

# 基于深度学习的眼底疾病筛查诊断系统的初步研究

陈婷丽<sup>1</sup>, 王静<sup>1</sup>, 袁非<sup>2</sup>

引用: 陈婷丽, 王静, 袁非. 基于深度学习的眼底疾病筛查诊断系统的初步研究. 国际眼科杂志 2020;20(8):1452-1455

基金项目: 上海市卫生健康委员会卫生行业临床研究基金 (No. 20194Y0437)

作者单位: <sup>1</sup>(214065) 中国江苏省无锡市, 华东疗养院眼科; <sup>2</sup>(200032) 中国上海市, 复旦大学附属中山医院眼科

作者简介: 陈婷丽, 女, 毕业于南京医科大学, 在读硕士研究生, 主治医师, 研究方向: 眼底病。

通讯作者: 袁非, 男, 博士研究生, 主任医师, 教授, 眼科主任, 博士研究生导师, 研究方向: 有晶状体眼和无晶状体眼的内屈光矫正手术及功能性视力的重建, 糖尿病视网膜病变防治。 yuanfei20@yahoo.com

收稿日期: 2019-12-17 修回日期: 2020-06-30

## 摘要

目的: 评估基于深度学习的眼底疾病筛查人工智能诊断系统的应用价值。

方法: 收集 2018-07/12 在我院就诊的患者 1345 例 2690 眼, 通过分析比较眼科专家诊断及基于多层深度卷积神经网络学习的人工智能诊断系统的一致性, 确定人工智能诊断系统的准确率、特异性和敏感性。

结果: 人工智能诊断系统的准确率为 62.82%, 所纳入患者的诊断结果有 1~5 (1.38±0.67) 个诊断, 其中 1 个诊断的准确率为 56.09%, 2 个诊断的准确率为 77.96%, 3 个诊断的准确率为 84.61%, 4 个诊断的准确率为 86.95%, 5 个诊断的准确率为 60.00%; 无明显异常及豹纹状眼底的一致性 Kappa 值分别为 0.044、0.169, 敏感性分别为 3.00%、99.6%, 特异性分别为 99.7%、14.2%, 其余诊断的一致性 Kappa 值高达 0.57~1.00, 敏感性高达 65.1%~100%, 特异性高达 93.0%~100%。

结论: 基于多层深度卷积神经网络学习的人工智能诊断系统能较好地诊断眼底疾病, 有望成为基层医疗的有效筛查工具。

关键词: 人工智能; 眼底照相; 深度学习; 卷积神经网络

DOI:10.3980/j.issn.1672-5123.2020.8.34

## A preliminary study of a deep learning assisted diagnostic system with an artificial intelligence for detection of retina disease

Ting-Li Chen<sup>1</sup>, Jing Wang<sup>1</sup>, Fei Yuan<sup>2</sup>

Foundation item: Shanghai Municipal Health Commission Clinical Research Fund (No.20194Y0437)

<sup>1</sup>Department of Ophthalmology, Huadong Sanatorium, Wuxi 214065, Jiangsu Province, China; <sup>2</sup>Department of Ophthalmology,

Zhongshan Hospital, Fudan University, Shanghai 200032, China

Correspondence to: Fei Yuan. Department of Ophthalmology, Zhongshan Hospital, Fudan University, Shanghai 200032, China. yuanfei20@yahoo.com

Received:2019-12-17 Accepted:2020-06-30

## Abstract

• AIM: To evaluate the application value of artificial intelligence diagnosis system for fundus disease screening based on deep learning.

• METHODS: A total of 1 345 patients (2 690 eyes) in our hospital were recruited from July 2018 to December 2018. The accuracy, specificity, consistency and sensitivity of the artificial intelligence diagnosis system were determined by comparison with ophthalmologist diagnosis and artificial intelligence diagnosis system which based on multi-layer deep convolution neural network learning.

