

· 临床研究 ·

糖尿病视网膜病变人工智能机器人辅助诊断系统的建立及应用

高韶晖¹ 金学民¹ 赵朝霞¹ 于伟泓² 陈有信² 孙宇辉³ 丁大勇³¹河南省人民医院眼科 河南省立眼科医院 河南省眼科研究所, 郑州 450003; ²中国医学科学院 北京协和医院眼科, 北京 100730; ³北京致远慧图科技有限公司 100086

通信作者: 金学民, Email: jinxuemin64@163.com

【摘要】 目的 建立和评估基于深度学习理论的糖尿病视网膜病变(DR)人工智能(AI)机器人辅助诊断系统。**方法** 联合包括北京协和医院等国内 8 家医院眼底专业医师对 25 297 张糖尿病患者眼底图像病变进行标记, 训练和建立一个深度学习框架同时识别 DR 病变的 AI 机器人辅助诊断系统, 即“嵩岳”机器人系统。依据 DR 分级和标识眼底病变, 构建是否罹患 DR、是否需要转诊 DR、是否重度非增生性和增生性 DR 以及是否增生性 DR 4 种模型, 建立一个基于病变识别技术的 DR 筛选系统, 应用受试者工作特征曲线(ROC 曲线)、敏感度、特异度进行 AI 诊断性能的数据分析。**结果** “嵩岳”系统中是否罹患 DR 模型的敏感度为 96.0%, 特异度为 87.9%, 曲线下面积(AUC)为 0.920; 需要转诊 DR 模型的敏感度为 90.4%, 特异度为 95.2%, AUC 为 0.925; 是否为重度非增生性和增生性 DR 模型的敏感度为 72.7%, 特异度为 96.2%, AUC 为 0.845; 增生性 DR 模型的敏感度为 73.5%, 特异度为 97.3%, AUC 为 0.855。**结论** “嵩岳”AI 机器人辅助诊断系统具有精确高效的 DR 诊断性能, 具有良好的临床应用价值。

【关键词】 糖尿病视网膜病变; 人工智能; 机器人; 辅助诊断; 深度学习**基金项目:** 河南省科技攻关项目(162102410004); 河南省立眼科医院影像与人工智能研发平台

DOI: 10.3760/cma.j.issn.2095-0160.2019.08.016

Validation and application of an artificial intelligence robot assisted diagnosis system for diabetic retinopathy

Gao Shaohui¹, Jin Xuemin¹, Zhao Zhaoxia¹, Yu Weihong², Chen Youxin², Sun Yuhui³, Ding Dayong³¹Department of Ophthalmology, Henan Provincial People's Hospital, Henan Eye Hospital, Henan Eye Institute, Zhengzhou 450003, China; ²Department of Ophthalmology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China; ³Beijing Visionary Intelligence Technology Ltd., Beijing 100086, China

Corresponding author: Jin Xuemin, Email: jinxuemin64@163.com

[Abstract] Objective To evaluate the performance of an artificial intelligence (AI) assisted diagnosis system for diabetic retinopathy (DR) based on deep learning theory. **Methods** Diagnostic performance of a robot assisted diagnosis system called SongYue for DR was trained by using 25 297 retinal images tagged by fundus doctors from multiple hospitals in China. Four types of DR detection model consisting of abnormal DR, referable DR, severe non-proliferative and proliferative DR as well as proliferative DR according to fundus lesions identification were established. The ability of the system to distinguish DR was determined by using receiver operator characteristic (ROC) analysis, sensitivity and specificity of the system. **Results** SongYue system achieved an area under the ROC curve (AUC) of 0.920 for successfully distinguishing normal images from those DR with a sensitivity of 96.0% at a specificity of 87.9%. The AUC of SongYue for referable DR was 0.925, sensitivity was 90.4%, and specificity was 95.2%. For severe non-proliferative and proliferative DR, AUC was 0.845, sensitivity was 72.7%, and specificity was 96.2%. For proliferative DR, AUC was 0.855, sensitivity was 73.5%, and specificity was 97.3%. **Conclusions** SongYue robot assisted diagnosis system has high AUC, sensitivity and specificity for identifying DR, showing good clinical applicable benefits.

