

# 眼底阅片人工智能系统在糖尿病视网膜病变筛查中的临床价值评价

李萌<sup>1</sup> 王耿媛<sup>1</sup> 夏鸿慧<sup>2</sup> 唐晓颖<sup>3</sup> 冯子卿<sup>2</sup> 姚永屹<sup>2</sup> 黄义劲<sup>3</sup> 范卫<sup>2</sup> 袁哲<sup>3</sup> 袁进<sup>1</sup>

<sup>1</sup>中山大学中山眼科中心 眼科学国家重点实验室,广州 510060;<sup>2</sup>广东省肇庆市高要区人民医院眼科 526040;<sup>3</sup>南方科技大学电子与电气工程系,深圳 518055

通信作者:袁进,Email:yuanjincornea@126.com

**【摘要】** 目的 比较基于眼底阅片的人工智能(AI)系统与医生阅片在糖尿病视网膜病变(DR)筛查中的诊断效率和准确性,评价AI系统的临床应用价值。方法 采用诊断试验研究方法,取2017年3月至2018年11月肇庆市高要区人民医院眼科就诊患者彩色眼底照片13 683张作为图像库,通过迁移学习+NASNet架构算法,对2~3位眼科高级职称医生精确标注的4 465张彩色眼底照片(正常彩色眼底照片2 510张,DR彩色眼底照片1 955张)学习,建立针对DR的AI系统(ZOC-DR-V1)。1 000张确诊的彩色眼底照片(正常彩色眼底照片300张,不同程度DR彩色眼底照片700张),分别交由AI系统组和医生组(包含眼科医生组和内分泌科医生组,均包含初级、中级和高级职称),分别进行阅片,记录AI系统和医生组的单张阅片时间和总耗时,比较AI系统与不同级别医生阅片准确率和效率的差异。将1 000张图片以第500张为界分成前段和后段2个阅片阶段,分别统计医生组和AI系统组的分段诊断符合率和分段平均阅片时间。结果 眼科AI系统(ZOC-DR-V1)完成训练后测试集的诊断符合率为94.7%,受试者工作特征曲线(ROC)曲线下面积(AUC)为0.994。在人机对抗中,内分泌科初、中、高级职称医生诊断符合率分别为94.0%、91.4%和93.4%,眼科组初、中、高级职称医生诊断符合率分别为92.7%、94.4%和95.6%,AI系统诊断符合率为95.2%。AI系统与眼科高级医生阅片诊断符合率比较,差异无统计学意义( $\chi^2=0.182, P=0.749$ )。内分泌科组初、中、高级职称医生单张平均阅片时间分别为(4.63±1.87)、(3.74±3.47)和(5.71±3.47)s,总耗时分别为1.29、1.04和1.58h;眼科医生组初级、中级、高级职称医生单张阅片时间分别为(7.25±6.58)s、(5.18±5.01)和(5.18±3.47)s,总耗时分别为2.02、1.44和1.44h;AI系统单张平均阅片时间和总耗时分别为(1.62±0.67)s和0.45h。AI系统单张平均阅片时间明显少于各医生组,差异均有统计学意义(均 $P<0.01$ )。内分泌科初级职称、眼科初级职称和眼科中级职称医生前段与后段阅片区间诊断符合率比较,差异均有统计学意义( $\chi^2=11.986, 6.517, 10.896$ , 均 $P<0.05$ );内分泌科中级职称和高级职称医生、眼科初级职称医生后段阅片区间的平均阅片时间较前段明显缩短,差异均有统计学意义( $t=4.175, 8.189, 5.160$ , 均 $P<0.01$ );AI系统分段阅片诊断符合率及平均阅片时间均保持稳定,差异均无统计学意义( $\chi^2=3.151, P=0.103; t=0.038, P=0.970$ )。结论 基于眼底阅片的眼科AI系统诊断符合率可达眼科专业高级职称医生水平,且平均阅片时间短,阅片质量稳定,可为大规模DR人群筛查提供新的方法与平台。

**【关键词】** 人工智能;糖尿病视网膜病变/诊断;筛查;阅片

基金项目:国家重点研发计划项目(2017YFC0112400、2017YFC0112405)