• RESULTS: The accuracy of artificial intelligence diagnosis system is 62.82%. There are 1-5 (1.38±0.67) diagnoses among the patients, among which the accuracy of one diagnosis is 56.09%, the accuracy of two diagnosis is 77.96%, the accuracy of three diagnosis is 84.61%, the accuracy of four diagnosis is 86.95%, and the accuracy of five diagnosis is 60.00%; The consistency kappa value without obvious abnormality and leopard pattern fundus was 0.044 and 0.169 respectively. The sensitivity was 3.00% and 99.6% respectively, the specificity was 99.7% and 14.2% respectively. The consistency Kappa value of other diagnosis was as high as 0.57-1.00, the sensitivity was as high as 65.1%-100%, and the specificity was as high as 93.0%-100%.

• CONCLUSION: This study shows that the artificial intelligence diagnosis system based on multi-layer deep convolution neural network learning is a reliable alternative to diagnose retina diseases, and it is expected to become an effective screening tool for primary medical treatment.

• KEYWORDS: artificial intelligence; fundus photography; deep learning; convolution neural network

Citation: Chen TL, Wang J, Yuan F. A preliminary study of a deep learning assisted diagnostic system with an artificial intelligence for detection of retina disease. *Guoji Yanke Zazhi (Int Eye Sci)* 2020;20(8):1452-1455

## 0 引言

糖尿病性视网膜病变、视网膜大动脉瘤和黄斑变性等眼底疾病常常严重危害患者视力, 而患者在疾病早期常无明显自觉症状。因此, 此类疾病的筛查和早期诊治对于患

者的预后具有重要的意义。随着分级诊疗的普及,眼底病变的筛查工作一直成为基层眼科、社区医院的重心,而探索一种经济快捷、高效的眼底病变筛查方式,一直以来是广大眼科医师和人工智能工程人员追求的目标。随着人工智能的快速兴起,实现眼底病变的智能化筛查成为目前研究的热点<sup>[1-2]</sup>。人工智能在糖尿病性视网膜病变<sup>[3-4]</sup>、青光眼<sup>[5-6]</sup>、黄斑变性<sup>[7-8]</sup>等眼科疾病的应用具有很高的敏感性和特异性。本研究拟通过人工智能诊断系统对真实世界中的患者进行诊断,以评估其在眼底疾病筛查诊断中的应用价值。

## 1 对象和方法

**1.1 对象** 收集 2018-07/12 在我院眼科体检的疗养员 1345 例 2690 眼,其中男 830 例 1660 眼,女 515 例 1030 眼,年龄 16~77(平均 45.06±10.70)岁,最佳矫正视力为 0.04~1.5。纳入标准:患者眼底照相成像清晰且患者同意本研究对其眼底照相进行相关数据分析。排除标准:(1)因眼外伤、角膜混浊、白内障、玻璃体混浊(积血)等导致屈光间质透明度下降,影响眼底清晰成像者;(2)因全身疾病无法配合眼底照相检查者。本研究经医院伦理委员会批准,经患者同意并签署知情同意书。

**1.2 方法** 所有体检者均由同一检查者采用免散瞳眼底彩色照相机进行眼底拍照,分别拍摄以黄斑为中心和以视盘为中心的眼底照片。采用以多层深度卷积神经网络算法为基础的人工智能诊断系统对所上传的眼底图像进行分析,获取人工智能诊断报告,作为人工智能诊断组。同时,该患者由两位眼底专科医师会诊,彼此分别独立对患眼进行裂隙灯下 90D 前置镜检查,分别得出诊断报告。将完全相同的诊断结果作为最终人工诊断结果,若出现诊断差异性,交由眼底专科主任医师进行最终诊断的确定,以上结果作为专家诊断组。本研究中主要诊断包含有 21 个临床常见眼底病所需诊断:(0)未见明显异常;(1)玻璃膜疣(黄斑区外);(2)眼底动脉硬化;(3)年龄相关性黄斑变性(ARMD)玻璃膜疣;(4)豹纹样眼底;(5)疑似白内障眼底/图片质量不佳;(6)杯盘比偏大;(7)其他黄斑变性;(8)黄斑前膜;(9)其他视神经病变;(10)不明异常-就诊/观察;(11)大玻璃膜疣/色素沉着;(12)视网膜零星出血;(13)视网膜有髓神经纤维;(14)非增殖性糖尿病性视网膜病变轻度;(15)非增殖性糖尿病性视网膜病变中度;(16)非增殖性糖尿病性视网膜病变重度;(17)增殖性糖尿病性视网膜病变;(18)视网膜分支静脉阻塞;(19)视网膜大动脉瘤;(20)视网膜激光斑。以专家诊断组为金标准,计算人工智能诊断组的准确率、一致性、敏感性和特异性。

