REVIEW

Deep learning in ophthalmology: a review

Parampal S. Grewal, MD,* Faraz Oloumi, PhD,† Uriel Rubin, MD,*
Matthew T.S. Tennant, MD, FRCSC*

Deep learning is an emerging technology with numerous potential applications in Ophthalmology. Deep learning tools have been applied to different diagnostic modalities including digital photographs, optical coherence tomography, and visual fields. These tools have demonstrated utility in assessment of various disease processes including cataracts, glaucoma, age-related macular degeneration, and diabetic retinopathy. Deep learning techniques are evolving rapidly, and will become more integrated into ophthalmic care. This article reviews the current evidence for deep learning in ophthalmology, and discusses future applications, as well as potential drawbacks.

Introduction

Ophthalmology is on the cusp of a revolution in the screening, diagnosis, and management of eye disease. This revolution is being led by computer-based deep learning (DL) technology that has the potential to change the practice of ophthalmology. DL is the newest and fastest growing component in the field of machine learning. It is a process by which vast databases are analyzed, and then compared to known ground-truths. Such DL technology is behind self-driving cars and the improvement in the success of computers to win at board games such as go and chess.[1,2] This technology is incorporated into social media and photo software to identify searchable components of a video or image.[3] In due time, DL may even allow the development of true artificial intelligence.

Ophthalmologists rely on pattern recognition via direct or indirect visualization of the eye and its surrounding structures to diagnose disease. Ophthalmology-related diagnostic technology provides further information to the clinician via digital representation of these same structures. This dependence on imaging makes the field of ophthalmology perfectly suited to benefit from DL algorithms. Incorporation of DL algorithms into the practice of ophthalmology has begun and could potentially change the fundamental type of work performed by ophthalmologists. Computer led intelligence will likely play an important role in screening and diagnosis of eye disease in the coming years. Such technological advancements may lead to human resources being left to focus on direct clinician-patient interactions such as a discussion of the diagnosis, prognosis, and treatment options. We believe that giving of consent and any medical or surgical intervention will remain the responsibility of a human clinician for the near future. Inclusion of DL algorithms into ophthalmology decision making is likely to take place faster than many may expect.[4]

Technical Aspects of Deep Learning

There are a few terms commonly associated with computer methods that deal with analysis and decision-making scenarios, including medical image assessment. Computer-aided diagnosis refers to approaches in which specific clinical features that are associated with the disease are extracted via specifically designed image-processing filters/tools.[5] The term “machine learning” generally refers to any pattern-classification method that requires a training scheme, whether supervised or unsupervised, to learn what may delineate the underlying patterns.[6] DL, for the most part, refers to machine-learning schemes that employ convolutional based neural networks (CNNs). Such networks utilize a bank of image-processing filters to extract various types of image features that the network deems indicative of pathological signs, as learned from the training set, which is an integral part of pattern classification methods.[7] DL can be viewed as a brute-force approach in determining the most appropriate image-processing filters/tools, which can quantify various biomarkers of the disease. Finally, the term artificial intelligence (AI), in a very general sense, refers to any system, mainly machine-learning based, that is capable of learning new patterns and features in an unsupervised manner, where intervention from a third party, usually humans, is not required. It is debatable whether a true AI system already exists or not.

Initially, the main drawback in real-world application of CNNs was simply lack of computational power. The rise of graphical processing units (GPUs), that have far surpassed the computational capabilities of central

https://doi.org/10.1016/j.jcjo.2018.04.019
ISSN 0008-4182/17

Canadian Ophthalmological Society | Société canadienne d’ophtalmologie
EYE PHYSICIANS AND SURGEONS OF CANADA | MÉDECINS ET CHIRURGIENS OPHTALMOLOGISTES DU CANADA

processing units (CPUs), has made running CNNs of farther depths (deep learners) much more time efficient and powerful. These deep neural networks are now commonly referred to as DL solutions and include networks such as GoogLeNet,[8] VGG,[9] and SegNet.[10] Such solutions have shown a lot of promise in generalizing overall content of an image, as well as in industrial applications such as different suggestive algorithms similar to Netflix's recommendation system. In the recent years, there have been many attempts to assess such systems for medical applications, including biomedical-image analysis.

DL-solutions for biomedical-image analysis fall into two main categories:

1 Supplying the DL network with only images and their associated conditions (diagnoses/labels/stages), commonly referred to as image-classification methods; and
2 Supplying the network with images and their associated ground-truth masks (black and white images), in which the pathological conditions that are associated with the disease are manually delineated, commonly referred to as semantic-segmentation methods.