DOI:10.3760/ema.j.issn.2095-0160.2019.08.015

## Clinical evaluation of artificial intelligence system based on fundus photograph in diabetic retinopathy screening

Li Meng<sup>1</sup>, Wang Gengyuan<sup>1</sup>, Xia Honghui<sup>2</sup>, Tang Xiaoying<sup>3</sup>, Feng Ziqing<sup>2</sup>, Yao Yongyu<sup>2</sup>, Huang Yijin<sup>3</sup>, Fan Wei<sup>2</sup>, Yuan Zhe<sup>3</sup>, Yuan Jin<sup>1</sup>

<sup>1</sup>Zhongshan Ophthalmic Center, Sun Yat-sen University, State Key Laboratory of Ophthalmology, Guangzhou 510060, China; <sup>2</sup>Department of Ophthalmology, Zhaoqing Gaoyao People's Hospital, Zhaoqing 526040, China; <sup>3</sup>Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China  
Corresponding author: Yuan Jin, Email:yuanjincornea@126.com

**【Abstract】 Objective** To study the efficiency and accuracy of artificial intelligence (AI) system based on fundus photograph in diabetic retinopathy (DR) screening, and evaluate the clinical application value of AI system.

**Methods** A diagnostic trial was adopted. Total of 13 683 color fundus photos were collected in Zhaoqing Gaoyao

People's Hospital from March, 2017 to November, 2018. The AI system for DR (ZOC-DR-V1) was established, based on transfer learning + NASNet algorithm, by training 4 465 precisely labeled fundus images (2 510 normal, and 1 955 with any stage of DR). One thousand confirmed fundus images (300 normal and 700 with any stage of DR), diagnosed by AI (AI group) and doctors (3 ophthalmologist doctors and 3 endocrinologist doctors) (doctor group), respectively. Ophthalmologist group and endocrinologist group were both composed of primary, intermediate and senior physicians. The mean reading time of each image and the total time of 1 000 images were recorded. The accuracy and efficiency of AI system and doctor groups were compared. The reading process was divided into two stages. The diagnostic coincidence rate and the average reading time of each group between the two parts were calculated and compared. This study protocol was approved by Ethic Committee of Zhongshan Ophthalmic Center, Sun Yat-sen University (No. 2017KYPJ104). **Results** After training, the diagnostic coincidence rate of AI system (ZOC-DR-V1) in test set was 94.7%, AUC was 0.994. In this "man-machine to war", the diagnostic coincidence rate of primary, intermediate and senior endocrinologist was 94.0%, 91.4% and 93.4%; the diagnostic coincidence rate of primary, intermediate and senior ophthalmologist was 92.7%, 94.4% and 95.6%; the diagnostic coincidence rate of AI system was 95.2%. There was no difference in the diagnostic coincidence rate between AI system and senior ophthalmologist ( $P = 0.749$ ). The mean reading time of each image of primary, intermediate and senior endocrinologists was  $(4.63 \pm 1.87)$ ,  $(3.74 \pm 3.47)$  and  $(5.71 \pm 3.47)$  seconds, and the total time of 1 000 images of primary, intermediate and senior endocrinologists was 1.29, 1.04 and 1.58 hours; the mean reading time of each image of primary, intermediate and senior ophthalmologists was  $(7.25 \pm 6.58)$ ,  $(5.18 \pm 5.01)$  and  $(5.18 \pm 3.47)$  seconds, and the total time of 1 000 images of primary, intermediate and senior endocrinologists was 2.02, 1.44 and 1.44 hours; the mean and total time of AI system was  $(1.62 \pm 0.67)$  seconds and 0.45 hours, and the reading time of AI system was significantly shorter than that of the doctor groups (all at  $P = 0.000$ ). The diagnostic coincidence rates between previous and posterior part of primary endocrinologist, primary and intermediate ophthalmologist were significantly different ( $\chi^2 = 11.986, 6.517, 10.896$ ; all at  $P < 0.05$ ), and the mean reading time in the posterior part was significantly shorter than that in the previous part of intermediate and senior endocrinologist and primary ophthalmologist ( $t = 4.175, 8.189, 5.160$ ; all at  $P < 0.01$ ). While the reading time of AI system remained stable throughout the process ( $\chi^2 = 3.151, P = 0.103; t = 0.038, P = 0.970$ ). **Conclusions** The ophthalmic AI system based on fundus images has a good diagnostic efficiency, and its diagnostic coincidence rate can compare with senior ophthalmologist, providing a new method and platform for large-scale DR screening.