**统计学分析:**采用统计学软件 SPSS 21.0 进行统计分析。计量资料以均数±标准差表示,计数资料以眼数表示。通过 Kappa 检验比较人工智能诊断组与专家组结果的一致性。其中 Kappa 值>0 且≤0.20 为极低的一致性,>0.20 且≤0.40 为一般的一致性,>0.40 且≤0.60 为中等的一致性,>0.60 且≤0.80 为高度一致性,>0.80 为几乎完全一致。Kappa 检验中  $P<0.05$  表示两者之间的一致性具有统计学意义。以专家组结果为金标准,计算人工智能诊断组的准确性、敏感度和特异性,其中涉及四个统计参数:真阳性(true positive, TP)、假阳性(false positive, FP)、真阴性(true negative, TN)、假阴性(false negative, FN),计算公式如下:准确率=(TP+TN)/(TP+FP+FN+TN),敏感性=TP/(TP+FN),特异性=TN/(FP+TN)。

## 2 结果

**2.1 人工智能系统诊断准确率** 在本研究中,所纳入患者的诊断结果有 1~5(1.38±0.67)个诊断,人工智能诊断组所有诊断完全一致的准确率为 62.82%,其中仅 1 个诊断的准确率为 56.09%,2 个诊断的准确率为 77.96%,3 个诊断的准确率为 84.61%,4 个诊断的准确率为 86.95%,5 个诊断的准确率为 60.00%。在仅 1 个诊断的结果中,由于豹纹状眼底诊断产生差异的为 606 眼(71.63%),去除豹纹状眼底诊断差异后,该组诊断准确率为 87.54%。

**2.2 人工智能系统诊断的一致性** 在本研究中,由于患者的诊断结果大于等于 1 个诊断,因此,我们根据不同诊断对结果分别进行一致性分析。人工智能诊断组和临床专家组一致性分析结果见表 1,在玻璃膜疣(黄斑区外)、ARMD 玻璃膜疣、黄斑前膜、不明异常-就诊/观察、近视萎缩弧、视网膜有髓神经纤维、分支静脉阻塞、非增殖性糖尿病性视网膜病变(轻度、中度、重度)、视网膜大动脉瘤、视网膜激光斑诊断中,Kappa 值>0.80,具有几乎一致性。在疑似白内障眼底/图片质量不佳、动脉硬化、杯盘比增大、其他黄斑变性、大玻璃膜疣/色素沉着、视网膜零星出血、增殖性糖尿病性视网膜病变诊断中,Kappa 值>0.60 且≤0.80,具有高度一致性。在其他视神经病变中,Kappa 值>0.40 且≤0.60,具有中等一致性,在无明显异常和豹纹状眼底中,Kappa 值>0 且≤0.20,具有极低一致性。

**2.3 人工智能系统诊断的敏感性和特异性** 在本研究中,人工智能系统诊断的敏感性结果见表 1,在玻璃膜疣(黄斑区外)、ARMD 玻璃膜疣、豹纹状眼底、黄斑前膜、不明异常-就诊/观察、近视萎缩弧、视网膜零星出血、视网膜有髓神经纤维、分支静脉阻塞、视网膜大动脉瘤、视网膜激光斑、非增殖性糖尿病性视网膜病变(轻度、中度、重度)诊断中,敏感性>0.80;在动脉硬化、疑似白内障眼底/图片质量不佳、杯盘比增大、其他黄斑变性、增殖性糖尿病性视网膜病变、大玻璃膜疣/色素沉着、其他视神经病变诊断中,敏感性>0.60 且≤0.80;在无明显异常诊断中,敏感性>0 且≤0.20,其中在专家组认为无明显异常而人工智能组认为有异常的 501 例中,人工智能组诊断为豹纹状眼底改变的占 98.6%,为 494 例。人工智能系统诊断的特异性结果见表 1,豹纹状眼底特异性为 0.142,其余诊断特异性均大于 0.90,具有较高的特异性。

