visual field data

Humphrey visual fields (24-2) were available in PDF format. For the pituitary group, an optical recognition

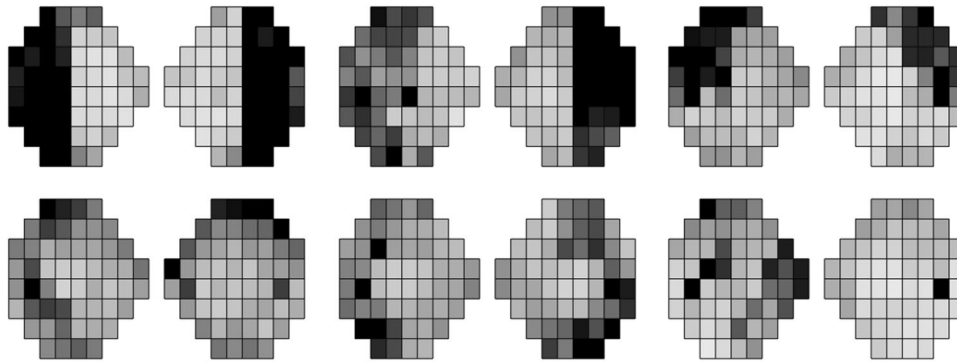


Fig. 1 Examples of bilateral field representation for three pituitary fields (top) and glaucomatous fields (bottom). For each pair, the left field represents the left eye and the right fields the right eye. The fields are oriented as the patient would see the world, i.e. temporal loss is

shown temporally, superior loss superiorly. Predominantly temporal defects are shown in the three pituitary field representations, whereas the glaucomatous fields show more variation in defect shape, position, severity and laterality

programme was developed in Matlab (Mathworks Inc., Nantick, MA, USA) to extract the numerical values and laterality from each PDF. No errors were detected in a random comparison of extracted values from 20 PDFs. Search and replace was undertaken to remove the symbol “<” from the entry “<0”. An equivalent, proprietary optical character recognition technique was used to extract the raw data from the glaucomatous group. For this study, the raw data from the numeric decibel map was considered. Values were transcribed into a matrix, with data from both eyes coupled for a single visit to clinic. Each pair of visual fields was labelled as a glaucomatous or pituitary defect. Each entry could be transformed into a bilateral graphical representation of field loss (Fig. 1 shows examples from the pituitary and glaucoma groups).

Machine learning

Machine learning was performed in Matlab using the Neural Networks Toolbox (Mathworks Inc., Nantick, MA, USA). Feed-forward back-propagation artificial neural networks were created. The input layer had one node per point on the bilateral visual field representation giving 108 inputs in total (54 per eye). A hidden layer size of 10 nodes was chosen. There was one output node, the activity in which would represent suspicion of chiasmal field defect. Sigmoid hidden and softmax output neurones were employed, and training was undertaken with scaled conjugate gradient back-propagation. Owing to the simple structure of the network, CPU training could be undertaken with each network trained taking <20 s.

Two assessments of this network structure were made. First, the performance of the network was tested using a training set of 70% of the total available bilateral field representations with 15% each used for validation and post-training testing. Inclusion of 15% of the fields in a

validation set allows “overfitting” to be detected. As training progresses, there is a risk that the network will effectively memorise the fields it has seen rather than relying on categorisation strategies that can be adapted to unseen fields. During training, overfitting is detected by analysing the error rate in the validation set. When the error rate among the validation set begins to rise during training (even if the error rate on the training set is still decreasing), then training is terminated and the network weights and biases that give minimum error in the validation set are used. Useful metrics of post-training performance include confusion matrices (false and true positives and negatives), and ROC curves.

Second, a “needle-in-a-haystack” task was simulated. A fresh artificial neural network was trained (same structure and protocols as above) using the same ratios for training, validation, and testing datasets as above. One of the 121 pituitary fields was withheld entirely from the training process. The trained network was then presented with all of the glaucomatous bilateral field representations ($n = 907$, the “haystack”) and the single pituitary field (the “needle”) held back during the training process. This process of training and interrogation was repeated 20 times for each of the 121 pituitary fields. Performance of each of the 2240 artificial neural networks generated in this way was characterised by the rank position of the pituitary field in terms of suspicion of non-glaucomatous diagnosis among the 907 distractor glaucoma fields.