[Key words] Artificial intelligence; Diabetic retinopathy/diagnosis; Screening; Reading

Fund program: National Key R&D Program of China (2017YFC0112400, 2017YFC0112405)

DOI:10.3760/cma.j.issn.2095-0160.2019.08.015

糖尿病视网膜病变(diabetic retinopathy, DR)是糖尿病常见的微血管并发症。根据国际糖尿病联盟(International Diabetes Federation, IDF)报道,中国糖尿病患者人数约有 1.1 亿<sup>[1]</sup>。中国糖尿病患者并发 DR、非增生性 DR(non-proliferative DR, NPDR)和增生性 DR(proliferative DR, PDR)的患病率分别达 18.45%、15.06% 和 0.99%<sup>[2]</sup>。随着糖尿病病程的延长,DR 患病风险逐渐增高,病情加重<sup>[3]</sup>。此外,由于经济条件、医疗资源分布和受教育程度等因素的影响,中国 DR 患病率分布呈农村地区高于城市地区的特点<sup>[4]</sup>。由于医院专科设置和医疗资源分布等原因,许多首诊于内分泌科的糖尿病患者未能得到规范的 DR 诊治和健康建议。另外,在偏远山区和农村地区,受限于经济、卫生、教育条件,很多糖尿病患者并不能认识到 DR 的危害,往往在视功能严重受损后才寻求眼科医生的救助。2017 年美国糖尿病学会(American Diabetes Association, ADA)指出,对于糖尿病患者进行定期随

访、早期检查,并对威胁视力的 DR 进行治疗,能够预防 98% DR 引起的视力丧失<sup>[5]</sup>。眼底照相是 DR 筛查的重要手段,不仅可以观察患者的眼底情况,还可以将保存的眼底照片用于随访比对、远程阅片会诊。目前,糖尿病住院患者 DR 筛查多通过眼科会诊进行,内分泌科医生对于 DR 的诊断水平参差不齐;同时,社区糖尿病患者 DR 筛查工作量大,而远程阅片的时效性和质量控制不尽如人意。2017 年斯坦福联合研究团队发表 Nature 封面文章,运用深度学习网络对大量人类皮肤损伤疾病图片进行训练分析可对皮肤癌进行诊断,准确率可达 91%<sup>[6]</sup>,这一研究成果标志着人工智能(artificial intelligence, AI)算法在医学领域进入应用型研究阶段。众多学者在 DR 的 AI 诊断算法上进行了探索,研究报道 AI 算法对 DR 的诊断灵敏度和特异度均可达 90% 以上<sup>[7-10]</sup>,但基于 AI 算法在 DR 筛查中的具体应用效能及其与专科医生阅片相比的优越性尚未见文献报道。本研究中将 AI 算法与专业医生在 DR

筛查中阅片诊断速度、正确率和稳定性进行对比,进一步说明 AI 系统的应用价值。

## 1 资料与方法

### 1.1 AI 数据图像集的收集及建立

采用诊断试验研究方法,取 2017 年 3 月至 2018 年 11 月肇庆市高要区人民医院眼科就诊患者彩色眼底照片 13 683 张作为图像库,包括正常眼底 8 089 张,DR 5 594 张(根据国际 DR 分期标准,其中轻、中、重度 NPDR 及 PDR 分别为 181、3 336、1 438 和 639 张),为扩瞳后 Topcon TRC-50DX 照相机拍摄的后极部以黄斑中心凹为中心 35° 范围的照片,每张图像辅以基本病例资料,包括性别、年龄、全身病史(糖尿病病程)、视力、裂隙灯显微镜检查结果,经 2 名高级职称眼科医生使用由中山眼科中心自主研发彩色眼底照相阅片分类软件 ZOC-Screen-V2.0(图 1A)分别阅片,2 名医生诊断意见一致时为确定诊断,意见不一致时交由第 3 名高级职称医生阅片。对确诊为 DR 的彩色眼底照片进行病灶标注(图 1B)。本研究经中山大学中山眼科中心医学伦理委员会批准(审批号:2017KYPJ104)。



图 1 彩色眼底照相阅片分类软件(ZOC-Screen-V2.0)及标注效果图 A:彩色眼底照相阅片分类软件(ZOC-Screen-V2.0)界面 B:图像病灶标注 绿色圆圈标注硬性渗出,蓝色圆圈标注软性渗出,红色圆圈标注微血管瘤

