The use of deep learning technology for the detection of optic neuropathy

Mei Li¹, Chao Wan²^  

¹Department of Ophthalmology, Yanan People's Hospital, Yanan, China; ²Department of Ophthalmology, the First Hospital of China Medical University, Shenyang, China

Contributions: (I) Conception and design: C Wan; (II) Administrative support: C Wan; (III) Provision of study materials or patients: Both authors; (IV) Collection and assembly of data: Both authors; (V) Data analysis and interpretation: M Li; (VI) Manuscript writing: Both authors; and (VII) Final approval of manuscript: Both authors.

Correspondence to: Chao Wan. Department of Ophthalmology, the First Hospital of China Medical University, 155 Nanjing North Street, Heping District, Shenyang 110001, China. Email: hundred2005@126.com.

Abstract: The emergence of computer graphics processing units (GPUs), improvements in mathematical models, and the availability of big data, has allowed artificial intelligence (AI) to use machine learning and deep learning (DL) technology to achieve robust performance in various fields of medicine. The DL system provides improved capabilities, especially in image recognition and image processing. Recent progress in the sorting of AI data sets has stimulated great interest in the development of DL algorithms. Compared with subjective evaluation and other traditional methods, DL algorithms can identify diseases faster and more accurately in diagnostic tests. Medical imaging is of great significance in the clinical diagnosis and individualized treatment of ophthalmic diseases. Based on the morphological data sets of millions of data points, various image-related diagnostic techniques can now impart high-resolution information on anatomical and functional changes, thereby providing unprecedented insights in ophthalmic clinical practice. As ophthalmology relies heavily on imaging examinations, it is one of the first medical fields to apply DL algorithms in clinical practice. Such algorithms can assist in the analysis of large amounts of data acquired from the examination of auxiliary images. In recent years, rapid advancements in imaging technology have facilitated the application of DL in the automatic identification and classification of pathologies that are characteristic of ophthalmic diseases, thereby providing high quality diagnostic information. This paper reviews the origins, development, and application of DL technology. The technical and clinical problems associated with building DL systems to meet clinical needs and the potential challenges of clinical application are discussed, especially in relation to the field of optic nerve diseases.

Keywords: Deep learning (DL); optic nerve; artificial intelligence (AI); fundus image; optical coherence tomography (OCT)

doi: 10.21037/qims-21-728
View this article at: https://dx.doi.org/10.21037/qims-21-728

^ ORCID: 0000-0002-7414-0773.
Introduction

Optic nerve abnormalities are a common presentation in ophthalmology clinics. Diagnosis and monitoring the progression of the disease requires visual assessment of the shape of the optic disc, which traditionally relies heavily on the professional experience and knowledge of the doctors. Although trained ophthalmologists can easily identify most optic nerve abnormalities, emergency physicians, neurologists, general practitioners, and other non-ophthalmic health care providers, who also have an important role in screening eye diseases, often have limited ability and experience in using an ophthalmoscope to examine the optic disc. This can lead to high rates of misdiagnoses and missed diagnoses of optic disc abnormalities. Therefore, intelligent tools that can screen and manage medical data safely and effectively, and provide scientific and high-quality diagnoses are urgently needed.

Recent advances in artificial intelligence (AI) and the collation of large medical data sets have stimulated great interest in the development of deep learning (DL) algorithms. Compared with subjective evaluation and other traditional methods, DL algorithms can identify optic nerve diseases faster and more accurately in diagnostic situations (1). This paper critically details the latest applications of DL models in optic neuropathy, discusses their advantages and limitations, and focuses on the inherent challenges of such models in screening, diagnosis, and progression detection. After a brief overview of DL and a comparison with traditional machine learning classifiers, the application of DL models in optic nerve-related clinical detection will be discussed.

AI, machine learning, and DL

AI is an important branch of computer science. It develops, studies, simulates, and extends human intelligent behavior (2). Machine learning is a crucial supporting technology for the realization of AI. It obtains rules from data analysis, establishes specific algorithms to identify patterns in data, and uses formulas to predict new samples (3).

In the 1980s, the development of machine learning algorithms rapidly escalated. The most successful early results were obtained through the statistical method of machine learning (4). This includes linear and logistic regression, classification, k-nearest neighbor, decision tree, random forest, kernel-based methods such as support vector machines (SVMs), and the like (5). In previous studies, SVMs have been used to skillfully combine the measurement of the optic disc edge and the optic cup area to achieve a glaucoma classification (6,7). However, although SVMs and other traditional algorithms performed satisfactorily in the field of problem-solving, these approaches cannot guarantee the improvement of human expertise through knowledge engineering technology. In fact, although many traditional machine learning technologies have been able to simulate the diagnosis of glaucoma and perimetry, they still need common sense information to solve many practical situations in the clinic. Subsequently, DL has been shown to be an effective method for constructing and training neural networks to solve complex problems. These algorithms automatically extract features to form solutions with complex perception and understanding, which can be copied and reused in combination with components from other solutions. One of the main benefits of DL is that it simplifies the requirements for professional knowledge (8).

People can apply the original data directly to the DL model instead of manually identifying the relevant features from the data. These functions may be more efficient than those that are manually managed and may be close to or beyond the level of human capability. However, as a trade-off, the features of automatic learning may not be easy to understand or explain.