[Key words] Diabetic retinopathy; Artificial intelligence; Robot; Assisted diagnosis; Deep learning**Fund program:** Technological Project of Henan Province (162102410004); Image and Artificial Intelligence

Research and Development Platform of Henan Eye Hospital

DOI:10.3760/cma.j.issn.2095-0160.2019.08.016

据IDF的最新统计,2017年全世界约有4.25亿糖尿病患者,中国约有1.14亿,居世界首位,其中超过1/3的患者并发糖尿病视网膜病变(diabetic retinopathy, DR),其发病率呈逐年快速增长且有年轻化的趋势^[1]。DR的早期诊断对患者的预后起着关键作用,因而完善DR的早期诊断是亟待解决的重大问题。目前DR诊断主要依靠专业眼底病医生进行,不仅工作量大,耗时费力,且存在主观偏倚,尤其对发展中国家来说,由于受技术设备的限制,更影响了其深入开展^[2-4]。近来以深度学习为基础的人工智能(artificial intelligence, AI)技术逐渐兴起并用于基于眼科图像数据库AI的辅助诊断^[5-9],为此我们研发了利用卷积神经网络技术全面评估DR且能转诊分类的AI机器人,并对其临床应用价值进行评价。

1 资料与方法

1.1 资料

河南省人民医院牵头并联合包括北京协和医院在内的国内8家医院,与中国人民大学数据工程与知识工程教育部重点实验室、北京致远慧图科技有限公司合作研制出对多种眼底疾病,包括DR做出AI辅助诊断的“嵩岳”机器人医生。本系统运用基于深度学习的EyeWisdom AI引擎对美国EyePACS DR网站Kaggle DR评测数据集21 512张和北京协和医院内分泌科的3 785张(其中2型糖尿病患者1 634例,1型糖尿病患者251例,参考文献^[1]的诊断标准)共25 297帧眼底图像进行标注,累积病变标注次数达1 003 996次。图像视角均为45°(黄斑中心凹为中心),来自Kaggle DR的图像眼底摄像系统包括Canon CR1/DGi/CR2和Topcon NW,内分泌科图像的相机型号为Canon CR2 AF,均为自然瞳孔下眼底图像。

1.2 方法

1.2.1 图像质量控制及病变标注 DR彩色眼底照相质量标准参照文献^[10]的标准,在Kaggle DR评测数据集中人工筛选出部分质量好和部分无法标注的低质量图像,以此为基础,训练一个深度卷积神经网络作为图像质量分类器。通过该分类器将88 702张Kaggle中的21 839张低质量的图像剔除,同时研究团队进行更细致的图像质量甄别,进一步删除不符合标准的图像。

为了训练病变检测模型并检验其有效性,本研究中对数据作如下划分。从上述25 297张图片中,任取

2 248张图片作为DR病变测试集。剩余的23 049张用于训练DR学习模型。为了训练基于病变检测方法的需要转诊的DR筛查模型并评估其有效性,从病变测试集中任取1 648张作为筛查训练集,剩余600张作为筛查测试集。

虽然Kaggle数据库配有DR分级,但部分图片分级不够准确,本研究中对剔除了低质量的图片后,研究团队成员对符合的图像进行重新标注。为确保标注质量,1张图片3位医师同时标注,对存在分歧的标注意见,按多数意见执行,若仍不能达成一致则提交由5名资深眼底医生组成的审阅小组共同讨论,形成终审意见,并安排1名合格的质量抽检员随机对标注内容和分级程度进行对比检查,对有争议的图片,同样提交审阅小组共同讨论。最终共有213张图片因标注争议提交审阅小组共同商定并最终达成标注分级共识。此外,因图像在拍摄过程中可能会受到光线、屈光间质、瞬目及眼球运动的影响,所有纳入的眼底图像均配备有增强图,以便更清晰直观地显示病灶特点,保证标注质量(图1,2)。



图1 DR眼底拍摄图示例 图2 DR眼底增强图示例

1.2.2 DR分级和病变标注 DR分级依据International Clinical Diabetic Retinopathy Scale的标准分为无DR、轻度非增生性DR(non-proliferative DR, NPDR)、中度NPDR、重度NPDR和增生型DR(proliferative DR, PDR)(DR0~DR4)^[11],其中中度、重度NPDR和PDR为需要转诊的DR,无和轻度NPDR为不需要转诊的DR^[12],标注包括微血管瘤、视网膜内出血、硬性渗出、棉绒斑、静脉串珠、玻璃膜疣、纤维增生、玻璃体积血、视网膜前膜、视网膜脱离等在内的共28项眼底病变。对已行眼底激光的DR患者,根据有无增生期表现,分为激光后稳定/不稳定,对有合并高血压、黄斑变性等其他眼底病变的图像,在合并症选项

上标注有/无。因考虑到仅凭 1 张彩色眼底照相的二维图片无法确定是否存在黄斑水肿,因此,阅片分级中对于糖尿病黄斑水肿未做要求,在本研究中所提到的需要转诊的 DR 中不包含糖尿病黄斑水肿。根据 DR 不同分级,通过建立是否罹患 DR、需转诊 DR、是否为重度 NPDR 和 PDR 以及是否为 PDR 4 种模型。