### 1.2 方法

**1.2.1 AI 算法的建立** 通过迁移学习+NASNet 架构算法,选取 4 465 张图像清晰、诊断明确并经过精准标记的彩色眼底照片进行学习,包括正常眼底照片 2 510 张,DR 眼底照片 1 955 张,建立针对 DR 的 AI 系统(ZOC-DR-V1)。取 952 张彩色眼底照片(正常眼底 417 张,DR 535 张)测试 AI 系统灵敏度、特异度和诊断符合率,绘制受试者工作特征曲线(receiver operator characteristic curve, ROC 曲线)。

**1.2.2 AI 临床评价** 另取 1 000 张成像清晰、诊断明确的彩色眼底照片作为人机对抗用图,其中正常彩色眼底照片 300 张,DR 眼底照片 700 张。正常彩色眼底照片纳入标准:(1)彩色眼底照片正常;(2)双眼屈光度  $\leq \pm 6.00$  D。正常彩色眼底照片排除标准:(1)有家

族性或遗传性视网膜病变;(2)有眼外伤病史和/或除低中度屈光不正和/或轻度年龄相关性白内障除外的其他眼部疾病史者;(3)有可能引起眼部不良反应的药物用药史,如氯喹、羟氯喹、氯丙嗪、利福平等者;(4)有可能影响视网膜的系统性疾病者。DR 彩色眼底照片纳入标准:(1)确诊 1 型糖尿病 5 年以上患者;(2)确诊 2 型糖尿病患者;(3)确诊妊娠期糖尿病患者;(4)符合 DR 诊疗指南中糖尿病视网膜病变分期中任意一期和/或糖尿病黄斑水肿;(5)除糖尿病外,无可能影响视网膜的系统性疾病。DR 彩色眼底照片排除标准:(1)屈光间质混浊,彩色眼底照相不能清晰成像者;(2)合并其他眼底病和/或其他严重影响视功能疾病者;(3)因智力、精神等原因不能配合检查者。将 6 名研究人员根据其工作性质分为 AI 系统组和医生组(包括眼科医生组和内分泌科医生组,均包含初级、中级和高级职称各 1 名),各组均对所有彩色图片进行阅片,记录 AI 系统组和医生组的单张阅片时间和总耗时,比较 AI 阅片系统与不同级别医生阅片的诊断符合率和效率。将 1 000 张图片按阅片顺序平均分为 2 段:前段为第 1~500 张;后段为第 501~1 000 张,分别统计医生组和 AI 系统组的分段诊断符合率和分段平均阅片时间。

### 1.3 统计学方法

采用 SPSS 22.0 统计学软件进行统计分析。计量资料的数据经 S-W 检验证实呈正态分布,以  $\text{mean} \pm \text{SD}$  表示,经 Levene 检验证实方差齐,AI 系统与不同级别医生单张平均阅片时间比较采用单因素方差分析,两两比较采用 LSD-*t* 检验。诊断符合率以百分数表示,AI 系统与眼科高级医生阅片诊断符合率、AI 系统和不同级别医生前段与后段阅片诊断符合率比较均采用卡方检验。 $P < 0.05$  为差异有统计学意义。

## 2 结果

### 2.1 AI 系统的性能评价

通过迁移学习+NASNet 架构算法和深度学习,测试集灵敏度为 96.89%,特异度为 93.57%,诊断符合率为 94.7%,ROC 曲线下面积

(area under curve, AUC)为 0.994(图 2)。

### 2.2 3 个组阅片诊断灵敏度、特异度和诊断符合率比较

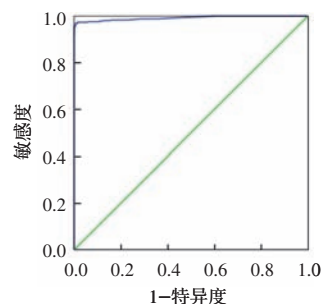


图 2 AI 系统(ZOC-DR-V1)ROC 曲线



在人机对抗中, AI 系统与眼科高级医生阅片诊断符合率比较, 差异无统计学意义 ( $\chi^2 = 0.182, P = 0.749$ ) (表 1)。

表 1 AI 系统与不同级别医生阅片灵敏度、特异度及符合率比较

组别	灵敏度 (%)	特异度 (%)	诊断符合率 [% (n/n)]
内分泌科医生组初级职称	93.57	95.00	94.0 (940/1 000)
内分泌科医生组中级职称	97.42	77.33	91.4 (914/1 000)
内分泌科医生组高级职称	91.86	97.00	93.4 (934/1 000)
眼科医生组初级职称	89.71	99.67	92.7 (927/1 000)
眼科医生组中级职称	98.14	85.67	94.4 (944/1 000)
眼科医生组高级职称	96.00	94.67	95.6 (956/1 000)
AI 系统组	96.43	92.33	95.2 (952/1 000)