The DL model is a type of artificial neural network which is composed of multiple layers of artificial neurons (9). It involves a huge data set and faster hardware. The data is input into the network and processed to achieve the desired results. Its performance can be enhanced by changing internal parameters repeatedly using back propagation algorithms. DL can process more complex data and improve the algorithm of neural network performance, generalization ability, and cross-server distributed training ability, all of which are superior to the performance of shallow artificial neural networks (10). With the help of well-labeled, large data sets, DL can be used for accurate classification of medical images, and it has shown unique advantages in a variety of medical disciplines, including ultrasound (11), dermatology (12,13), pathology (14), radiology (15,16), and ophthalmology. In the field of ophthalmology, the DL system (DLS) has been developed for the detection of various diseases, such as diabetic retinopathy, age-related macular degeneration, glaucoma, and retinopathy of prematurity and cardiovascular diseases. These DLS are often comparable to the performance of human clinicians (17).
Optic nerve in fundus photographs

Identification, location, and extraction of the optic disc

Accurate positioning and segmentation of the optic disc on the fundus image is essential for the detection of optic neuropathy and other lesions. In traditional machine learning, there are three methods for locating the optic disc (18). The first is based on the characteristics of the optic disc; the second is based on the information of the vascular structure; and the third is the combination of the above two methods. The method based on the characteristics of the optic disc mainly relies on identification of an area with high brightness, approximate circular shape, and large internal gray contrast to locate the optic disc. This method is completely dependent on the characteristics of the optic disc itself, and thus, when there are lesions in the image that can lead to changes in the brightness and shape of the optic disc, the accurate identification of the optic disc cannot be guaranteed. The method of locating the optic disc based on the vascular structure relies on locating the optic disc and the blood vessels converging on the optic disc, as well as the characteristics of those blood vessels. This latter method improves upon the first method to a certain extent and has good robustness for fundus lesions. However, it depends on the accuracy of the vascular segmentation and it may be difficult to locate the optic disc accurately when the distribution of the vascular network is irregular or when the integrity of the vascular network is damaged due to the lesions.

In the field of DL, there are two types of research related to the optic disc: one is related to the detection of the optic disc in the whole fundus image (19) and the other is related to the classification of the optic disc in the cropped image (20). While there has been much research on the optic disc classification algorithm, the target detection of the optic disc has not been thoroughly studied (21).

At present, there are some limitations in the use of DL for optic discs. The optic disc is generally used as the target object in a photo of the fundus, and while it is not large, it is very prominent and contains rich features. However, even the most advanced object detection algorithms have difficulty detecting small defects. For example, in most fundus photos, the targets of interest, such as microaneurysms, bleeding points, hard exudation, cotton-wool spots, and drusen, are relatively small, occupying only a small area of the image and thus, may be easily overlooked and difficult to capture (22). Although these lesions are small, their clinical risk and potential for harm can be significant. In addition, fundus images contain lesions that are morphologically diverse, such as hemorrhage, epiretinal membrane, and atrophy. The influence of comorbidity pathologies on the function of these algorithms is unclear as most of the algorithms have been trained and tested on data sets that exclude other eye conditions (such as retinal diseases and high myopia) (23). In addition, the problem with identifying the optic disc using Single Shot Multi-Box Detector and You Only Look Once is that the bounding box size is not uniform. To overcome this limitation, Rogers et al. (24) used the cloud-based AI system Pegasus to evaluate fundus photography using a set of convolutional neural network (CNN), with each dedicated to different tasks such as identifying special marks (optic disc, macula) and different types of fundus pathological lesions. The purpose of this design was to extend to any fundus photo that contains an optic disc. First, one CNN is used to recognize and extract the optic disc, and then the standardized format is input into another CNN to classify it. When developing a disease detection model, it is necessary to standardize the resolution of the images. After cropping the bounding box area, the resolution of the other images must be adjusted to the lowest resolution image, rendering it more efficient to cut out the images around the optic disc with the same resolution.

Segmentation of the optic cup and the optic disc

The regions of the optic disc and cup contain key pixel features. A large cup to disc ratio (CDR) is an important manifestation of optic nerve damage caused by glaucoma. Accurate calculation of the CDR is the basis of glaucoma screening (25). According to the method of DL, the minimum bounding box of the two regions is identified, the optic disc and optic cup regions are well segmented, and the diameter of the optic cup and the optic disc region is measured to calculate the CDR.

At present, optic disc segmentation is roughly divided into three categories (26). The first is based on the region growing method (27) where appropriate seed points are set in the optic disc. The region growing operation is then performed on the seed points according to the brightness and edge characteristics of the optic disc to obtain the optic disc contour. The key objective of this method is to determine the initial seed point, which is usually based on the optic disc positioning and is thus limited to the accuracy of the optic disc positioning. The second method is based on the template method, which uses a circle or ellipse to fit the boundary of the optic disc. Alternatively, the Hough transformation is used to determine the boundary of the...
optic disc (28). However, since the optic cup is also shown as a bright circular area, this method can result in incomplete segmentation of the optic disc. In addition, some lesions (such as hard exudation) with similar brightness and shape as the optic disc will also render this method unsuccessful. The third method is based on the active contour model, which first determines the initial contour of the optic disc and then obtains the boundary of the optic disc by using the continuous evolution of the contour driven by external constraints and internal energy (29).