Results

Characteristics of the visual field representations

Figure 1 shows a selection of the pituitary and glaucomatous bilateral visual field representations. Mean dB

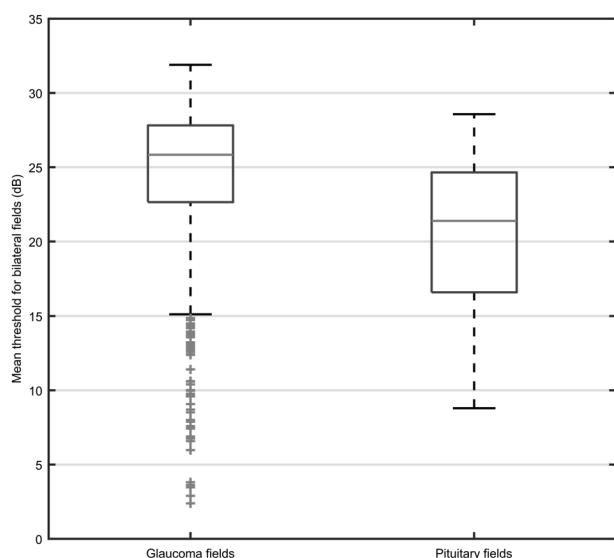


Fig. 2 Box and whisker plot of mean threshold (dB) across all bilateral field representations for glaucomatous fields (left) and pituitary fields (right). Field loss was greater ($p < 0.0001$) in the pituitary group, though mean field loss fell within the limits of the glaucomatous field distribution

threshold across fields for both eyes in the pituitary group was 20.3 dB. The corresponding values for the glaucomatous fields was 24.4 dB. The mean field loss across the bilateral representations was significantly greater in the pituitary group ($p < 0.0001$, unpaired *t*-test). However, there was significant spread in both glaucomatous ($SD = 5.0$ dB) and pituitary ($SD = 5.2$ dB) groups, and resulting overlap means that mean amount of field loss alone was not a reliable indicator of glaucomatous or pituitary aetiology. Figure 2 visualises the characteristics of mean field loss in the two groups.

Performance of the neural network

Confusion matrices for the task of diagnosing each field as glaucomatous or pituitary are shown in Fig. 3, as are the receiver operating characteristic plots for this task. Performance shown in Fig. 3 is representative of many repeats of neural network training. Better and worse performance could be obtained by randomly changing the constituents of the training, validation, and testing groups. For the example shown in Fig. 3, sensitivity was 95.9% and specificity and 99.8% specificity across the whole test population.

The “needle-in-a-haystack” analysis ranks each bilateral field according to the likelihood of being a pituitary field. As there was a single pituitary “needle” and 907 glaucomatous fields in the “haystack”, a rank of 1 means that a given field is the most likely candidate to be a pituitary field, and a rank of 908 corresponds to the least likely. Across the 2420 artificial neural networks generated (20 repeats for each of

121 pituitary fields withheld during training), the median rank given to the pituitary field was 1st (mean rank was 14th being skewed by the few fields poorly identified and having high ranks). Figure 4 shows the performance of the artificial neural networks in this task in more detail. Figure 5 shows the fields in the pituitary which ranked outside the top 5 in terms of suspicion of being pituitary.

Discussion

Enthusiasm for machine learning and artificial intelligence in ophthalmology tends to focus on their potential to improve upon and automate specific aspects of a patient’s workup for a known diagnosis. In glaucoma this includes assessment of the optic disc [15, 16], detection of field progression [11, 17], and analysis of an OCT scan [12]. Relatively neglected has been consideration of how these technologies will simulate the more general processes a human clinician performs. This leads to reasonable objections: when a machine that knows only glaucoma is managing a patient, what if that patient has, or develops, a different diagnosis? There is a need to consider how these new technologies can be adapted to use in clinic, and we must also be cautious not to rely too heavily on the isolated judgements made by machine learning approaches even if the underlying methodology is powerful [18].

