Review

Current status and future trends of clinical diagnoses via image-based deep learning

Jie Xu¹, Kanmin Xue ², Kang Zhang³,4✉

¹ Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Science Key Lab, Beijing, China
² Royal Victorian Eye and Ear Hospital, Melbourne, Victoria 3002, Australia
³ Faculty of Medicine, Macau University of Science and Technology, Taipa, Macau
⁴ Institute for Genomic Medicine, UC San Diego, La Jolla, CA, USA

✉ Corresponding author: Kang Zhang; e-mail: kang.zhang@gmail.com

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). See http://ivyspring.com/terms for full terms and conditions.

Received: 2019.07.01; Accepted: 2019.07.28; Published: 2019.10.12

Abstract

With the recent developments in deep learning technologies, artificial intelligence (AI) has gradually been transformed from cutting-edge technology into practical applications. AI plays an important role in disease diagnosis and treatment, health management, drug research and development, and precision medicine. Interdisciplinary collaborations will be crucial to develop new AI algorithms for medical applications. In this paper, we review the basic workflow for building an AI model, identify publicly available databases of ocular fundus images, and summarize over 60 papers contributing to the field of AI development.

Key words: artificial intelligence, deep learning, machine learning, ophthalmology

Introduction

Artificial intelligence (AI) has recently experienced an era of explosive growth across many industries, and healthcare is no exception [1]. AI will have particular utility in healthcare and will dramatically change the diagnostic and treatment pathways for many, if not most, diseases. Regardless of the specific technique, the general aim of these technologies in medicine is to use computer algorithms to uncover relevant information from data and to assist clinical decision making [2]. In many developed countries and China, the application of AI technology in healthcare has developed quickly, at least in part because it enhances human resources and abilities and improves the accuracy of medical treatment. As many countries that support the development of advanced technologies welcome the incoming era of AI, they will begin to develop the necessary specifications of governance by law, regulation, technology, and standards to fully optimize this developing field of technology.

Ophthalmology is a discipline that is highly dependent on technological development. Modern ophthalmology currently makes full use of mechanical, electrical, magnetic, acoustic, optical, and other imaging technologies, and it will lead in fully implementing and adapting new technological developments such as AI. Ophthalmologists should enthusiastically embrace the development of AI technology and use it to promote advances in ocular medicine as much as possible.

Workflow of deep learning

AI is broadly used in both the technical and popular lexicons to encompass a spectrum of learning, including but not limited to machine learning, representation learning, deep learning, and natural language processing [1]. Deep learning is making major advances in solving problems that have resisted the best attempts of the AI community for many years. It is very good at discovering intricate structures in high-dimensional data and is therefore applicable to multiple medical domains [3].
learning discovers intricate structure in large data sets by using multiple intermediate layers positioned between the input and output layers, allowing each level to learn to transform its input signal into the following layer (Fig 1). The application of deep learning, particularly in images of the retina include classification, e.g., detection of diabetic retinopathy (DR) and diabetic macular edema (DME) in fundus photographs [4]; segmentation, e.g., segmentation of the lungs [5], brain [6], cell mitosis [7]; and prediction, e.g., prediction of myopia development and progression [8]. The workflow of deep learning can be defined in three stages: (1) pre-processing of the image data; (2) training of the model, validation, and model testing; and (3) evaluation (Fig. 2). Data pre-processing is a critical step that is necessary to build accurate machine learning models. The pre-processed work includes noise reduction, data normalization, feature selection, and extraction [9].

For training a model, we initially split the model into three sections: data training, validation, and testing. The training set enables the model to learn to fit the data parameters of the classifier. The validation set is used to prevent overfitting, and the test set is used to evaluate the performance of the trained model. Evaluation is an integral part of the development process. It helps to determine if the model will do a good job of predicting the target on new and future data.