Usually, the traditional machine learning method and the DL method are combined to segment the visual disc. A novel optic disc sensing integration network for automatic glaucoma screening was proposed by Fu et al. (30). This method integrates the whole-field fundus photo and the deep context of the local optic disc area. The depth flow of the four different layers and modules are known as global image flow, guidance network of segmentation, local disc area flow, and disc pole transform flow. Finally, the output probability of different flows is integrated into the filter results. Using this method, experiments were carried out on two different glaucoma datasets, and the network was found to be superior to other most other advanced algorithms. Wang et al. (31) put forward a DL framework ranging from coarse to fine, based on the classical CNN, called the U-net model, to accurately recognize the optic disc. The network was trained on the fundus image and the gray-scale vascular density map, thereby generating two different segmentation results. The combined results from the overlapping strategies were used to identify the local image blocks (disk candidate regions) and input into the U-net model for further segmentation. Based on the collected data set and 2,978 test images from six common data sets, the developed framework achieved 89.1% and 93.9% mean intersection over union and dice similarity coefficients, compared to 87.4% and 92.5% obtained by the unique U-net model, respectively. The contrast with the available methods showed that the proposed DL framework is reliable and has relatively high performance in automatic optic disc segmentation.

Ko et al. (32) evaluated and detected the performance of a CNN framework for vertical CDR (VCDR) classification using 944 fundus photographs, including 465 non-glaucomatous optic neuropathy (NGON) eyes and 479 glaucomatous optic neuropathy (GON) eyes. Based on stratified sampling, the images were divided into a training set consisting of 763 images and a test set consisting of 181 images. The accuracy, sensitivity, and specificity of the CNN classifier were 95.0%, 95.7%, and 94.2%, respectively, compared with 92.8% accuracy obtained with the ensembles model. The area under the curve (AUC) was 0.992. Recently, efforts have been made to use DL algorithms to detect and classify glaucoma. Gao et al. (33) propose a network called the recurrent fully convolution network (RFC-Net), which is used to segment the optic disc and the optic cup automatically and capture more advanced data and subtle edge information. The RFC-Net can minimize the loss of spatial information. The effectiveness and generalization of the proposed method was evaluated using the DRISHTI-GS1 data set. By designing effective data preprocessing, the number and diversity of training samples could be increased such that the training set contained 500 images and the test set consisted of 510 images. Compared with the original full convolutional network method and other advanced methods, the proposed method achieved better segmentation performance.

Park et al. (21) studied and compared the performance of the most advanced DL architecture in detecting optic disc and VCDR in fundus images. The training data set was composed of 1959 eyes with normal fundus, glaucoma, and other optic neuropathy (in which VCDR >0.4 accounted for 94.3%) which were randomly divided into a training data set and a verification data set at a ratio of 9:1. A total of 204 eyes (95 healthy and 109 glaucoma-affected eyes) were used as the measurement data set. Three different architectures, namely, You Only Look Once V3, ResNet, and DenseNet, were compared in terms of accuracy of detection, processing time, diagnostic performance, graphics processing unit (GPU) effect, and image resolution. It was found that with an increase in the input image distinguishability, the accuracy of grading, positioning deviation, and diagnostic performance were modified. In addition, the optimal architecture varied with distinguishability. Thus, when balancing speed and accuracy, the choice of architecture may depend on the aims of the researcher.

In addition to the training data set, a perfect generator can provide significant annotated data as a supplementary segmentation information network. This can improve the final predicted segmentation results. Most of the existing DL techniques cannot achieve ideal segmentation performance because a large amount of pixel-level annotation data cannot be obtained during the period of training. To overcome this limitation, Liu et al. (34) proposed an effective and efficient joint segmentation method on the optic disc and the optic cup based on a semi-supervised conditional generative adversarial network.
model and preserve the detailed contour of the small
accelerate the convergence of a small target segmentation
was used to effectively
calculate the CDR. In the training process, a network based
optic disc and optic cup in the fundus images to accurately
sampled from the ORIGA-650 data set for training and 131
the ORIGA-650 dataset, of which, 519 were randomly
achieved with the ORIGA-650 and the REFUGE data
segmentation results for the optic disc and the optic cup
between the fundus image and the relevant segmentation
Marked and unmarked data were used to improve
the performance of the segmentation. The most advanced
results for the optic disc and the optic cup
were achieved with the ORIGA-650 and the REFUGE data
sets. Jiang and colleagues (35) considered that the blood
vessels passing through the optic disc region in the fundus
image would affect the detection of the boundary box.
They proposed to remove the blood vessels in advance to
further improve the overall performance, assumed that the
shape of the optic disc and optic cup region was an ellipse,
and applied the DL method to achieve the most advanced
segmentation performance of the optic disc and optic cup
on the ORIGA-650 database.