## 3 讨论

近年来,随着人工智能的快速发展,人工智能在临床中的应用日益广泛<sup>[2,9-10]</sup>。深度学习算法是对人工智能神经网络的发展,其中包括多种算法,诸如:受限波尔兹曼机、卷积网络、堆栈式自动编码器等<sup>[11]</sup>。本研究中,人工智能主要以深度学习为基础,通过多层深度卷积神经网络的训练,自动获取眼底图片中的异常病灶信息,进而得出相应的诊断。

卷积神经网络<sup>[12]</sup>(convolutional neural networks, CNN)是近几年以来发展迅速并且受到各学科领域广泛关注的一种深度学习的模型。早期的卷积神经网络主要用于手写识别,随着其不断发展并且广泛应用在图像处理、物体分类及语音处理等领域。卷积神经与传统的神经网络相比,具有以下优点:(1)卷积神经网络中的神经元采用局部感知连接,相比较传统神经网络全连接的方式很大程度上降低了网络计算的复杂度<sup>[13]</sup>;(2)卷积神经网络权值共享的特殊方式使其布局更加接近生物网络,在图像处理上极

表1 人工智能诊断系统的一致性、敏感性和特异性

临床诊断	Kappa 值	敏感性	特异性
近视萎缩弧	0.898	0.865	0.998
玻璃膜疣(黄斑区外)	0.847	0.906	0.977
ARMD 玻璃膜疣	0.810	0.896	0.973
黄斑前膜	0.863	0.863	0.998
其它视神经病变	0.570	0.800	0.998
不明异常-就诊/观察	0.814	1	0.998
视网膜激光斑	1	1	1
视网膜有髓神经纤维	0.857	0.818	0.999
视网膜分支静脉阻塞	1	1	1
视网膜大动脉瘤	1	1	1
疑似白内障眼底/图片质量不佳	0.723	0.717	0.996
无明显异常	0.044	0.030	0.997
豹纹状眼底	0.169	0.996	0.142
动脉硬化	0.795	0.763	0.982
杯盘比增大	0.758	0.651	0.997
其他黄斑变性	0.631	0.764	0.995
大玻璃膜疣/色素沉着	0.798	0.769	0.998
视网膜零星出血	0.776	0.823	0.998
轻度 NPDR	0.867	0.920	0.998
中度 NPDR	0.889	0.911	0.997
重度 NPDR	0.855	0.886	0.996
PDR	0.717	0.780	0.992

具优势;(3)卷积神经网络利用多卷积核操作,实现能够对于复杂图像信息进行多特征抽取,进一步提高了网络的抽象能力;(4)卷积神经网络中含有下采样层,能够有效地减少每层样本的数量,进一步提升模型的性能。

以往研究主要集中在人工智能诊断系统对于单病种的诊断一致性、准确性、敏感性和特异性的研究<sup>[14]</sup>。在真实世界中,患者的诊断往往并非单一,因此,我们通过人工智能诊断系统分析门诊患者的眼底照相,进而评估其一致性、准确性、敏感性和特异性。

人工智能诊断组对于患者的完整诊断的诊断准确性为 62.82%,其中 4 个诊断的诊断准确率最高为 86.95%,其次为 3 个诊断的准确率为 84.61%,2 个诊断的准确率为 77.96%,单个诊断准确率仅为 71.63%,5 个诊断的准确率最低为 60%。诊断个数是指同一患者同一眼别的诊断名称个数,由于单一患眼具有 5 个诊断的数量很少,在纳入患者中仅具有 5 眼别,其中 3 眼别的诊断完全一致,诊断准确率为 60%。在临床中仅靠眼底照相对单眼做出 5 个诊断的患者数量很少,这也是本研究存在的不足之处,我们也将在今后扩大样本量,从而完善人工智能系统对于多诊断的眼底照相的准确率研究,但是在本研究中,对于具有 5 个诊断的眼底照相,人工智能诊断系统具有 60%的准确率,也说明了人工智能眼底病变诊断系统在未来的临床工作中具有很好的应用前景。