Common open database of retina images

Many public databases have been published, and most include instructions for use by researchers in analysis and testing. For diseases of the retina, the databases include basic pathological features that usually provide information about the seven layers of the retina and about the choroid and sclera. This type of information is compiled by a process referred to as “segmentation”, which historically has been achieved by manual image processing, but increasingly it is done by computer algorithms. This information allows comparison of the performance of different algorithms analyzing the same fundus image, with reference to the reliable implementation of a gold-standard procedure [10]. Commonly used fundus databases includes DRIVE [11], STARE [12], Image-Ret [13,14], e-ophtha [15], HEI-MED [16], Retinopathy Online Challenge [17], Messidor [18], RIM-ONE [19], and DRION-DB [20]. Among them, DRIVE, STARE, Image-Ret, and Messidor are used mostly to diagnose DR, while DRION-DB and RIM-ONE are used mostly for segmentation of the optic nerve head in the diagnosis of glaucoma (Table 1).
the disease in the early stage to prevent blindness. The continuous monitoring of DR is essential to control complication of diabetes mellitus. Early detection and recognition. Recent advances of deep learning in the field of computer vision and pattern conditions images.

by training deep models with millions of annotated image classification. Many deep learning models have automatic detection of DR has attracted a lot of attention. Most automated methods use fundus images as an input. These photographs are examined for the presence of lesions indicative of DR, including microaneurysms, hemorrhages, exudates, and cotton wool spots. The application of automated image analysis to fundus images may reduce the workload and costs by minimizing the number of photographs that need to be manually graded [25]. Gulshan et al [4] were the first to present a deep learning algorithm for the detection of DR in retinal fundus photographs. In 2 validation sets of 9963 images and 1748 images, at the operating point selected for high specificity, the algorithm had 90.3% and 87.0% sensitivity and 98.1% and 98.5% specificity for detecting referable diabetic retinopathy, defined as moderate or worse diabetic retinopathy or referable macular edema by the majority decision of a panel of at least 7 US board-certified ophthalmologists. Subsequently, Ting et al [26] developed a deep learning system to detect multiple related eye diseases, including DR, possible glaucoma, and age-related macular degeneration(AMD), the Area under the receiver operating characteristic curve (AUC) of 0.936 for referable DR, sensitivity and specificity were 90.5% and 91.6%, for vision-threatening diabetic retinopathy, AUC was 0.958, sensitivity and specificity were 100% and 91.1%. More recently, deep learning was applied to automated segmentation of optical coherence tomography (OCT) images. Kermann et al [27] developed an OCT imaging diagnostic tool based on a deep learning framework for screening patients with AMD, DME, and drusen. The classifier distinguishing DME images from normal images achieved an accuracy of 98.2%, with a sensitivity of 96.8% and specificity of 99.6%. In April 2018, the first AI diagnostic system to receive US Food and Drug Administration (FDA) approval for marketing was IDx-DR, the case of IDx-DR highlights one of the earliest successes of an AI-based technology completing the regulatory process in the United States.