However, most methods regard the division of the
optic cup and optic disc as two independent tasks and do
not fully consider the clear relationship between them.
By considering the characteristics of the optic cup in the
optic disc and the position between the optic cup and the
optic disc, segmentation of the optic disc and cup can be
achieved accurately. Bian et al. (25) proposed a depth neural
network based on anatomy to implement the segmentation
of fundus images. They also used the 400 images from the
REFUGE official training data set and 650 images from
the ORIGA-650 dataset, of which, 519 were randomly
sampled from the ORIGA-650 data set for training and 131
for validation. The model could accurately segment the
optic disc and optic cup in the fundus images to accurately
calculate the CDR. In the training process, a network based
on cascaded attention mechanisms was used to effectively
accelerate the convergence of a small target segmentation
model and preserve the detailed contour of the small
target accurately. The method was verified in the MICCAI
REFUGE fundus image segmentation contest, and the
dice score was 93.31% in optic disc segmentation and
88.04% in optic cup segmentation. It can also be used to
realize manual and semi-automatic interactive segmentation
schemes and automatic segmentation algorithms. However,
there are some limitations to this method. In contrast to
the one-stage method, which divides the optic disc and the
optic cup, the two-stage method proposed by the author is
not sufficiently efficient. The end-to-end model proposed
does not perform preprocessing and post-processing, and
as such it saves a lot of time. In addition, the prediction
of optic disc in this study may affect the results of the optic
cup. If the division of the optical cup in the first stage is
incorrect, erroneous information may be provided to the
second stage.

Due to the significant overlap between the optic cup
and the neuroretinal marginal region, it is a challenge to
obtain the CDR value automatically with high accuracy and
robustness. Based on a semi-supervised learning scheme,
Zhao et al. (36) proposed a direct CDR estimation method
which uses a CNN called MFPPNet for unsupervised
feature representation of the fundus image and random
forest regression to obtain the CDR value. Bypassing the
middle segmentation allows the CDR value to be directly
regressed based on the feature representation of the optic
disc through DL technology. This method was verified
using the glaucoma data sets Direct-CSU and ORIGA-650,
achieving a lower average CDR error of 0.0563 and an AUC
of 0.905 on the data set of 421 fundus images. Calculating
the CDR value enables advanced CDR estimation and
satisfactory glaucoma screening to be performed.

However, it should be noted that there are still significant
challenges before these algorithms can be applied to the
complex real world. For example, the photographic quality
of color fundus images is not uniform. If DL algorithms
are trained with photos captured by different cameras, the
performance may be poor. Moreover, the current diagnostic
performance results are only based on VCDR and it should
be noted that during the diagnosis of glaucoma, VCDR is
not the only factor used by ophthalmologists to identify an
abnormal optic disc. The optic disc of the fundus image can
also provide extra information associated with glaucoma
diagnosis, such as the sign of lamina punctate, vascular
system bayonet, retinal nerve fiber layer (RNFL) defect,
opptic disc hemorrhage, optic disc tilt, and peripapillary
atrophy. Therefore, strictly speaking, the purpose of
current research is not to evaluate glaucoma diagnostic
performance, but rather, to evaluate the performance of target recognition and classification algorithms represented by VCDR. Nonetheless, the ability to diagnose glaucoma using only VCDR is an important tool for health and allied health professionals.

**Application of DL in optic neuropathy of fundus images**

**GON**

Yang et al. (37) developed an AI classification algorithm called the ResNet-50 that aims to evaluate the efficacy of the DL method in distinguishing NGON from GON by using image recognition. Analysis of 3,815 fundus images showed 93.4% sensitivity and 81.8% specificity. The average precision-recall AUC was 0.874, which showed excellent performance.

In a retrospective single-center study, Al-Aswad et al. (38) graded 110 color fundus photographs using the DLS Pegasus. The consistency between Pegasus and the gold standard was 0.715, while the highest consistency between ophthalmologists and the gold standard was 0.613. In fact, the Pegasus achieved better results than 5 of the 6 ophthalmologists in terms of diagnostic performance. There was no statistically significant difference between Pegasus and the ‘best case’ consensus between the ophthalmologists.

There are some limitations to the use of DLS for the identification of GON. Li et al. (39) collected 48,116 fundus photos retrospectively and developed a DLS for the automatic classification of GON on color fundus photos. The DLS had an AUC of 0.986, sensitivity of 95.6%, and specificity of 92.0%, demonstrating that the DLS could detect reference GON with high sensitivity and specificity. However, the most common causes of false-negative and false-positive results occurred when a physiological large optic cup or the optic disc of a pathological myopia patient could not be accurately identified.

Moreover, in clinical practice, poor patient cooperation, small pupil, or refractive medium opacity will limit the performance of optical disc recognition. Ha et al. (40) proposed a DL method to improve the resolution and readability of fundus optic disc photos through contrast, color, and brightness compensation. Each high-resolution original fundus photo was converted into two counterparts, a reduced ‘low-resolution fundus photo’ and a ‘compensated high-resolution fundus photo’. This was produced using a customized image post-processing algorithm to enhance the visibility of the optic disc edge and the surrounding retinal vessels. Super resolution GAN (SR-GAN) is then used to learn the difference between the two counterparts. Finally, by inputting high-resolution photos into SR-GAN, ‘enhanced fundus optic disc photos’ can be obtained with four-fold magnification and overall color and brightness conversion.

**Abnormalities of the optic disc**

In the past few decades, a large number of fundus photographs have been collated in the screening network for diabetic retinopathy (41). Through DL, these data sets have been used to train automatic systems to detect diabetic retinopathy and other common pathologies (42). However, challenging cases such as papilledema and anterior ischemic optic neuropathy (AION) can limit the use of these automatic detection systems. This is because standard DL requires too many examples of these conditions. Quellec et al. (43) proposed a new, small sample learning framework based on a t-SNE visualization tool. The framework extends the CNN to detect rare cases in fundus images, such as optic disc edema, by learning with only a small number of samples. Experiments on 164,660 screening data sets from the OPHTDIAT screening network showed that an AUC greater than 0.8 was achieved in 37 of 41 cases (average AUC was 0.938). Therefore, this framework is superior to other frameworks in detecting rare cases. These richer predictions will facilitate the adoption of automatic ophthalmic pathological screening, which will completely change the clinical practice of ophthalmology.