在 21 个眼底病常用诊断中,我们发现除无明显异常及豹纹状眼底外,其余诊断的一致性 Kappa 值高达 0.57~1,敏感性高达 0.651~1,特异性高达 0.93~1。虽然在无明显异常及豹纹状眼底诊断中 Kappa 值较低,分别为 0.044、0.169,然而其 P 值小于 0.05,提示人工智能诊断组与专家组之间具有诊断一致性。但是其一致性非常低。

同时,我们发现豹纹状眼底的诊断一致性、特异性分别为 Kappa 值 0.169,特异性 0.142,说明这可能和我们在人工智能系统中关于豹纹状眼底的深度学习算法有关,单纯通过眼底照相可能存在一定的局限性,尚需通过深度学习结合患者的验光结果等进一步加强和完善豹纹状眼底的诊断一致性和特异性。通过数据分析发现,在单个诊断中,共 501 眼由临床专家组诊断为无明显异常,而人工智能诊断组认为其中有 494 眼为豹纹状眼底。当去除豹纹状眼底诊断时,单个诊断的准确率提升至 87.54%。因此,我们推测,该诊断系统可以随着深度学习增加从而得到进一步优化,具有很大的提升空间。

目前大多数人工智能在眼底病的研究主要集中在人工智能诊断系统对于糖尿病性视网膜病变的应用价值。He 等<sup>[3]</sup>和 Gargeya 等利用深度学习算法对糖尿病患者的眼底照相进行识别以诊断 DR,都获得了大于 90%的敏感性和特异性<sup>[12]</sup>。翁铭等<sup>[15]</sup>通过人工智能诊断系统对糖尿病患者进行诊断,并将 DR 分为轻度 NPDR、中度 NPDR、重度 NPDR 及 PDR 进行分析,得出人工智能诊断系统的敏感性为 0.82,特异性为 0.91。本研究结果显示,在患者诊断中糖尿病性视网膜病变(轻度 NPDR、中度 NPDR、重度 NPDR、PDR)的敏感性和特异性分别为 0.920、0.911、0.886、0.780 和 0.998、0.997、0.996、0.992,其中敏感性较以往研究类似,而特异性大于 99%。这可能由于筛查人群的不同,既往研究主要针对糖尿病患者的 DR 筛查,而本研究所纳入门诊患者中非糖尿病性视网膜病变患者占大多数,提示人工智能对于 DR 患者的诊断具有高度特异性,在筛查诊断中具有极低的假阴性,对于 DR 筛查工作具有重要的临床意义。

本研究结果显示,人工智能诊断系统对于本试验中大多数诊断均具有较高的敏感性和特异性,然而不同的眼底疾病都有其临床诊断金标准,单依赖于眼底照相可能无法做出精确的诊断。青光眼是第二大致盲疾病,人工智能诊断系统在青光眼方面的应用主要在检测视网膜神经纤维层厚度、视野和杯盘比等方面<sup>[16-17]</sup>。Diaz-Pinto 等<sup>[18]</sup>研究认为基于卷积神经网络的人工智能诊断系统通过眼底照相诊断青光眼的敏感性为 0.934,特异性为 0.858。本研究从眼底照相中通过识别杯盘比的大小确认其是否异常,具有高度一致性,且有较好的敏感性 0.651 和高度特异性 0.997。本研究敏感性下降的主要原因可能为 Diaz-Pinto 的研究对象为具有 56%青光眼患者的眼底图库,而本研究对象为门诊随机患者,青光眼确诊患者常伴有明显的杯盘比增大,从而有利于人工智能诊断系统进行确诊,提高诊断敏感性。临床上对于青光眼的诊断仍需视野、OCT 等检查支持,因此本研究中以杯盘比增大作为诊断,我们认为,随着人工智能诊断系统的深度学习的加强和优化,其对于人群筛查具有重大的临床意义。本研究结果显示,人工智能诊断系统在黄斑前膜、黄斑变性、ARMD 玻璃膜疣等黄斑疾病诊断抑或是视网膜大动脉瘤、分支静脉阻塞、视网膜零星出血中,都具有高度一致性、敏感性和特异性。本研究中的人工智能诊断系统目前主要通过眼底照相的细节识别处理,而对于该类眼底疾病,临床上常需要进一步行 OCT、眼底血管造影等检查以进一步明确诊断。因此,在后续人工智能诊断系统研究中,将会纳入更多眼科检查结果,通过多维度的深度卷积神经网络学习不断更新人工智能诊断系统,使之