Table 1. Summary of publicly available databases of ocular retinal images

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of images</th>
<th>Camera Model</th>
<th>Image Resolution (pixels)</th>
<th>Field of View</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRIVE [13]</td>
<td>40</td>
<td>Canon CR5</td>
<td>768×584</td>
<td>45°</td>
<td>Blood vessel segmentation</td>
</tr>
<tr>
<td>STARE [12]</td>
<td>400</td>
<td>Topcon trv-50</td>
<td>605×700</td>
<td>35°</td>
<td>Blood vessel segmentation; Optic disk detection</td>
</tr>
<tr>
<td>Image Ret [13,14]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIARETDB0</td>
<td>130</td>
<td>unknown</td>
<td>1500×1152</td>
<td>50°</td>
<td>Diabetic retinopathy detection</td>
</tr>
<tr>
<td>DIARETDB1</td>
<td>89</td>
<td>unknown</td>
<td>1500×1152</td>
<td>50°</td>
<td>Diabetic retinopathy detection</td>
</tr>
<tr>
<td>e-optha [15]</td>
<td>82</td>
<td>OPHDIAT Tele-medical network</td>
<td>2048×1360; 1440×960</td>
<td>-</td>
<td>Diabetic retinopathy detection</td>
</tr>
<tr>
<td>e-optha_MA</td>
<td>381</td>
<td>Zeiss Visucam PRO fundus camera</td>
<td>2196×1958</td>
<td>45°</td>
<td>Hard exudate detection; Diabetic macular edema assessment</td>
</tr>
<tr>
<td>HEI-MED [16]</td>
<td>169</td>
<td>unknown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retinopathy Online Challenge [17]</td>
<td>100</td>
<td>Canon CR5-45-NM</td>
<td>768×576; 1058×961</td>
<td>45°</td>
<td>Microaneurysms detection</td>
</tr>
<tr>
<td>Messidor [18]</td>
<td>1200</td>
<td>TopCon TRC NW6</td>
<td>1389×1383</td>
<td>45°</td>
<td>Diabetic retinopathy detection</td>
</tr>
<tr>
<td>RIM-ONE [19]</td>
<td>169</td>
<td>Nidek AFC-210 with a body of a Canon EOS 5D</td>
<td>2144 × 1424</td>
<td>-</td>
<td>Optic nerve head segmentation</td>
</tr>
<tr>
<td>DRIONS-DB [20]</td>
<td>110</td>
<td>Color analogue fundus camera digitized by HP-Photosmart-S20 scanner</td>
<td>600×400</td>
<td>-</td>
<td>Optic nerve head segmentation</td>
</tr>
</tbody>
</table>

Important research studies applying artificial intelligence to ophthalmic conditions

Image classification is a long-term research topic in the field of computer vision and pattern recognition. Recent advances of deep learning techniques have greatly facilitated the research of image classification. Many deep learning models have demonstrated performances comparable with well-trained human experts in the classification of natural images, and some have outperformed the experts. The flourish of deep learning-based image classification started with the work of AlexNet [21], where an eight-layered convolutional neural network (CNN) was designed for the classification task in ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Subsequently, a series of deeper neural networks continuously refreshed the record of ILSVRC, including GoogLeNet (22 layers) [22], VGGNet (16 or 19 layers) [23], and Deep Residual Net (18, 34, 50, 101, or 152 layers) [24]. These neural networks are the most widely used architectures that can achieve accurate classification for natural images by training deep models with millions of annotated images.

Diabetic retinopathy

Diabetic retinopathy is the most common organ complication and can manifest as the earliest sign of complication of diabetes mellitus. Early detection and continuous monitoring of DR is essential to control the disease in the early stage to prevent blindness. The automatic detection of DR has attracted a lot of attention. Most automated methods use fundus images as an input. These photographs are examined for the presence of lesions indicative of DR, including...
Glaucoma