Liu et al. (44) presented their research on high-performance DLS for semantic labeling of neuro-ophthalmic images using small data sets. The system modified the ResNet-152 deep CNN that was pretrained on ImageNet, which can distinguish between normal and abnormal optic discs and detect the optic disc laterality of the right and left eyes. The system could detect a variety of neuro-ophthalmic diseases, including but not limited to atrophy, AION, hypoplasia, and optic papilledema.

Bioussé et al. (45) compared the diagnostic performance of AI DLS and neuro-ophthalmologists in the classification of optic disc appearance. Evaluated on 800 randomly presented fundus images without clinical information, this DLS performs at least as well as two professional neuro-ophthalmologists in classifying optic disc abnormalities.

Optic disc edema is common in clinical practice, especially in patients with neurological diseases. It is often used to determine the presence of intracranial hypertension and to evaluate the progression of the disease. Due to the lack of ophthalmologists in many departments, it is often
necessary to use machine learning to evaluate optic disc edema. Ahn et al. (46) used machine learning to distinguish between patients with optic neuropathy, pseudopapilledema (PPE), and normal subjects. A model was designed and compared with the 3 most commonly used machine learning classifiers, GoogleNet Inception V3, 19 layers of super deep convolution network from visual geometry group, and 50 layers of deep residual learning (ResNet). The accuracy and area under the receiver operating characteristic curve (AUROC) were analyzed. It was concluded that machine learning technology could distinguish between PPE and oedema optic disc of optic neuropathy.

Milea et al. (47) trained, validated, and externally tested a DLS to classify the optic disc as normal, with papilledema, or with other abnormalities using 15,846 retrospectively collected fundus photographs. Using the clinical diagnosis of neuro-ophthalmologists as the reference standard, it was demonstrated that the DLS could distinguished between papilledema, a normal optic disc, and an abnormal optic disc without papilledema. Notably, photos obtained after mydriasis were more conducive to accurate recognition. Vasseneix et al. (48) trained a DLS to accurately classify the severity of papilledema in 2,103 mydriatic fundus photos of patients with elevated intracranial pressure, and the results were comparable to that of independent neuro-ophthalmologists.

As the structure of the optic disc is different between the nasal and temporal sides, the direction of the optic disc and whether the pathological lesion site is unilateral or bilateral are vitally important clinical data which can affect the diagnosis and determine the relative imaging and/or systematic evaluation. Several studies have shown that DLSs can accurately recognize the right or left eye in photos with optic discs (44,46). Furthermore, it can distinguish optic disc edema from normal optic disc with an average accuracy of 93%, and distinguish real optic disc swelling from pseudo-swelling with an accuracy of about 95%.

**Optic disc and systemic diseases**

Most DL studies on fundus photography are aimed at screening for retinal diseases and glaucoma. However, studies have shown that the information extracted from retinal fundus images can also be used to predict cardiovascular risk factors. Poplin et al. (49) trained data from 284,335 patients and validated the DL model on 2 independent data sets of 12,026 and 999 patients. The trained DL model used the anatomical characteristics of the optic disc to predict cardiovascular risk factors previously thought to be non-existence or those that could not be quantified in retinal images, such as age (mean absolute error within 3.26 years), gender (AUC =0.97), smoking status (AUC =0.71), systolic blood pressure (mean absolute error within 11.23 mmHg), and major adverse cardiac events (AUC =0.70).

Although we can already apply DL to the recognition of fundus optic neuropathy, it may be a challenging task to deeply understand how the machine classifies fundus photos into normal or diseased states. Therefore, further research on visualization convolution layers and filters are needed to understand how the machine classifies images, and larger data sets may be needed to help verify existing findings.

**Optic nerve in optical coherence tomography (OCT)**

**Positioning of the optic disc**

Thompson and colleagues (50) used the spectral-domain optical coherence tomography (SD-OCT) Bruch’s membrane opening minimum rim width (BMO-MRW) parameter as the reference standard for optical disc photo marking. BMO-MRW may be particularly effective in grading optic discs in challenging images. For example, in cases of high myopia, DL prediction is highly correlated with the actual BMO-MRW. The AUC for distinguishing glaucoma from normal eyes was 0.945 as predicted by DL and 0.933 as assessed by actual measurement (P=0.587). Similarly, the class activation diagram confirmed that the edge of the neuroretina is very important to the classification of the algorithm.

There are some limitations in optical disc positioning in OCT. In the case of optic disc swelling, segmentation of projected retinal vessels from the volume of the OCT is challenging due to shadow artifacts caused by swelling. Islam et al. (51) proposed that the vascular information from multiple projected retinal layers can significantly improve vascular visibility and designed a method based on DL to segment vessels, which involves simultaneously using three front images of the OCT as input. In the case of optic disc swelling, using multiple frontal images achieved better vascular segmentation compared with the traditional method using a single front image.