能够更为智能化的服务眼科医疗。

综上所述,基于多层深度卷积神经网络学习的人工智能诊断系统对于绝大多数眼底疾病可以达到较高的敏感性和特异性,适用于眼底疾病的筛查工作,特别有利于基层医院或社区医院对于眼底疾病的初步诊断。本研究中也存在一定的不足之处,比如眼底疾病的病种仍有一定的局限性,部分病种的样本量较小,部分眼底病变未采用其他检查做为诊断参考标准等。因此,我们将进一步扩大眼底病种,增大样本量,完善人工智能诊断模型,以期对于常见眼底疾病能够做出更为准确的诊断,推动人工智能眼底病变诊断系统的发展。

#### 参考文献

- 1 Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, *et al.* Artificial intelligence in retina. *Prog Retin Eye Res* 2018;67:1-29
- 2 Ting DSW, Pasquale LR, Peng L, *et al.* Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019; 103 (2): 167-175
- 3 He J, Cao T, Xu F, *et al.* Artificial intelligence-based screening for diabetic retinopathy at community hospital. *Eye (Lond)* 2020;34(3): 572-576
- 4 Kanagasalingam Y, Xiao D, Vignarajan J, *et al.* Evaluation of Artificial Intelligence-Based Grading of Diabetic Retinopathy in Primary Care. *JAMA Netw Open* 2018;1(5):e182665
- 5 Rogers TW, Jaccard N, Carbonaro F, *et al.* Evaluation of an AI system for the automated detection of glaucoma from stereoscopic optic disc photographs:the European Optic Disc Assessment Study. *Eye (Lond)* 2019;33(11):1791-1797
- 6 Phan S, Satoh S, Yoda Y, *et al.* Evaluation of deep convolutional neural networks for glaucoma detection. *Jpn J Ophthalmol* 2019;63(3): 276-283

- 7 Russakoff DB, Lamin A, Oakley JD, *et al.* Deep Learning for Prediction of AMD Progression: A Pilot Study. *Invest Ophthalmol Vis Sci* 2019;60(2):712-722
- 8 Schmidt-Erfurth U, Waldstein SM, Klmscha S, *et al.* Prediction of Individual Disease Conversion in Early AMD Using Artificial Intelligence. *Invest Ophthalmol Vis Sci* 2018;59(8):3199-3208
- 9 Devalla SK, Liang Z, Pham TH, *et al.* Glaucoma management in the era of artificial intelligence. *Br J Ophthalmol* 2020;104(3):301-311
- 10 赵金凤,吴桢泉,郑磊,等. 人工智能在眼科疾病诊疗的应用现状. *眼科新进展* 2019;39(5):495-500
- 11 Lin E, Tsai SJ. Machine Learning in Neural Networks. *Adv Exp Med Biol* 2019;1192:127-137
- 12 周飞燕,金林鹏,董军. 卷积神经网络研究综述. *计算机学报* 2017;40(6):1229-1251
- 13 常亮,邓小明,周明全,等. 图像理解中的卷积神经网络. *自动化学报* 2016;42(9):1300-1312
- 14 Kuwayama S, Ayatsuka Y, Yanagisono D, *et al.* Automated Detection of Macular Diseases by Optical Coherence Tomography and Artificial Intelligence Machine Learning of Optical Coherence Tomography Images. *J Ophthalmol* 2019;2019:6319581
- 15 翁铭,郑博,吴茂念,等. 基于深度学习的DR筛查智能诊断系统的初步研究. *国际眼科杂志* 2018;18(3):568-571
- 16 Zheng C, Johnson TV, Garg A, *et al.* Artificial intelligence in glaucoma. *Curr Opin Ophthalmol* 2019;30(2):97-103
- 17 Hemelings R, Elen B, Barbosa-Breda J, *et al.* Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning. *Acta Ophthalmol* 2020;98(1): e94-e100
- 18 Diaz-Pinto A, Morales S, Naranjo V, *et al.* CNNs for automatic glaucoma assessment using fundus images: an extensive validation. *Biomed Eng Online* 2019;18(1):29