注: AI: 人工智能

### 2.3 3 个组阅片时间比较

内分泌科组初、中、高级职称医生单张平均阅片时间分别为 (4.63±1.87)、(3.74±3.47) 和 (5.71±3.47) s, 总耗时分别为 1.29、1.04 和 1.58 h; 眼科医生组初级、中级、高级职称医生单张平均阅片时间分别为 (7.25±6.58)、(5.18±5.01) 和 (5.18±3.47) s, 总耗时分别为 2.02、1.44 和 1.44 h; AI 系统单张平均阅片时间和总耗时分别为 (1.62±0.67) s 和 0.45 h。各组平均阅片时间存在统计学差异 ( $F = 30.543, P < 0.001$ ), AI 系统单张平均阅片时间明显少于内分泌科初、中、高级职称

医生组以及眼科初、中、高级职称医生组, 差异均有统计学意义 ( $t = 5.075、8.967、17.296、23.509、22.415、9.018$ , 均  $P < 0.001$ ) (图 3)。

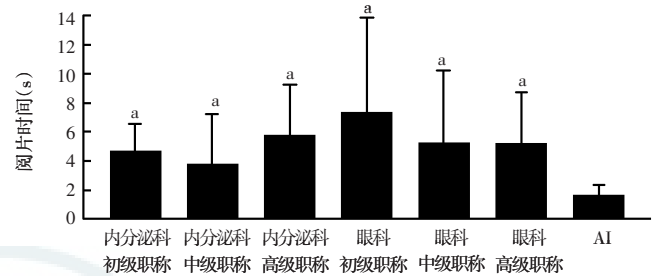


图 3 AI 系统与不同级别医生单张平均阅片时间比较 与 AI 系统组比较,  $^a P < 0.01$  (单因素方差分析, LSD- $t$  检验) AI: 人工智能

### 2.4 3 个组阅片结果稳定性比较

医生组中内分泌科初级职称医生、眼科初级职称和眼科中级职称医生前段和后段阅片区间诊断符合率比较, 差异均有统计学意义 ( $\chi^2 = 11.986、6.517、10.896$ , 均  $P < 0.05$ ) (表 2)。内分泌科中级职称和高级职称医生、眼科初级职称医生后段阅片区间的平均阅片时间较前段明显缩短, 差异均有统计学意义 ( $t = 4.175、8.189、5.160$ , 均  $P < 0.01$ )。AI 系统分段阅片诊断符合率及平均阅片时间均保持稳定, 差异均无统计学意义 ( $\chi^2 = 3.151, P = 0.103; t = 0.038, P = 0.970$ )。

表 2 AI 系统和不同级别医生前段与后段阅片诊断符合率比较 [% (n/n)]

时间段	总例数	内分泌科医生诊断符合率			眼科医生诊断符合率			AI 系统诊断符合率
		初级职称	中级职称	高级职称	初级职称	中级职称	高级职称	
前段	500	91.4 (457/500)	92.8 (464/500)	91.8 (459/500)	90.6 (453/500)	96.8 (484/500)	94.2 (471/500)	94.0 (470/500)
后段	500	96.6 (483/500)	90.0 (450/500)	95.0 (475/500)	94.8 (474/500)	88.8 (444/500)	96.8 (484/500)	96.4 (482/500)
$\chi^2$ 值		11.986	2.494	4.153	6.517	10.896	3.933	3.151
P 值		0.001	0.142	0.055	0.015	0.001	0.066	0.103

注: AI: 人工智能 ( $\chi^2$  检验)

表 3 AI 系统及不同级别内分泌科、眼科医生前段与后段单张平均阅片时间比较 (mean±SD, s)

时间段	例数	内分泌科医生单张平均阅片时间			眼科医生单张平均阅片时间			AI 系统单张平均阅片时间
		初级职称	中级职称	高级职称	初级职称	中级职称	高级职称	
前段	500	5.53±2.60	4.72±1.00	7.58±0.88	8.47±0.78	5.41±0.54	5.64±0.58	1.62±0.67
后段	500	3.74±0.52	2.76±0.30	3.83±0.52	6.03±0.72	4.95±0.46	4.71±1.66	1.62±0.67
$t$ 值		1.627	4.175	8.189	5.16	1.449	1.176	0.038
P 值		0.104	<0.01	<0.01	<0.01	0.148	0.24	0.97