Missed pituitary tumours in the glaucoma clinic are a well-known clinical pitfall [3, 4]. When they present with progressive visual field loss, they can only be distinguished from glaucoma if they present with other symptomatology (many will not) or by consideration of visual fields (preferably bilateral). Other clinical findings suggestive of chiasmal compression might be suggestive to the neuro-ophthalmologist (for example colour vision and RAPD), but these might not arouse suspicion in a high-volume glaucoma clinic staffed by non-specialists. Most automated assessments of visual fields consider only a single eye, though there are algorithms that can detect a neurological field loss despite this [13, 14]. Machine learning will undoubtedly contribute significantly to visual field interpretation in future manned and automated clinics. As yet, there has been little consideration of how machine learning can be applied to detect non-glaucomatous diagnoses where bilateral field representations are required.

The results we present here support the adaptability of feed-forward back-propagation artificial neural networks, a well-established paradigm in machine learning, to the task of detecting pituitary visual field defects in a realistic setting. Among our visual fields from pituitary patients, some showed classic bitemporal hemianopias with a sharp distinction between the normal nasal fields and lost temporal fields. Most, however, were less classic with only partial,

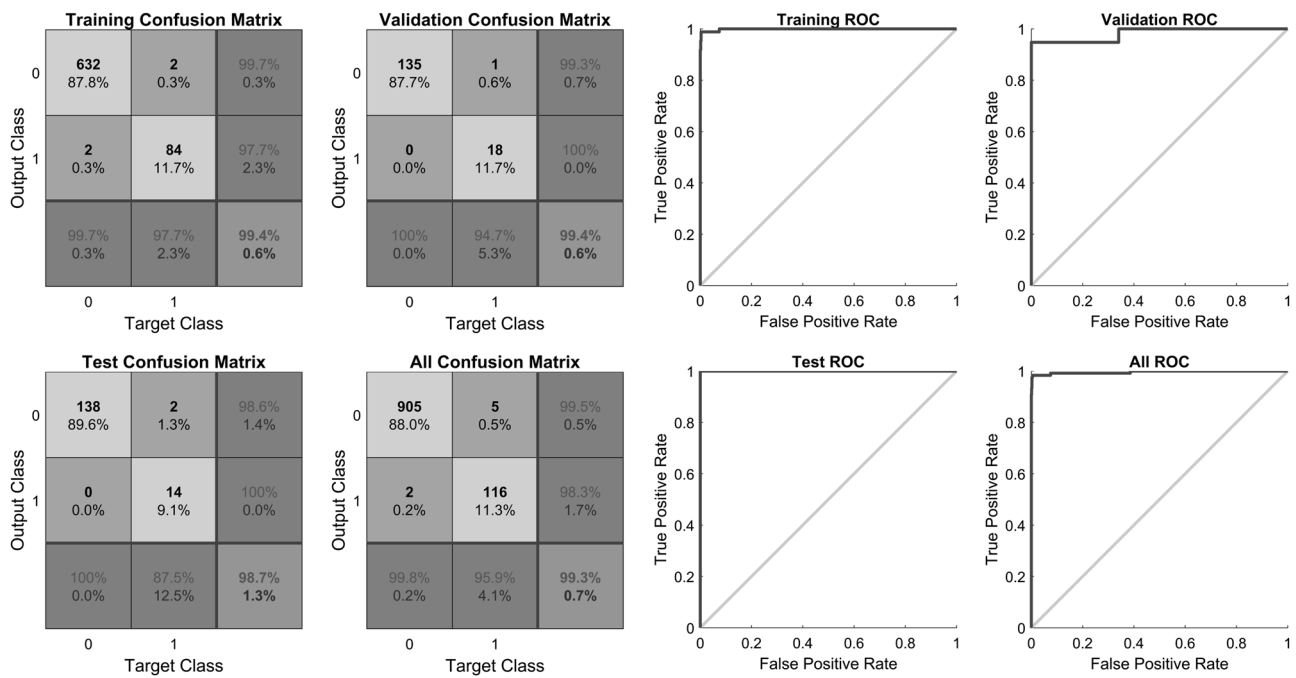


Fig. 3 Confusion matrices for training, validation, and test groups are shown on the left, as well as a confusion matrix for all data combined. Class 1 are pituitary fields, class 0 glaucomatous. Green squares are true negatives and positives, red are false positives and negatives. The bottom left “test confusion matrix” shows the performance of an artificial neural network confronted with fields that were not used

during the training process (i.e. not in the training or validation group). In this example (which is representative of many repeats—neither the best nor the worst performing network generated), the test matrix shows two false negatives where pituitary fields were classified as glaucomatous, and no false positives. The respective receiver operating characteristic curves are shown for all groups on the right

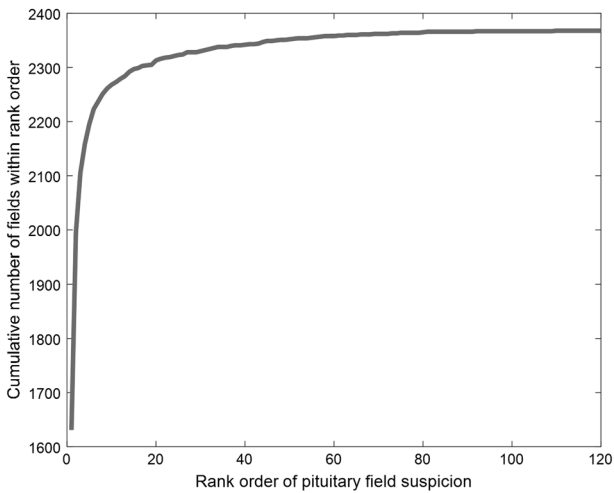


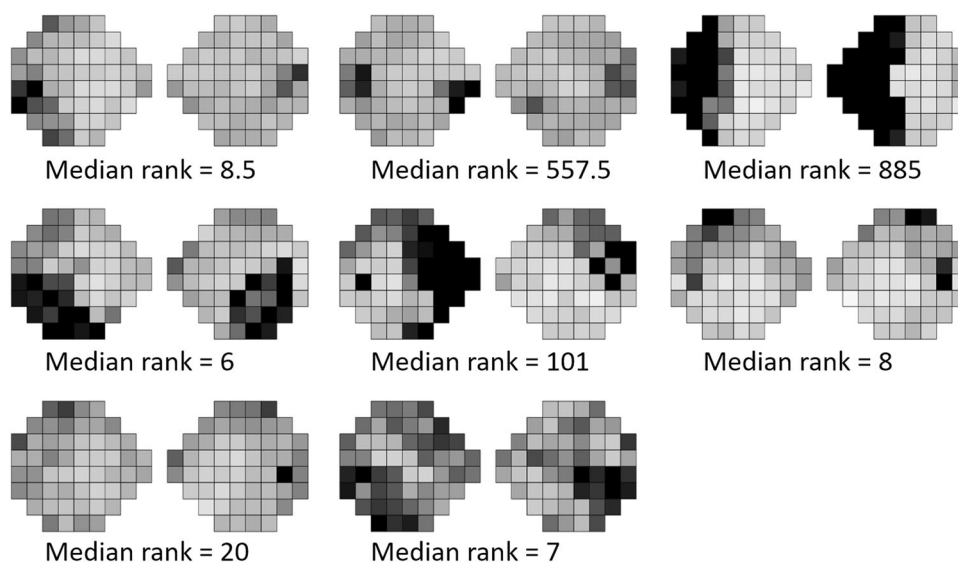
Fig. 4 Cumulative plot of rank (abscissa) within which the 2420 artificial neural networks (ordinate) classify the pituitary visual field. 1631 out of 2420 networks identified the pituitary field as the most likely (rank #1, 67%) to be of pituitary aetiology; 2195 identified it as one of the most likely five (rank #5 or better, 91%); and 2268 as one of the most likely 10 (rank #10 or better, 94%)