Glaucoma is a group of eye diseases that damage the optic nerve and can result in irreversible vision loss and blindness and is the second leading cause of blindness worldwide. It is estimated that the disease affected 60.5 million people in 2010, and this figure is expected to reach 79.6 million by 2020 [28]. Currently, there is no cure for glaucoma, and vision loss, once it has occurred, is permanent. However, early detection and treatment are helpful to slow or stop the disease progression and can protect against serious vision loss. Many researchers have studied how to diagnose glaucoma automatically based on retinal images. These studies can be separated into two types. The first type outputs the glaucoma diagnosis results directly through deep learning models. Li et al [29] trained a CNN on LabelMe datasets for glaucoma diagnosis. In the validation dataset, this DL system achieved an AUC of 0.986 with sensitivity of 95.6% and specificity of 92.0%. The most common reasons for false-negative grading were glaucomatous optic neuropathy with coexisting eye conditions, including pathologic or high myopia, DR, and AMD. The leading reason for false-positive results was having other eye conditions, mainly including physiologic cupping. The second type of studies uses deep learning models to segment the glaucoma related tissues such as optic disc and optic cup, and then calculates medical measures (e.g., cup-to-disc ratio) for diagnosis. Previous studies have used various special forms of perimetry to discriminate preperimetric glaucoma from healthy eyes [30]. Asaoka et al [31] applied a DL method to differentiate the visual fields of pre–plus disease or worse, the sensitivity and specificity were 100% and 94%, respectively. In addition to detection and monitoring common blinding eye diseases, deep learning is also being expanded to the field of rare diseases, such as congenital cataracts and retinopathy of prematurity (ROP) in newborns. Long et al [34] constructed a CNN-based computer-aided diagnosis framework (CC-Cruiser) to classify and grade congenital cataract. In the clinical trial, CC-Cruiser achieved 98.25% accuracy with the identification networks; 100%, 92.86% and 100% accuracy for opacity areas, densities and locations, respectively, with the evaluation networks; and 92.86% accuracy with the strategist networks. Brown et al [35] reported the results of a fully automated DL system that could diagnose plus disease, the most important feature of severe ROP, for diagnosis of plus disease, the algorithm achieved a sensitivity of 93% with 94% specificity. For detection of pre–plus disease or worse, the sensitivity and specificity were 100% and 94%, respectively. In addition, retinal microvascular changes and retinopathy provide important clinical indicators for predicting the occurrence, development, therapeutic effect and prognosis of cardiovascular and cerebrovascular diseases. Poplin et al [36] using deep-learning models trained on data from 284,355 patients and validated on two independent datasets of 12,026 and 999 patients, they predicted cardiovascular risk factors not previously thought to be present or quantifiable in retinal images, such as age (mean absolute error within 3.26 years), smoking status (AUC = 0.97), systolic blood pressure (mean absolute error within 11.23 mmHg) and major adverse cardiac events (AUC = 0.70). Current AI studies using deep learning techniques for DR, AMD, glaucoma, cataract, and anterior ocular segment diseases are summarized in Table 2 [4,26,27,29,31,34,37-63].
## Table 2. Summary of influential papers on ophthalmic image analysis