注: AI: 人工智能 (配对  $t$  检验)

### 3 讨论

DR 的筛查手段包括检眼镜检查、彩色眼底照相检查、激光扫描检眼镜 (scanning laser ophthalmoscope, SLO) 检查、光相干断层扫描 (optical coherence tomography, OCT)、多光谱眼底分层成像 (multiple spectrum imaging,

MSI) 等。检眼镜检查需要专业的眼科医生进行, 在眼科医疗资源不足的偏远或基层地区, 难以保证 DR 筛查的进行以及检查的准确性, 且不能留存患者影像资料, 不利于随访。SLO、OCT、MSI 等仪器设备昂贵, 难以在基层推广, 普及程度不高。彩色眼底照相仍然是目前 DR 筛查主要的检查手段<sup>[12]</sup>。随着医院信息化

建设逐步完善和互联网技术逐渐普及,我国已建立多个依托于区域高水平眼科医疗机构的眼科读片中心。远程眼科阅片系统的建立在一定程度上缓解了眼科医疗资源不足地区的眼科就诊压力,但在大量眼底图片涌入区域读片中心时,读片的时效性难以保证。同时,读片的准确性很大程度上取决于读片医生专业水平和责任心<sup>[13]</sup>,不同医生的读片水平参差不齐,尤其在高强度读片压力下,误诊率和漏诊率增加。AI 算法的发展为解决这些问题提供了途径。目前针对 DR 筛查的 AI 算法研究大部分基于彩色眼底照片,一方面前期的临床大数据积累以彩色眼底照片为主;另一方面,算法的实施和应用也以彩色眼底照片为基础,更具有推广性。

本研究使用的 AI 系统采用迁移学习+NASNet 架构算法,可在使用小批量数据的情况下,经过较短时间的训练学习取得相对较好的效果。近 3 年来文献报道的基于彩色眼底照片的 DR AI 辅助诊断系统灵敏度和特异度分别为 87%~99% 和 91.4%~99.0%,AUC 为 0.989~0.991<sup>[14-17]</sup>。本研究使用的 AI 系统测试集诊断符合率为 94.7%,AUC 为 0.994,可满足临床筛查需求。本研究中未对 DR 进行分级诊断,一方面考虑到基于单视野彩色眼底照片进行分级与实际临床分级存在差异;另一方面,目前的 DR AI 辅助诊断系统主要适用于 DR 筛查而不是确诊。

本次评价中 AI 系统诊断符合率与医生组中诊断符合率最高者——眼科高级职称医生的水平非常接近,灵敏度和特异度也接近。诊断符合率反映整体诊断正确率,灵敏度是衡量试验检测出患者的能力,特异度则反映将正常人正确地判别为无病者的能力。临床上不仅要求诊断方法具有高诊断符合率,同时也需要高灵敏度和特异度<sup>[18]</sup>。医生在诊断疾病时有时会基于自己的经验和诊断习惯,偏保守或偏激进,导致灵敏度偏低或者特异度偏低。AI 系统基于大量经临床金标准确诊彩色眼底照片的学习,可最大程度地避免个人诊断习惯造成的偏倚。

AI 系统在阅片时间上也具有明显优势。本研究中为了模拟现实工作中医生的工作时间和工作量,要求医生组在 1.5 h 内完成 1 000 张彩色眼底照片阅片,结果有 2 名医生未在规定时间内完成阅片,阅片总耗时最长的医生的阅片总时长是 AI 系统阅片总时长的 4.89 倍。AI 系统的运行速度与计算机硬件配置、处理系统、AI 算法复杂程度、图形处理器部署等因素有关。本研究中采用的 AI 系统为离线版本,所用计算机处理器为 Intel Core i5-6300U,在 Windows 10 64 位系统环境

下运行,在优化配置后,AI 系统运行速度仍有提升空间。相比于医生阅片,AI 系统在高工作强度、长时间持续工作条件下,例如区域集中阅片、大样本人群筛查等,更具有优势。