**CDR**

To train neural networks with objective VCDR values, the
optic disc parameters measured with Cirrus high-definition OCT (HD-OCT) (Carl Zeiss Meditec, Dublin, CA) can be used for calculations. Since optic disc regions can change up to five times, there is no absolutely defined pathological VCDR. However, Mwanza et al. (52) reported that VCDR showed good glaucoma diagnostic performance using Cirrus HD-OCT (AUROC ≥0.951), and that there was no statistical difference between average and clock hour RNFL thickness (AUROC =0.950 and 0.957, respectively). In addition, another study demonstrated that the optic disc parameters of Cirrus HD-OCT for glaucoma showed brilliant reproducibility. In particular, the intraclass correlation coefficients of VCDR were ≥0.921 (53). Studies have compared Cirrus HD-OCT with confocal laser ophthalmoscope and reported a strong correlation between the two modes of VCDR (54,55). Although the optic disc parameters of HD-OCT and confocal laser ophthalmoscope are comparable in the diagnosis of glaucoma, the parameters cannot be used interchangeably. Moreover, there may be errors in the parameters measured by OCT (56). However, Cirrus HD-OCT can provide highly reliable and repeatable VCDR data for neural networks, and it is expected that these data may be more accurate than the VCDR measured by ophthalmologists (57).

RNFL

SD-OCT has become the most widely used tool in the diagnosis of glaucoma structural damage. Measurement of the optic disc and the RNFL is commonly used to monitor disease diagnosis and progression in clinical practice. However, the traditional SD-OCT structural damage assessment requires to segment the region of interest to extract measurements, such as the RNFL thickness. This measurement process is completed by the OCT software, but tends to be inaccurate. The results showed the presence of segmentation errors and artifacts in the RNFL scanning, and the average probability was 0.90±0.17 and 0.12±0.22, respectively (58). In busy clinical practice, although it is feasible to correct segmentation errors by manual inspection, the actual operation is time-consuming and very difficult. Another challenge is the interpretation of the SD-OCT scanning, which requires the analysis of many parameters and different regions and it is not easy for clinicians to integrate all the different parameters from the RNFL thickness, the topographical parameters of the optic disc, and the macular assessment. The presence of many parameters increases the chance of making type I errors, that is, accidental anomalies. This led to the concept of ‘red disease’, that is, the incorrect diagnosis of glaucoma based on red results obtained from one or several parameters of the SD-OCT scan, without other confirmed clinical features (59).

Considering these limitations of OCT interpretation, a DL model can provide an alternative method to quantify the pathologic changes and structural damage without relying on the features defined by the automatic segmentation software. As mentioned above, provided there is enough data, the DL algorithm can automatically learn features from it. Therefore, these models can use the original SD-OCT images without inputting predefined features. Along with these ideas, instead of the traditional method of calculating RNFL by quadrant, a DL algorithm without segmentation can be trained to predict RNFL thickness. No segmented prediction is highly correlated with conventional RNFL thickness, and the average absolute error of a high-quality image is about 2 μm. Importantly, the DL model can extract reliable RNFL thickness information from images with conventional segmentation failure. To avoid errors in the traditional segmentation of RNFL, which may affect the accuracy of SD-OCT in detecting glaucoma damage, Thompson et al. (60) developed a segmentation-free DL algorithm for evaluating glaucoma damage using full circle B-scan images from SD-OCT. A CNN is trained to distinguish glaucoma from normal eyes without a segmentation line by using SD-OCT circular B-scan. The segmentation-free DL algorithm is superior to the traditional RNFL thickness parameter in the diagnosis of glaucoma damage in OCT scanning, especially in the early stages of the disease. Petersen et al. (61) also reported that the DL model for detecting glaucoma using non-segmented SD-OCT performed better than the RNFL thickness parameters extracted by automatic segmentation.

Some reports have examined the role of OCT in diagnostic decisions when combined with other relevant clinical information. The machine-to-machine (M2M) DL algorithm trained by Jammal et al. (62) using RNFL thickness parameters of SD-OCT was applied to a subset of 490 fundus photographs obtained from 370 subjects. The Spearman correlation between human classification and RNFL thickness predicted by the M2M DL was compared with the global index of standard automatic perimetry (SAP). The AUC and the partial AUC of the clinical significance specific area (85~100%) were used to compare the ability of each output to distinguish between eyes with a repeatable glaucoma SAP defect and eyes with a normal visual field.
The algorithm was trained to quantify RNFL damage in the fundus photographs. The performance of the M2M DL algorithm in detecting eyes with repetitive glaucoma visual field loss was comparable or even superior to that of human diagnostic performance. Indeed, this DL algorithm may replace glaucoma screening by health or allied health professionals. In the M2M model, the DL algorithm is used to train the color fundus photos and label them with the objective quantitative reference standards. The RNFL thickness is measured by SD-OCT.