<table>
<thead>
<tr>
<th>Disease</th>
<th>Procedures/examinations</th>
<th>Data sets</th>
<th>Deep learning techniques</th>
<th>Performance</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keratoconus</td>
<td>Pentacam</td>
<td>194 normal, 454 keratoconus, 67 forme fruste, 28 astigmatic, 117 after refractive surgery</td>
<td>SVM</td>
<td>Acc: 98.9%; Sen: 99.1%; Spe: 98.5%; AUC: 0.99</td>
<td>Hidalgo et al. [37]</td>
</tr>
<tr>
<td></td>
<td>Dual Scheimpflug</td>
<td>177 normal, 148 keratoconus</td>
<td>Decision Tree</td>
<td>Sen: 100%; Spe: 99.5%</td>
<td>Smadja et al. [38]</td>
</tr>
<tr>
<td></td>
<td>Pentacam HR</td>
<td>30 normal, 15 unilateral keratoconus, 30 bilateral keratoconus</td>
<td>FNN</td>
<td>Bilateral keratoconus versus normal; AUC: 0.99; Sen: 100%; Spe: 95%</td>
<td>Kovacs et al. [39]</td>
</tr>
<tr>
<td>Pterygium</td>
<td>Anterior segment</td>
<td>2,692 non-pterigym, 325 pterygium</td>
<td>Shape features + SVM/ANN</td>
<td>Acc: 91.27%; AUC: 0.956; Sen: 88.7%; Spe: 88.3%</td>
<td>Zaki et al. [40]</td>
</tr>
<tr>
<td></td>
<td>photographed images</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cataract</td>
<td>Slit-lamp image</td>
<td>5,379 images with decimal grading scores ranging from 0.1 to 5.0.</td>
<td>CRNN</td>
<td>70.7% exact integral agreement ratio (RI); 88.4% decimal grading error 5 1.0 (Re5.0).</td>
<td>Gao et al. [41]</td>
</tr>
<tr>
<td></td>
<td>Slit-lamp image</td>
<td>476 normal, 410 cataract</td>
<td>DCNN</td>
<td>Cataract vs Normal: Acc:98.87%; Sen:98.78%; Spe:98.95%</td>
<td>Long et al. [34]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>767 normal, 472 cataract (246 mild cataract,128 moderate cataract, and 98 severe cataract)</td>
<td>SVM, BPNN</td>
<td>Acc: 93.2% for detection,84.5% for grading; Sen:94.2% for detection,74.6-89.3% for grading; Spe:91.5% for detection,90.4-98.9% for grading</td>
<td>Yang et al. [42]</td>
</tr>
<tr>
<td></td>
<td>Slit-lamp image</td>
<td>476 normal, 410 pediatric cataract</td>
<td>CNN, SVM</td>
<td>Acc, Sen, and Spe: classification (97.07%, 97.28%, and 96.83%); three-degree grading area (89.02%, 86.63%, and 90.75%); density (92.68%, 91.05%, and 93.94%); location (89.28%, 82.70%, and 93.08%)</td>
<td>Liu et al. [43]</td>
</tr>
<tr>
<td>POAG</td>
<td>Fundus image</td>
<td>Training set:125,189; Validation set:71,896</td>
<td>DLS</td>
<td>AUC: 0.942; Sen: 96.4%; Spe: 87.2%</td>
<td>Ting et al. [26]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>Training set: 31,745; Validation set: 8,000</td>
<td>DCNN</td>
<td>AUC: 0.98; Acc: 92.9%; Sen: 95.6%; Spe: 92.0%; AUC: 0.986</td>
<td>Li et al. [29]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>589 normal, 837 glaucoma</td>
<td>CNN</td>
<td>Acc: 98.13%; Sen: 98%; Spe: 98.3%</td>
<td>Raghavendra et al. [44]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>30 normal, 30 open-angle glaucoma</td>
<td>SVW</td>
<td>Acc:91.67; Sen:90%;Spe:93.33%</td>
<td>Krishnan et al. [45]</td>
</tr>
<tr>
<td></td>
<td>Visual field</td>
<td>Training set:225; Test set: 129</td>
<td>ANN</td>
<td>AUC: 0.890; Sen: 78.3%; Spe: 89.5%</td>
<td>Oh et al. [46]</td>
</tr>
<tr>
<td></td>
<td>Visual field</td>
<td>266 normal, 72 mild, 86 moderate, 86 severe glaucoma</td>
<td>SVM</td>
<td>Acc: 93.1%; Sen: 89.75%; Spe: 96.2%</td>
<td>Acharya et al. [47]</td>
</tr>
<tr>
<td></td>
<td>Visual field</td>
<td>Normal/glaucoma Fundus images/85/39; Normal/glaucoma SLO images: 46/19</td>
<td>RIFM</td>
<td>Acc for Fundus images: 94.4%,SLO images: 93.9%,Sen for Fundus images: 92.3%,SLO images: 89.5%,Spe for Fundus images: 93.5%,SLO images: 93.5%</td>
<td>Haleem et al. [48]</td>
</tr>
<tr>
<td></td>
<td>Visual fields</td>
<td>53 glaucoma eyes, 108 normal eyes</td>
<td>FNN</td>
<td>AUC: 92.6%, The sensitivity was 77.8%,54.6%, and 50.0%, respectively, at the specificity of 90%, 95%,and 99%</td>
<td>Asaoka et al. [31]</td>
</tr>
<tr>
<td>DR</td>
<td>Fundus image</td>
<td>Training set:76,370; Validation set:112,648</td>
<td>DLS</td>
<td>For referable DR: AUC: 0.936; Sen: 90.5; Spe: 91.6%; For vision-threatening DR: AUC: 0.958; Sen: 100%; Spe: 91.1%</td>
<td>Ting et al. [26]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>Development Data Set (EyePACS in the United States and 3 eye hospitals in India): 120,175 Validation Data Set (EyePACS-1: 9,963; Messidor-2: 1,748)</td>
<td>DCNN</td>
<td>AUC: 0.99 for EyePACS;0.990 for Messidor;Sen: 90.3% for EyePACS;87% for Messidor;Spe: 98.1% for EyePACS;98.5% for Messidor</td>
<td>Guishan et al. [4]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>170 DR, 170 normal</td>
<td>PNN-GA, SVM quadratic kernels</td>
<td>PNN-GA: Acc:99.41%,Sen:99.41% S Spe:98.41%; SVM; Acc:99.12 % Sen:98.82%,Spe:99.41%</td>
<td>Ganesan et al. [49]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>EyePACS; 75,137 DR images; External validation: MESSIDOR 2 and E-Ophtha</td>
<td>DCNN</td>
<td>AUC 0.94 for Messidor 2, 0.95 for E-Ophtha,Sen 93 for Messidor 2.87% for E-Ophtha,Spe 90% for Messidor 2.94% for E-Ophtha</td>
<td>Gargeya et al. [50]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>Training set: 327 diabetic patients; Validation set: 725 diabetic patients</td>
<td>LASSO</td>
<td>Acc: 89.2%; AUC: 0.90; Sen: 75%; Spe: 89.6%</td>
<td>Oh et al. [51]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>Training set: 400; Testing set: 9,954</td>
<td>Ensemble of classifiers with hidden Markov chain for context information, trained by genetic algorithm</td>
<td>Sen: 92.2%; Spe: 90.4%</td>
<td>Tang et al. [52]</td>
</tr>
<tr>
<td></td>
<td>Fundus image</td>
<td>Messidor-2 dataset: 1,748</td>
<td>CNN</td>
<td>Referable DR: AUC: 0.980; Sen: 96.8%; Spe: 87%; Vision threatening DR: AUC: 0.989; Sen: 100%; Spe: 90.8%</td>
<td>Abramoff et al. [53]</td>
</tr>
<tr>
<td>Fundus image</td>
<td>4,445 DR, 5,494 normal</td>
<td>DIARETDB1, FAZ, MESSIDOR Private dataset: 750 (Normal: 150, mild NPDR: 150, moderate NPDR: 150, severe NPDR: 150, PDR: 150)</td>
<td>DCNN</td>
<td>Acc: 0.81</td>
<td>Takahashi et al. [54]</td>
</tr>
<tr>
<td>Fundus image</td>
<td>4,445 DR</td>
<td></td>
<td>DNN</td>
<td>AUC: 0.924; Sen: 92.18%; Spe: 94.50%</td>
<td>Abbas et al. [55]</td>
</tr>
</tbody>
</table>
Status of AI applications in clinical diagnoses