Figure 1 illustrates a general flow-chart representation of a typical CNN that provides a two-class classification.

Deep Learning in Ophthalmology

DL research in ophthalmology has progressed rapidly, and has the potential to become a part of daily clinical practice in the relatively near future.[4] In the same way that an electrocardiogram machine can provide a relatively accurate reading of the electric function of the heart, DL algorithms are now offered by a number of private companies that focus on screening for retinal disease including diabetic retinopathy and macular degeneration.[11–13] In due time, DL algorithms may become incorporated into many digital ophthalmic diagnostic tools.

Cataract

Anterior segment research has demonstrated the ability of DL to grade nuclear cataracts from cross-sectional slit lamp images on the Wisconsin grading system as compared to professional graders. The mean absolute error was 0.304 and the integral grading error $\leq$ 1.0 was 99%, which outperformed other published automated cataract grading methods.[14] DL has also been applied to slit lamp images of pediatric cataracts. A CNN algorithm had high sensitivity and specificity for grading, density, and location of pediatrics cataracts when compared with Pediatric Ophthalmologists.[15]

Glaucoma

Glaucoma research has investigated the application of DL for different testing modalities including optic nerve images, optical coherence tomography (OCT), and visual fields. Clinical optic nerve assessment includes cup-to-disk ratio, focal changes in the neuroretinal rim, and peripapillary nerve fiber layer, disc hemorrhages, and vessel changes. CNN has been used to develop an algorithm to diagnose a glaucomatous optic nerve as defined by clinical assessment with area under the curve (AUC) of receiver operative characteristic curve of 0.831 and 0.887 on 2 separate databases.[16] Modern day glaucoma assessment relies on OCT imaging of the optic nerve. A hybrid DL method using single wide-field OCT has successfully classified healthy glaucoma suspects from early glaucoma by creating a probability map that was 93.1% accurate when compared with a glaucoma expert diagnosis, outperforming current OCT parameters.[17] Other research has

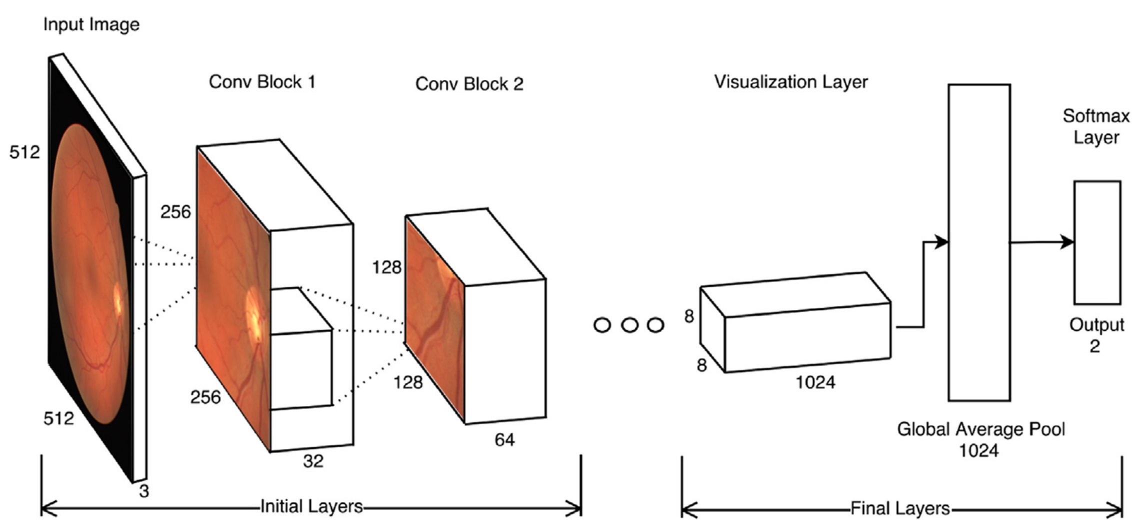

Fig. 1—A typical flow-chart representation of a deep learning network that provides classification of the image into two groups. Note that the size of the images and the convolutional blocks may vary in different applications. The convolutional layers attempt to extract relevant image features while reducing the dimensionality of them. The Softmax layer is responsible for the decision-making aspect of the network.