When evaluating color fundus photos, the pretrained M2M DL model is used to efficiently predict the thickness of RNFL and thus, the damage of glaucoma can be quantified. Medeiros et al. (63) proposed a new method using quantitative SD-OCT data to train the DL algorithm to quantify the structural damage of glaucoma in optic disc photos. The DL CNN was trained to evaluate the optic disc images and predict the average RNFL thickness of SD-OCT. There was a significant correlation between the RNFL thickness measurement value predicted by the DL model and the actual value measured by SD-OCT (r=0.832; P<0.01). The performance of the M2M model and SD-OCT RNFL thickness model in differentiating glaucoma from normal eyes was similar (the AUCs were 0.940 and 0.944, respectively; P=0.724). For 95% specificity, the sensitivity of the predictive measure was 76%, while the sensitivity of the actual SD-OCT measure was 73%. For 80% specificity, the predictive measurements were 90% sensitive, while the actual SDO-CT measurements were 90% sensitive. The photo-trained DL network had an overall accuracy of 83.7% in replicating such classifications. Visual class activation maps or heat maps were used to highlight the most important areas in the images to predict the RNFL damage in the real image used for prediction, thereby allowing more detailed analysis of the prominent areas.

The optic disc pit (ODP), first described by Wiethe in 1882, is a round or oval cavity or depression in the optic disc (66). It is a rare ophthalmic finding with an incidence ranging from about 1 in 500 to 1 in 11,000. While in many cases, ODP can remain clinically asymptomatic, visual field defects can be very severe and in extreme cases, it can result in complete blindness. Between 25% to 75% of ODP patients develop disc depression-associated maculopathy or ODP maculopathy, accompanied by retinoschisis, inner retinal atrophy, serous macular detachment, and significant visual loss. Before the development of OCT, it was impossible to image ODP in vivo. Maertz et al. (67) proposed a new and promising technique, namely, ultrahigh-speed swept-source megahertz-OCT (SS-MHz-OCT) that can be used to display the optic papilla imaging of ODP or ODP maculopathy eyes by a-scans at 1.68 million per second. 3D volume reconstruction of the ODP and high-resolution images from a single densely sampled MHz-OCT data set can be used to generate a 3D rendering of the optic disc region to study ODP characteristics.

However, it is worth noting that in general, a class activation map does not have sufficient resolution to accurately locate small areas related to classification. The lack of accuracy is due to the convolution layer construction of the DL model, which leads to the decline of final layer sampling. In addition, the efficiency of class activation maps is largely determined by the model used and the quantity and quality of available training data. Although DL algorithms may indeed capture information that is not evident to the human eye, it is necessary to be aware of the resolution limitations of these class activation maps.

**Stereoscopic optic disc image**

A study by Maetschke and colleagues (65) developed a DL algorithm which can distinguish glaucoma eyes from healthy eyes by using the original and undivided OCT volume of the optic disc. The performance of the algorithm was superior to the traditional SD-OCT parameter. The AUC was 0.94, while the AUC of a logistic regression model combined with SD-OCT parameters was 0.89. The class activation map appeared to highlight the areas in the volume of OCT which have been clinically identified as important areas for the diagnosis of glaucoma, especially the retinal nerve margin, optic cup, lamina cribrosa, and its surrounding areas. The class activation map provides a better understanding of the CNN by highlighting the areas causing high activation in the real image used for prediction, thereby allowing more detailed analysis of the prominent areas.

The optic nerve in computed tomography and magnetic resonance (MR) images

Head and neck (HaN) cancer (HNC), including oral cancer,
Salivary gland cancer, sinus and nasal cancer, nasopharyngeal cancer, oropharyngeal cancer, hypopharyngeal cancer, and laryngeal cancer are amongst the most common types of cancer worldwide (68). Radiotherapy is one of the basic treatment methods for HNC (69). To minimize complications after treatment, it is necessary to provide a precise spatial description for the target volume and organs at risk (OAR), so as to direct a highly conformal radiation dose to tumor cells while preserving healthy tissues. Therefore, accurate depiction and segmentation of the target volume and the OAR from medical images is a vital step in the effective planning of radiotherapy in patients with HNC. For example, in nasopharyngeal carcinoma, the eye, optic nerve, and optic chiasm must be accurately delineated (70). Since manual rendering is a tedious and time-consuming task which is complicated by internal and/or inter-observer variability, computer automatic segmentation has been developed as an alternative. In the past decade, there has been a growing interest in medical imaging and radiotherapy planning. The emerging trend is to transform the automatic segmentation of HaN OAR from a map-based approach to a DL-based approach (71).

Zhu et al. (72) proposed an end-to-end, atlas-free, 3D convolution DL framework for fast and automatic full volume HaN anatomical segmentation. It provides a feasible solution to describe OAR from computed tomography (CT) images. The model could improve the segmentation accuracy and simplify the automatic segmentation of pipelines. With this method, the OAR could be depicted in a fraction of a second in the HaN CT scans.

Ibragimov et al. (73) proposed an algorithm based on DL to segment OAR in HaN CT images. A CNN was used to train the consistent intensity pattern of OAR in CT images, and the OAR was segmented in a previously invisible test CT image. The performance of the CNN in the segmentation of the spinal cord, mandible, parotid gland, submandibular gland, larynx, pharynx, eyeball, optic nerve, and optic chiasm was verified in 50 CT images. The segmentation results ranged from 37.4% of optic chiasm dice similarity coefficient to 89.5% dice similarity coefficient of the mandible. At the same time, it was suggested that the combination of MR images may be beneficial to some OAR with unclear boundaries.

For the comparison of CT and MR images, Tong et al. (74) developed a novel method based on GAN with shape constraint (SC-GAN) for automatic HaN OAR segmentation on CT and low field MR imaging. The full convolution DenseNet with deep supervision was used as the segmentation network of voxel prediction. Compared with CT, MR imaging has higher accuracy in automatic segmentation of optic nerve. The low field MR images obtained with the MR-guided radiotherapy system could support the accurate and automatic segmentation of OAR soft tissue for adaptive radiotherapy.