On April 2, 2019, the FDA issued a discussion paper that proposed a regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as medical devices (SaMD) [64]. This document explains the principles for developing artificial intelligence software, the new framework for equipment, the principles of the total product lifecycle (TPLC) certification methodology, and examples of potential real-world AI software modifications that may or may not be allowed under the proposed framework. The idea of the proposal was that with appropriately tailored regulatory oversight, AI/ML-based SaMD will deliver safe and effective software functionality that improves the quality of care that patients receive.

To date, two AI algorithms have been fully approved by the FDA and used clinically. One is IDx-DR for detecting DR, and the other is Viz.AI for analyzing images for indicators associated with a stroke. These two devices are described as “locked” algorithms, meaning that they can only be modified by the manufacturer at intervals for the purpose of “training” with new data, followed by manual verification and validation of the updated algorithm. However, there is much promise beyond locked algorithms that is ripe for health care applications. These opportunities require careful oversight to ensure that the benefits of these advanced technologies outweigh the risks to patients. These machine learning algorithms can continually evolve and are often referred to as “adaptive” or “continuously learning” algorithms. Adaptive algorithms can learn from new user data presented to the algorithm through real-world use. The FDA is exploring a framework that would allow for modifications to algorithms to be made from real-world learning and adaptation, while ensuring that the safety and effectiveness of the software as a medical device is maintained [65].