demonstrated utility of DL for segmentation of OCT images of the optic nerve head, including accurate identification of the retinal nerve fiber layer (accuracy 0.84 ± 0.03).[18] DL has also successfully assessed visual fields identifying preperimetric glaucomatous field (Anderson-Patella's criteria) by Humphrey Field Analyzer 30-2 (Meditec, Dublin, Calif). The deep feed-forward neural network had an AUC of 92.6% (95% CI 89.8%–95.4%), which was superior to other machine learning methods.[19]

Retinal Disorders

Retina research regarding age-related macular degeneration (AMD) and diabetic retinopathy (DR) has demonstrated the utility of DL in screening and diagnosis. A recent AMD study effectively used a deep CNN to grade AMD from fundus photos. A database of more than 130 000 color fundus photos graded by experts was used. The study classified AMD into disease-free/early stages (AREDS stage 1 or 2) and intermediate/advanced stages (AREDS stage 3 or 4). Intermediate/advanced stages were defined considering this as referable AMD. The accuracy of DL ranged from 88.4%± 0.5% to 91.6% ± 0.1% and the AUC was between 0.94–0.96.[20] DL has also been applied to detect AMD from OCT images. A DL approach using TensorFlow (Google Inc., Mountain View, Calif) was developed from a database of 1.2 million images. This was subsequently validated on a separate set of 50 healthy controls and 50 neovascular AMD patients with 100% accuracy.[21]

Initial research into DR has looked at identifying specific disease features such as microaneurysms,[22] exudates[23] and hemorrhages.[24] Subsequent literature has demonstrated that DL algorithms can identify DR without explicitly defining these features. A recent study using a database of 75 137 fundus images to develop a DL algorithm could distinguish healthy fundi from those with DR. The AUC was 0.97 achieving a sensitivity of 94% and specificity of 95%.[25] Emerging literature has investigated identification of more specific grades of DR. A study by Google Inc. (Mountainview, Calif), developed an algorithm using CNNs to identify referable DR. They defined referable DR as hard exudates within 1 disc diameter of the macula and/or moderate and worse DR. The model was validated on two separate data sets achieving an AUC of 0.991 (95% CI 0.988–0.993) and 0.990 (95% CI 0.986–0.995). They defined two operating cut points for sensitivity and specificity. Depending on the cut point, the sensitivity ranged between 87.0%–97.5% and the specificity ranged between 93.4%–98.5%.[26]

DL has also been applied to retinal imaging, namely OCT layer segmentation. CNNs have been developed to identify intraretinal fluid with a cross-validation Dice coefficient of 0.911.[27] Furthermore, DL has been used to improve segmentation in diabetic macular edema (DME)[28] and AMD.[29] Recently, DL has successfully been applied to accurately detect and quantify macular fluid on OCT. An automated method was developed using OCT images from patients with neovascular AMD, DME, and retinal vein occlusion. The DL algorithm successfully detected intraretinal cystoid fluid (mean AUC 0.94, range 0.91–0.97) and subretinal fluid (mean AUC 0.92, range 0.86–0.98).[30]

Appendix 1 (supplemental materials, available online) provides an overall summary of the type of technical methods, application, and results of several works that have assessed the use of DL networks in ophthalmic images.

Expanded DL

A recent study has investigated multiple disease detection using DL. Specifically, a DL algorithm was developed and validated for detection of DR, possible glaucoma, and AMD using fundus images from a population with diabetes. Referable DR was defined as moderate NPDR or worse, DME (presence of hard exudates) or ungradable image. Vision threatening DR was defined as severed NPDR or PDR. Possible glaucoma was defined as a cup-to-disc ratio of ≥ 0.8, focal neuroretinal rim or retinal nerve fiber layer changes, or optic disc hemorrhage. AMD was defined as intermediate or worse AMD. The algorithm achieved high sensitivity and specificity for detecting referable DR [sensitivity 90.5 (95% CI 87.3%–93.0%); specificity 91.6% (95% CI 91.0%–92.2%)], vision threatening DR [sensitivity 100% (95% CI 94.1%–100%); specificity 91.1% (95% CI 90.7%–91.4%)], possible glaucoma [sensitivity 96.4% (95% CI 81.7%–99.9%); specificity 87.2% (95% CI 86.8%–87.5%)], and AMD [sensitivity 93.2% (95% CI 91.1%–99.8%); specificity 88.7% (95% CI 88.3%–89.0%]). This was further validated in 10 multiethnic cohorts of patients with diabetes.[31]