医生在阅片时,工作效率会受注意力集中程度、思维活跃程度、熟练程度等因素影响。本研究中将阅片过程人为地划分为前段和后段,结果显示 AI 系统在前段和后段阅片过程中阅片时间和诊断符合率均保持稳定,但医生组整体上后段平均阅片时间少于前段平均阅片时间,可能是由于前段阅片过程处于熟悉、适应过程,且阅片时间宽裕,而后段阅片过程中医生已经熟悉图片的特征、掌握了阅片技巧,且越往后越临近规定的阅片完成时限,医生有意加快了阅片速度。另一方面,在阅片熟练程度增加和时间压力的双重作用下,医生的前、后段阅片诊断符合率呈现出 2 种不同的趋势。但总体来讲,医生阅片的效果存在波动性。

本研究中纳入的彩色眼底照片均为质量良好、病灶明显、诊断明确的图片,以减小图像在诊断上的争议。此种情况下,不同级别医生之间对图像的判别差异不大,阅片的差异并不十分明显。就 AI 系统而言,本次测试诊断符合率也较我们内部测试集测试结果略有提高,尽管不存在统计学差异,但仍一定程度上说明图像质量对于 AI 系统的诊断结果起到影响作用。在实际应用过程中,也需要就彩色眼底照片图像质量制定一定的标准,以保证 AI 系统诊断的可靠性。

AI 系统的建立基于对已由金标准确诊的海量影像资料进行学习、训练,当数据库确定后,AI 系统开发的耗时主要集中于算法研发。以 Google 旗下 DeepMind 公司研发的 AlphaGo 为例,2016 年 AlphaGo 与围棋世界冠军、职业九段棋手李世石进行围棋人机大战,以 4 比 1 的总比分大获全胜<sup>[19]</sup>。在此基础上,DeepMind 进一步推出 AlphaGo Zero,AlphaGo Zero 以 100% 胜率打败之前版本 AlphaGo。版本更新所花费时间仅 1 年余,算法训练时间仅 3 d<sup>[20]</sup>。并且随着硬件不断改善、算法逐步优化,这一时间还将缩短。而在现实中,1 名住院医师通常需要经历本科期间 5 年、住院医师规范化培训 3 年,共 8 年的学习时间,相较于现实中医生培养周期来说,AI 算法训练周期显著缩短。同时,AI 算法可复制,具有通用性和推广性。

综上所述,本研究结果表明基于眼底阅片的眼科 AI 系统诊断符合率可达眼科专业高级职称医生水平,且平均阅片时间短,阅片质量稳定,可为大规模 DR 人群筛查提供新的方法与平台。本研究中仍存在一定的局限性:首先,未考虑 AI 系统对多种型号眼底相机采

集图像识别的通用和兼容性能;其次, AI 系统对 DR 进行分期诊断的效率需要进一步研究;未来 AI 系统在研发和落地应用中仍需要更进一步多中心、多层次临床研究验证。