and not always symmetric, loss. This corresponds to the clinical observations that bitemporal hemianopia is the classic but by no means exclusive pattern of field defect seen in pituitary lesions [19, 20]. The pituitary fields on which the artificial neural networks performed most poorly

are shown in Fig. 5. These fields are far from classic for a pituitary mass and may not be diagnosed as such by many doctors. As we hoped to represent a real-world scenario, where training dataset could be selected from electronic medical records with minimal human supervision, we resisted the temptation to prune the pituitary field dataset down to classical fields.

This work has limitations. It is unclear, for example, when applied to a real clinic where the rate of pituitary tumours will be low, whether too many false positives would arise. We do not, however, envisage this approach as forming part of a fully automated pathway for some time—detection of a potential chiasmal defect would not automatically lead to an expensive and possibly unnecessary MRI scan, but rather flag the patient record for review by a senior clinician. Based on the results we describe here, a senior clinician reviewing five fields out of a clinic population of 1000 would give the opportunity to detect the majority of pituitary tumours (91% based on an incidence here of 1.1 pituitary lesion per thousand patients in a clinic—the true incidence of pituitary lesions in a glaucoma clinic is not known). We also did not include structural data, for example from an OCT. Undoubtedly this would improve accuracy but would also limit the benefits to those patients who had access to structural imaging—many clinics in less developed countries will not have these machines, and

Fig. 5 The fields from the pituitary group most poorly detected among the glaucomatous distractors. The median rank given to each field represents the rank given by 20 artificial neural networks asked to detect the pituitary field among 907 glaucomatous distractors. All fields where the median rank was worse than 5 are shown. Several fields from pituitary patients rank just outside the top 5. Very poorly ranking fields have patterns that are not typically seen in pituitary disease, and likely represent co-pathology



reliable scans are not possible in all patients (for example where cataract makes automated analysis inaccurate).

We also concentrate only on a single type of neurological field defect—variants of bitemporal hemianopia due to chiasmal compression. There are many other patterns of field loss that suggest a non-glaucomatous diagnosis: among these are altitudinal defects caused by ischaemic optic neuropathy, congruous and incongruous homonymous field loss from post-chiasmal lesions, and scotomata from optic neuritis. A follow-on study will consider these patterns, and there is no reason a priori to believe that performance will not translate. The major challenge will be in building a training dataset for those more uncommon patterns of field loss. Even a very large centre, for example, may only have a few examples of a junctional scotoma to contribute to a machine learning approach.

The machine learning paradigm we employ, the traditional feed-forward back-propagation neural network, might seem dated given the current success of deep neural networks. However, it is an appropriate tool for the task. The amount of data we had available, and the relative simplicity of the task of detecting specific patterns of field loss, make deep learning a rather large hammer with which to hit a small nail. In future work, where many more visual fields will be available, and combination with structural measures would be desirable, then a more complex machine learning paradigm, such as deep learning (or whatever comes next) would certainly be worthwhile. The intention of the current work is not to try to exceed the performance of the trained human (who will form his diagnostic suspicion using many data sources), but simply to show that this particular task (using visual fields alone) can be performed well in a semi-automated fashion.

Until machine learning can furnish us with a generalised intelligence capable of learning to perform medicine like a

human does, we will likely have to rely on a battery of algorithms that are expert at specific aspects of patient care. Our work shows that machine learning adapts well to the interpretation of bilateral visual field representations and can detect an important pathology to distinguish from glaucoma. Follow on projects will expand the work to categorisation of all visual field patterns and introduce structural measures to further refine accuracy.

Summary

What was known before

- Machine learning is a powerful technique for analysing patient data.
- Computer algorithms are successful at distinguishing between glaucomatous and neurological field defects.