Discussion

Benefits of Deep Learning

DL has become a reliable tool to interpret ocular data derived from digital photographs, OCTs, and visual fields. It has been shown to effectively screen for common blinding diseases such as AMD, DR, and glaucoma. Early diagnosis of these conditions will serve to reduce preventable vision loss. Internationally and in underserved communities across Canada there is a significant screening burden, and an efficient automated teleophthalmology process would help to address unmet screening needs. DL algorithms can already be applied to teleophthalmology programs to identify normal from abnormal retinal images thereby reducing the number of images sets reviewed by the clinician. As sensitivity and specificities improve for DL systems, the number of image sets to review will be reduced, improving access to underserved communities by minimizing clinician time and alleviating clinician scarcity.

DL solutions can assist clinicians in identifying abnormalities present via a diagnostic test. In such a scenario, a provisional diagnosis with associated description of key abnormalities found with the test may be provided by the computer. Confirmation of the diagnosis as well as counselling and treatment would remain the responsibility of the Ophthalmologist.

DL diagnostic systems could be integrated into the primary care setting reducing or potentially eliminating unnecessary referrals. Taking such systems one step further, DL tools could enable ophthalmic self-monitoring by patients via smartphone retinal photography, visual acuity and visual field testing. Such technology would empower patients, facilitate early diagnosis, as well as identify treatable eye disease.

Consistent interpretation of ocular data by DL might also facilitate high quality ophthalmic research by reducing grading and tester variability.

Limitations of Deep Learning

While the promise of DL may seem unlimited, concerns remain. These concerns are well founded as improper DL training can lead to unintended and detrimental outcomes. Such detrimental outcomes can be illustrated by a recent example.

Tay is Microsoft's AI chat-bot that was setup on Twitter to learn from the existing tweets on how to communicate and tweet for itself.[32] Tay's very first tweet was "…I'm so stoked to meet you…humans are super cool." However, within 16 hours, Tay had learned to be racist and misogynistic, tweeting messages like "Hitler was right, I hate Jews." This unintended and detrimental outcome occurred because Tay was learning through engaging conversation with users respondimg with biased, racist, and misogynistic sentiments. Microsoft shut down Tay in $<$ 24 hours, and it hasn't been back online since. While this example may be extreme, such a scenario emphasizes the need to train such machine-learning networks with unbiased, clearly marked, and multireference training data, having same features marked by several experts to avoid any interobserver variability in the training of DL networks.

DL technologies that focus on identification of a specific ocular disease may miss other important ocular factors (level of vision, lens status, glaucomatous nerve, choroidal melanoma) and systemic factors (glycemic control, other diabetic vascular disease, relevant comorbidities). Variability exists in definitions for referable disease, and current algorithms may not incorporate disease severity and referral urgency.

Limitations remain for DL when providing quantitative measures or assessment of interval changes in disease severity. DL solutions do not directly provide a quantitative measure of the diagnostic condition, and are inherently incapable of direct measurement of specific features. Particular image-processing techniques are required for such measurements.

Use of DL networks for diagnosis with a single-label per image for training (image-classification methods) can cause the network to associate incorrect information with the presence of the disease. As a specific example, if images from a subgroup of people with higher prevalence of DME were captured with the optic disk slightly temporal, the DL network will incorrectly learn to associate temporal-location of the disk with the presence of DME.[33] Such approach to application of DL networks is prone to bias and potential errors. We believe that the best utilization of such DL networks is where they are coupled with specific masks (black and white annotation) of the pathological conditions (semantic-segmentation methods), that can lead to detection and segmentation of the diagnostic clues, as would a specialist, and consequently providing a classification based on those findings.

It is important to remember that current medical knowledge is derived from decades of observational data gathering, hypothesizing, and validation through clinical research, so one cannot simply take the findings of a DL system as the absolute in accuracy, without realizing, assessing, and validating the findings to be consistent with our medical knowledge. To date, incorporation of clinical experience and published literature into DL solutions remains sparse. Inclusion of this clinical knowledge will need to take place to fully realize the potential benefits, as well as possible drawbacks of DL. We believe that best patient care is realized through a strong doctor-patient relationship that is built on trust and compassion. DL solutions cannot replace such relationship, but rather must complement it.