Since the outline of the OAR directly affects the planning optimization and local dose distribution, it also influences the evaluation and efficacy of the radiotherapy regimen. Guo et al. (75) evaluated the dosimetric effect of automatic OAR segmentation based on DL on nasopharyngeal carcinoma treatment. There was no correlation between geometric measurement of automatic segmentation and dosimetric differences. Furthermore, automatic segmentation based on DL showed good therapeutic effect on tumor tissues.

**Discussion**

Previous studies have suggested that classifier models trained using both structural and functional tests may have better discrimination ability compared to machine learning classifiers trained using only structural or functional tests. Similarly, DL models that have had combinatorial training may also show superior performance. However, most notably, many eye diseases, such as glaucoma, lack a perfect reference standard or gold standard. Therefore, it is difficult to establish an appropriate unbiased study to evaluate the accuracy of a proposed novel diagnostic method that relies on a combination of structural and functional tests. In such cases, it is important to consider the clinical needs and settings of the innovation application. To develop and extend a DL model that can accurately replicate the diagnosis of diseases by an ophthalmologist, a reasonable combination of the fundus phase, a reference standard graded visual field, and SD-OCT images and printouts can be used, possibly in combination with other clinical texts and imaging information. Such a model can have a huge impact in clinical practice and increase the performance of the clinical diagnosis of diseases.

In a recent study, Mariottoni et al. (76) proposed a set of relatively uncomplicated structural and functional parameters which can be integrated in an objective approach as a reliable reference standard for the research and development of glaucoma AI diagnosis models. According to the RNFL evaluation of SD-OCT and the visual field evaluation of standard automatic visual field examination, the diagnostic criteria of glaucoma optic...
neuropathy should include the corresponding structural and functional damage. They summed up a set of criteria using well-developed global and local parameters, requiring topographic correspondence between the neural structural and functional impairments, which improved the specificity of the diagnosis. They also developed a DL model which uses fundus photos to distinguish glaucoma from normal eyes and classifies them according to objective reference standards. The total AUC of the model was 0.92, with a sensitivity of 77% and a 95% specificity. It is worth noting that an objective reference standard combining SD-OCT and SAP data may avoid laborious, prolonged expert scoring, and it may also increase the comparability of integrating various equipment and diagnostic research on different types of populations.

With unparalleled progress in computer technology and imaging technology, medical imaging has developed from an auxiliary and supporting discipline to an indispensable means of examination for clinical and research applications in modern medicine. DL is an exciting technology. High precision models show that DL can use relatively small data repositories to effectively learn increasingly complex information and high degree of generalization images. To some extent, AI can completely change the diagnosis and management of diseases by classifying difficult images for clinical experts and viewing a large number of images quickly. Compared with artificial evaluation, AI has significant advantages in data mining, information integration, data processing, and interactive query.

DL has great prospects in the diagnosis and treatment of optic nerve diseases. DL models have been shown to use simple images to detect and quantify optic nerve damage, thereby facilitating low-cost, rapid, and immediate screening of diseases. Furthermore, DL has been shown to optimize the damage estimate of the original SD-OCT and visual field data, which can facilitate the application of these tests in clinical diagnosis and treatment. Although AI and DL have made great progress in the diagnosis and treatment of optic nerve diseases, most medical applications using this technology are still in its infancy. Through the interdisciplinary cooperation of clinicians, engineers and designers, AI in health care may eventually accelerate the diagnosis and referral of ophthalmic diseases. The validation of novel diagnostic tests must meet strict performance confirmation procedures and tight quality control criteria. Furthermore, constant monitoring and evaluation must be performed, with particular focus on reference standards and the setting of test strategies in clinical practice. In neuro-ophthalmic diseases, the reference standards used in the clinical setting may vary greatly, depending on the application mode and purpose of the test. Similarly, the accuracy requirements for diagnostic tests may also vary, depending on whether they are used for community group screening or opportunistic screening, or for disease detection or surveillance in tertiary care centers.

At present, the application of DL in clinical practice still faces many challenges. Combining important clinical information relating to each patient in conjunction with information obtained from multimodal imaging will be more beneficial for clinical diagnosis and treatment planning. Future research should focus on the development of highly accurate AI systems capable of not only detecting optic nerve abnormalities but also predicting the nature of potential conditions. Large clinical real life data sets should be used to evaluate the efficacy of AI systems in terms of clinical deployment and cost-effectiveness, and to address whether DL can become an accurate and affordable solution for patients in the future.

Acknowledgments

Funding: This work was supported by grants from the Natural Science Foundation of Liaoning Province (2020-MS-01 to C Wan) and the Natural Science Foundation of Xinjiang Province (2020D01A122 to C Wan). This work was supported by the Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education.

Footnote

Conflicts of Interest: Both authors have completed the International Committee of Medical Journal Editors’ (ICMJE) uniform disclosure form (available at https://dx.doi.org/10.21037/qims-21-728). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with
the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References


2018;256:289-98.

Cite this article as: Li M, Wan C. The use of deep learning technology for the detection of optic neuropathy. Quant Imaging Med Surg 2022;12(3):2129-2143. doi: 10.21037/qims-21-728