利益冲突 所有作者均声明不存在利益冲突

## 参考文献

- [1] International Diabetes Federation. IDF Diabetes Atlas 8th Edition [EB/OL]. (2017-12-01) [2019-03-20]. <http://www.diabetesatlas.org>.
- [2] Song P, Yu J, Chan KY, et al. Prevalence, risk factors and burden of diabetic retinopathy in China: a systematic review and meta-analysis [J/OL]. *Glob Health*, 2018, 8(1): 010803 [2019-03-27]. <http://www.jogh.org/documents/issue201801/jogh-08-010803.pdf>. DOI: 10.7189/jogh.08.010803.
- [3] 王芳, 王春芳, 闫建林. 45 岁以上糖尿病患者中糖尿病视网膜病变的患病率调查及相关危险因素分析[J]. *中华实验眼科杂志*, 2013, 31(8): 783-787. DOI: 10.3760/cma.j.issn.2095-0160.2013.08.018.  
Wang F, Wang CF, Yan JL. Survey of diabetic retinopathy of the diabetic population over 45 years old [J]. *Chin J Exp Ophthalmol*, 2013, 31(8): 783-787. DOI: 10.3760/cma.j.issn.2095-0160.2013.08.018.
- [4] Liu L, Wu X, Liu L, et al. Prevalence of diabetic retinopathy in mainland China: a meta-analysis [J/OL]. *PLoS One*, 2012; 7: e45264. [2019-03-27]. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045264>. DOI: 10.1371/journal.pone.0045264.
- [5] Solomon SD, Chew E, Duh EJ, et al. Diabetic retinopathy: a position statement by the American Diabetes Association [J]. *Diabetes Care*, 2017, 40(3): 412-418. DOI: 10.2337/dc16-2641.
- [6] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks [J]. *Nature*, 2017, 542(7639): 115-118. DOI: 10.1038/nature21056.
- [7] Jin G, Xiao W, Ding X, et al. Prevalence of and risk factors for diabetic retinopathy in a rural Chinese population: the Yangxi Eye Study [J]. *Invest Ophthalmol Vis Sci*, 2018, 59(12): 5067-5073. DOI: 10.1167/iov.18-24280.
- [8] Zhang G, Chen H, Chen W, et al. Prevalence and risk factors for diabetic retinopathy in China: a multi-hospital-based cross-sectional study [J]. *Br J Ophthalmol*, 2017, 101(12): 1591-1595. DOI: 10.1136/bjophthalmol-2017-310316.
- [9] McGuire HC, Ji L, Kissimova-Skarbek K, et al. Type 1 diabetes mellitus care and education in China: The 3C study of coverage, cost, and care in Beijing and Shantou [J]. *Diabetes Res Clin Pract*, 2017, 129: 32-42. DOI: 10.1016/j.diabres.2017.02.027.
- [10] Lu J, Hou X, Zhang L, et al. Association between body mass index and diabetic retinopathy in Chinese patients with type 2 diabetes [J]. *Acta Diabetol*, 2015, 52(4): 701-708. DOI: 10.1007/s00592-014-0711-y.
- [11] Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search [J]. *Nature*, 2016, 529(7587): 484-489. DOI: 10.1038/nature16961.
- [12] Salz DA, Witkin AJ. Imaging in diabetic retinopathy [J]. *Middle East Afr J Ophthalmol*, 2015, 22(2): 145-150. DOI: 10.4103/0974-9233.151887.
- [13] 李建军, 徐亮, 路从磊, 等. 远程眼科阅片及时性及阅片报告详细性的初步研究 [J]. *眼科*, 2016, 25(1): 13-17. DOI: 10.13281/j.cnki.issn.1004-4469.2016.01.004.  
Li JJ, Xu L, Lu CL, et al. The timeliness of the image reading and the description details of the reading reports in teleophthalmology [J]. *Ophthalmol CHN*, 2016, 25(1): 13-17. DOI: 10.13281/j.cnki.issn.1004-4469.2016.01.004.
- [14] Ramachandran N, Hong SC, Sime MJ, et al. Diabetic retinopathy screening using deep neural network [J]. *Clin Exp Ophthalmol*, 2018, 46(4): 412-416. DOI: 10.1111/ceo.13056.
- [15] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs [J]. *JAMA*, 2016, 316(22): 2402-2410. DOI: 10.1001/jama.2016.17216.
- [16] Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning [J]. *Ophthalmology*, 2017, 124(7): 962-969. DOI: 10.1016/j.ophtha.2017.02.008.
- [17] Hemanth DJ, Anitha J, Son LH, et al. Diabetic retinopathy diagnosis from retinal images using modified hopfield neural network [J]. *J Med Syst*, 2018, 42(12): 247. DOI: 10.1007/s10916-018-1111-6.
- [18] Parikh R, Mathai A, Parikh S, et al. Understanding and using sensitivity, specificity and predictive values [J]. *Indian J Ophthalmol*, 2008, 56(1): 45-50. DOI: 10.4103/0301-4738.37595.
- [19] Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search [J]. *Nature*, 2016, 529(7587): 484-489. DOI: 10.1038/nature16961.
- [20] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge [J]. *Nature*, 2017, 550(7676): 354-359. DOI: 10.1038/nature24270.

(收稿日期: 2019-04-08 修回日期: 2019-07-06)

(本文编辑: 刘艳)

读者 · 作者 · 编者

## 本刊对论文中关键词的著录要求

本刊投稿的论文请分别在中英文摘要下方标引 3~5 个关键词以便于编制文献索引。关键词应选取能反映文章主题概念的词或词组, 中英文关键词应一致。投稿作者可登陆 <http://www.ncbi.nlm.nih.gov/mesh> 或 <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=mesh> 网站从美国国立医学图书馆的 MeSH 数据库中选取关键词, 其中文译名可参照中国医学科学院信息研究所编译的《医学主题词注释字顺表》。未被词表收录的新的专业术语(自由词)可直接作为关键词使用, 但应排序在最后。中医药关键词应从中国中医科学院中医药信息研究所编写的《中医药主题词表》中选取。关键词中的缩写词应按《医学主题词注释字顺表》还原为全称, 各关键词之间用“;”分隔。

(本刊编辑部)