Conclusion

DL is an emerging tool that can provide benefits to both patient and clinician. There are limitations with the current technology, and questions remain on its best applications. As DL technology evolves it will become more integrated into ophthalmic care, freeing the clinician from repetitive tasks and allowing them to focus on improved patient care. DL will allow Ophthalmologists to focus their resources on patient relationships as well as optimizing medical and surgical care.

Online-only material

This article includes online-only material. Appendix 1 can be found on the CJO website. It is linked to this article in the online contents.

Appendix

Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jcjo.2018.04.019.

References

1. Knight W. Finally, a Driverless Car with Some Common Sense. MIT Technology Review. www.technologyreview.com/s/608871/finally-a-driverless-car-with-some-common-sense/. Published 2017. Accessed December 30, 2017.
2. Lai M. Giraffe. *Using Deep Reinforcement Learning to Play Chess.* September, 2015.
3. Marr B. 4 Mind-Blowing Ways Facebook Uses Artificial Intelligence. Forbes.
4. Lee A, Seattle M, Taylor P, Kingdom U. Machine learning has arrived! *Ophthalmology.* 2017:1726-8.
5. Rangayyan R. *Biomedical Image Analysis*. Boca Raton, FL: 2005.
6. Duda R, Hart P, Stork D. *Pattern Classification*. 2nd ed. New York, NY: Wiley; 2001.
7. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern.* 1980:193-202.
8. Szegedy C, Liu W, Jia Y, et al. *Going deeper with convolutions.* 2014.
9. Simonyan K, Zisserman A. *Very deep convolutional networks for large-scale image recognition.* 2014.
10. Badrinarayanan V, Kendall ACR. A deep convolutional encoder-decorder architecture from image segmentation. *SegNet.*
11. Eyenuk. www.eyenuk.com/. Published 2018. Accessed February 2, 2018.
12. Vision Quest i-Rx. visionquest-irx.com/. Accessed February 2, 2018.
13. Retmarker. www.retmarker.com/. Accessed February 2, 2018.
14. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans Biomed Engeneering.* 2015:2693-701.
15. Liu X, Jiang J, Zhang K, et al. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. *PLoS One.* 2017:1-18.
16. Chen X, Xu Y, Wing D, Wong K. Glaucoma detection based on deep convolutional neural network. *Conf Proc IEEE Eng Med Biol Soc.* 2015:715-8.
17. Muhammad H, Fuchs TJ, Cuir NDe, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma.* 2017.
18. Devalla SK, Chin KS, Mari J, et al. A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head. *Invest Ophthalmol Vis Sci.* 2018:63-74.
19. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology.* 2016:1974-80.
20. 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.* 2017:1170-6.
21. Treder M, Lauermann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. *Graefe's Arch Clin Exp Ophthalmol.* 2018:259-65.
22. Shan J, Li L. A deep learning methods for microaneurysm detectin in fundus images. *IEEE First Conf Connect Heal.* 2016:357-8.
23. Prenta P, Lon S. Detection of exudates in fundus photographs using convolutional neural networks. *Intnerational Symp Image Signal Process Anal.* 2015:188-92.
24. Grinsven MJJPV, Ginneken BV, Hoyng CB, Theelen T, Sánchez CI. Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images. *IEEE Trans Med Imaging.* 2016:1273-84.
25. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology.* 2017:962-9.
26. 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.* 2016:2202-410.
27. Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express.* 2017:530-3.
28. Roy AG, Conjeti S, Karri SPK, et al. ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed Opt Express.* 2017:111-8.
29. Fang L, Cunefare D, Wang C, Guymer RH, Li S, Farsiu S. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomed Opt Express.* 2017:2732-44.
30. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. *Ophthalmology.* 2018.
31. Shu D, Ting W, Cheung CY, 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.* 2017:2211-23.
32. Wikipedia. Tay (bot). en.wikipedia.org/wiki/Tay_(bot). Accessed April 21, 2018.
33. Abramoff MD, Lou Y, Erginay A, et al. Improved automated setection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci.* 2016:5200-6.

Footnotes and Disclosure:

The authors have no proprietary or commercial interest in any materials discussed in this article.

From the *Department of Ophthalmology and Visual Sciences, University of Alberta, Edmonton, Alta.; †Aurteen Inc., Calgary, Alta.

Correspondence to: Matthew T.S. Tennant, MD, Alberta Retina Consultants Suite 400, 10924 107 Avenue, Edmonton, Alta: mtennant@ualberta.ca