What this study adds

- By considering bilateral field representations, simple machine learning techniques are highly successful in distinguishing field loss caused by chiasmal compression from that caused by glaucoma.
- This approach can support surveillance of patients in large virtual clinics and in the automated clinics of the future.

Acknowledgements The research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Funding PT's time was funded by the NIHR Biomedical Research Centre at Moorfields Eye Hospital and the Institute of Ophthalmology. No further funding was used for this study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Kotecha A, Brookes J, Foster P, Baldwin A. Experiences with developing and implementing a virtual clinic for glaucoma care in an NHS setting. *Clin Ophthalmol*. 2015;9:1915.
2. Kotecha A, et al. Developing standards for the development of glaucoma virtual clinics using a modified Delphi approach. *Br J Ophthalmol*. 2017;102:531–4. [bjophthalmol-2017-310504](https://doi.org/10.1136/bjophthalmol-2017-310504)
3. Drummond SR, Weir C. Chiasmal compression misdiagnosed as normal-tension glaucoma: can we avoid the pitfalls? *Int Ophthalmol*. 2010;30:215–9.
4. Kobayashi N, Ikeda H, Kobayashi K, Onoda T, Adachi-Usami E. [Characteristics of fifty cases of pituitary tumors in eyes diagnosed with glaucoma]. *Nihon Ganka Gakkai Zasshi*. 2016;120:101–9.
5. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev*. 1958; 65:386–408.
6. Goldbaum MH, et al. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci*. 1994;35:3362–73.
7. Madsen EM, Yolton RL. Demonstration of a neural network expert system for recognition of glaucomatous visual field changes. *Mil Med*. 1994;159:553–7.
8. DeepMind's AI beats world's best Go player in latest face-off | New Scientist. <https://www.newscientist.com/article/2132086-deepminds-ai-beats-worlds-best-go-player-in-latest-face-off/>. Accessed 12 July 2018.
9. IBM's Watson supercomputer crowned Jeopardy king—BBC News. <https://www.bbc.co.uk/news/technology-12491688>. Accessed 12 July 2018
10. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*. 2016;123:1974–80.
11. Yousefi S, et al. Unsupervised Gaussian mixture-model with expectation maximization for detecting glaucomatous progression in standard automated perimetry visual fields. *Transl Vis Sci Technol*. 2016;5:2.
12. Muhammad H, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma*. 2017;26:1.
13. Boland MV, et al. Evaluation of an algorithm for detecting visual field defects due to chiasmal and postchiasmal lesions: the neurological hemifield test. *Investig Ophthalmol Vis Sci*. 2011;52:7959.
14. McCoy AN, et al. Development and validation of an improved neurological hemifield test to identify chiasmal and postchiasmal lesions by automated perimetry. *Invest Ophthalmol Vis Sci*. 2014;55:1017–23.
15. Haleem MS, et al. A novel adaptive deformable model for automated optic disc and cup segmentation to aid glaucoma diagnosis. *J Med Syst*. 2018;42:20.
16. Omodaka K, et al. Classification of optic disc shape in glaucoma using machine learning based on quantified ocular parameters. *PLoS ONE*. 2017;12:e0190012.
17. Khalil T, Khalid S, Syed AM. Review of Machine Learning techniques for glaucoma detection and prediction. In: *Proceedings of the 2014 Science and Information Conference, SAI 2014*. p. 438–42. <https://doi.org/10.1109/SAI.2014.6918224>
18. Yuki K, et al. Predicting future self-reported motor vehicle collisions in subjects with primary open-angle glaucoma using the penalized support vector machine method. *Transl Vis Sci Technol*. 2017;6:14.
19. Lee IH, et al. Visual defects in patients with pituitary adenomas: the myth of bitemporal hemianopsia. *Am J Roentgenol*. 2015;205: W512–8.
20. Ogra S, et al. Visual acuity and pattern of visual field loss at presentation in pituitary adenoma. *J Clin Neurosci*. 2014; 21:735–40.