Main challenges in the application of AI

At present, there are at least four limitations in AI technology that is based on machine learning. First, most machine learning methods have too few training sets and verification sets. More image data training is needed to further improve accuracy, sensitivity, and specificity. Transfer learning is an approach more suitable when limited training data is available for the problem under consideration. In transfer learning, one can learn a complex model using data from a source domain where large-scale annotated images are available (e.g. natural images). Then, the model is further fine-tuned with data of the target domain where only a small number of annotated images are available (e.g. medical images) [66] (Fig 3).

The second limitation is that the examination/detection equipment used in different...
countries, regions, and medical institutions is not uniform; therefore, the acquired images have differences in quality and resolution, which will inevitably affect the accuracy of image analysis and thus affect the accuracy of the diagnosis. These differences will present certain obstacles in the wide-scale applications of AI technology. One solution to this problem is to unify and standardize the examination equipment. This will be difficult to achieve. Another method is to further improve AI machine learning methods at the framework and algorithm level so that they can be flexibly applied to images of different qualities while simultaneously ensuring the accuracy of intelligent diagnosis. This will increase the applicability of AI in different regions and medical institutions. However, a lot of research support is still needed in this area.

The third limitation is that the current machine learning methods for disease diagnosis lack “explanation capacity”. They do not have the ability to provide the clinician or other users with the reason for the diagnosis. The output result is based only on training and intensive learning. Thus, it is only a simple statement based upon the differences in the patient and normal images. There is no explanation for why the differences exist or the pathological basis of the differences. This, along with other issues, will affect to some extent the acceptance of these devices by doctors in clinical applications and could even provoke confusion among clinicians.

Finally, the fourth limitation in machine learning-based AI technology is that the diagnosis of some rare diseases is still unreliable. Because of the scarcity of these diseases, there are not enough cases to meet the requirements of the training and verification sets. It is difficult or impossible to ensure the accuracy of the learning model in diagnosing rare diseases. To improve the diagnosis of rare diseases, it will be necessary to optimize machine learning algorithms. This can be done by transitioning from reliance on the number of learning samples for accuracy training to utilizing combinations of various training modes and types.

Suggestions for the application and development of AI in medicine

1. Consolidate the data foundation of AI applications

AI must first collect a certain amount of data in the medical industry. The structure of current medical data is very complex, in part, because there is no uniformity in the standards for compilation, and this leads to widespread information islands. It is necessary to establish a mechanism for circulating and sharing medical data. Development of data desensitization methods will promote data standardization and normalization. These methods can be used to establish standard test data sets to consolidate the data foundations of AI applications.
2. Define the positioning of AI development

In medicine, AI aims to help doctors (rather than substitute for doctors) to reduce the morbidity and mortality rates of patients waiting for professionals. Because doctors will not be replaced by AI, the diagnostic result of AI is only a reference for a clinical diagnosis, and doctors will always be responsible for the result. Currently, AI products in medicine play only an auxiliary role in the clinic, such as the diagnosis of DR, cancer screening, medical image recognition, disease rehabilitation, and in other fields.

3. Formulate regulations and laws for the application of AI in medicine

To achieve the uniform standards necessary for effective medical AI applications, laws, regulations, and other levels of governance must be established at the national level. The implementation of the national standards in industry and in routine and research clinical settings will ensure that the technology can be made widely and quickly available in the safest and most rational way. This will prudently promote the application of AI in the medical field.

4. Strengthen data security of AI data applications

As with the collection of any personal and medical data, the risk of inadvertent or pirated disclosure is a major concern. To reduce these risks, it is necessary to strengthen the construction of privacy protection, desensitize the data, and collect the data according to different levels and different granularities to reduce the risk of privacy disclosure. The United States has extensive experience in privacy protection of medical data. While enhancing privacy protection, it encourages the rational access and meaningful use of data and makes a selected portion of it available for research in real time, open access databases.

5. Promote the cultivation of interdisciplinary talents

Future development of AI in medicine can be enhanced by focusing on the integration of disciplines such as medicine, information science, and engineering. This goal can be achieved by encouraging universities, research institutes, and enterprises to cooperate with each other, set up appropriate scholarships, and establish training bases and local pilot programs. Finally, an innovative talent introduction system and mechanism should be developed to attract highly talented students to carry out innovation and entrepreneurship in the field of medical AI, thus driving the further development of the field.

Future of AI application in clinic

Machine learning has shown its great potential in ophthalmology. Most of the current studies regarding intelligent diagnosis of eye diseases focus on dual classification problems, whereas many patients suffer from multiple categorical retinal diseases in the clinical setting. It is therefore necessary to have a model for detecting and distinguishing DR, AMD, glaucoma, and other retinal disorders simultaneously [67].

With a new generation of AI developed as a broad strategy, the applications of AI in the medical field will increase and improve. AI plays an important role in disease diagnosis and treatment, health management, drug research and development, precision medicine, etc. It can contribute significantly to solving problems of the uneven distribution of medical resources, reducing costs, and improving treatment efficiency. Applying AI helps to make up for the shortcomings of insufficient medical resources, enhance the fairness of medical services, and improve the construction of hierarchical diagnosis and treatment. In the future, AI will also offer important support for establishing an integrated medical service system. A qualified and efficient integrated medical service system can be built with the help of information-based systems.

Laws and regulations to define the legal status, responsibility sharing mechanisms, and supervision of automated systems are not yet enacted in China and most other countries. Given the complex ethical boundaries of medical AI application, the excessive control of medical AI will hinder innovation and development. On the other hand, the lack of management brings the risk of unclear subject responsibility in AI applications. Therefore, it is necessary to reasonably define AI in the medical field. The current laws on medical AI regulation are either non-existent or are in only the most primitive stages of development. There are no sound laws that regulate medical big data, the basis for medical AI. Further, there are no clear legal instructions regarding the ownership of AI data, the right to use it, privacy standards, data security, accountability norms, and whether laws can protect researchers, clinicians, and engineers from errors of innovation.

Summary

Deep learning has the ability to probe more deeply into and discern more discriminative features from extremely large datasets. It has been applied in many research and clinical fields that rely on medical image analysis, making breakthrough progress in those disciplines. Due to unique features in
ophthalmology, the diagnosis of eye diseases in clinical practice requires interpretation of many imaging studies for auxiliary diagnosis. However, detection resolution of the human eye is limited and so is human attention span. Proficiency levels of ophthalmologists also differ and it’s inevitable for human errors to occur. As documented in the existing literature review, most of the current deep learning methods representing the leading level are the use of supervised learning, especially the CNN-based framework. Preliminary researches mainly focused on pre-training CNN and taking CNN as feature extractor. These pre-training networks could be downloaded directly, and conveniently applied to the analysis of any medical images. In recent two years, end-to-end training CNN has become a prioritized application of deep learning to the analysis of medical images. In recent two years, end-to-end training CNN has become a prioritized approach for the analysis of medical images. However, obtaining well-annotated data used for supervised learning is another major challenge for the application of deep learning to the analysis of medical images. As annotated data were usually limited at present, how to utilize unannotated images to achieve a high diagnostic accuracy using a combination of unsupervised and supervised learning will be another important development direction. In addition, electronic medical records (EMRs) contain a wealth of clinical diagnostic and treatment information that can be extracted and used to form diagnoses using natural language processing and deep learning. This information can be used to supplement the image data to formulate a complete diagnosis mimicking a human physician [68]. Thus, in the near foreseeable future, AI relying on deep learning will combine image analysis with EMRs, further advancing the diagnostic power and ability to monitor disease progression and response to treatment in ways never before anticipated.

Acknowledgements

This study was funded in part by: the Beijing Hospitals Authority Youth Programme, code: QML20170206; The priming scientific research foundation for the junior researcher in Beijing Tongren Hospital, Capital Medical University (No. 2018-YIJ-ZZL-052).

Competing Interests

The authors have declared that no competing interest exists.

References

46. Oh E, Yoo TK, Hong S. Artificial neural network approach for differentiating open-angle glaucoma from glaucoma suspect without a visual field test. Invest Ophthalmol Vis Sci. 2015;56: 3957-3966.