1.2.3 “嵩岳”机器人 DR 深度模型算法 “嵩岳”机器人基于深度学习先进的 YOLO 源码解析,是端到端的实时目标检测系统,与大部分目标检测和识别方法不同,YOLO 将目标区域预测和目标类别预测整合于单个神经网络模型中,实现在准确率较高的情况下快速目标检测与识别,更加适合现场应用环境。YOLO 将单个神经网络应用于完整图像,该网络将图像划分为若干区域,然后预测每个区域的边界框和概率,这些边界框再由预测概率加权,且不依赖人工特性。

YOLO 的核心思想就是利用整张图作为网络的输入,直接在输出层回归 bounding box 的位置和 bounding box 所属的类别,其将每张图像分成 $S \times S$ 个网格,如果某个目标的中心落在这个网格中,则这个网格就负责预测这个目标。每个 bounding box 要预测中心点坐标(x, y)、框宽(w)、框高(h)和 confidence 共 5 个值,每个网格还要预测一个类别信息,记为 C 类。得到每个 box 的 confidence 分值后,滤掉得分低的 boxes,数据处理后得到最终检测结果。鉴于眼底图像

的某些微小病变难以与背景图像进行区分,本研究中将眼底图像的统一尺寸调整为 2496×2496 ,然后按 6×6 将图像均匀分为 36 张 416×416 的子图进行识别训练,避免尺寸较小的病变遗漏。同时为了分析模型预测结果与医生标注的一致性,本研究中将每张 45° 眼底图像划分成 $4.5^\circ \times 4.5^\circ$ 视野范围,共计 10×10 个小块,在每一小块上进行有无病变的比较,综合 100 个小块的结果,得到该眼底图像的检测分析结果(图 3)。



图 3 病变标注及模型预测图 A:标注图由研究团队标记产生 B:预测图以 bounding box 的形式产生

“嵩岳”机器人使用多层感知器作为分类器,本研究中综合考虑了病变与 DR 分级的相关性和病变正样本数量这 2 个因素,选定微血管瘤、视网膜内出血、硬性渗出、棉绒斑 4 种病变进行训练和预测。感知器输入层的神经元个数等于病变特征维度,即 4。输出层包含 2 个神经元,分别对应转诊和不转诊的概率值,中间层神经元个数根据经验设为 16(图 4)。

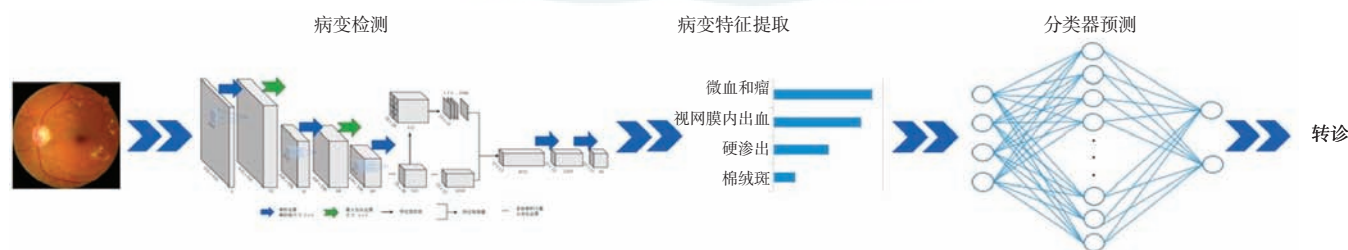


图 4 基于病变检测基础的筛查模型

1.3 统计学方法

采用 SPSS 22.0 统计学软件进行统计分析。应用受试者工作特征曲线(receiver operator characteristic curve, ROC 曲线)分析 AI 诊断性能,采用敏感度、特异度、阳性预测价值、阴性预测价值、曲线下面积(area under curve, AUC)等作为系统评价效能指标。

2 结果

2.1 “嵩岳”机器人辅助诊断系统各模型对 DR 的诊断性能

在 25 297 张 DR 眼底图像的标注中,无 DR 者 1 021 张,轻度 NPDR 者 7 712 张,中度 NPDR 者 13 116 张,重度 NPDR 者 2 347 张,PDR 者 1 101 张。“嵩岳”机器人辅助诊断系统中识别是否罹患 DR 模型的敏感度为 96.0%,特异度为 87.9%;识别是否需要转诊 DR 模型的敏感度为 90.4%,特异度为 95.2%;识别是否为重度 NPDR 和 PDR 模型的敏感度为 72.7%,特异度为 96.2%,识别是否为 PDR 模型的敏感度为 73.5%,特异度为 97.3%(表 1)。

2.2 “嵩岳”机器人 DR 辅助诊断系统各模型对 DR 诊断的 AUC

表 1 “嵩岳”机器人辅助诊断系统各模型对 DR 的诊断性能

模型	敏感度	特异度	阳性预测价值	阴性预测价值
是否罹患 DR	0.960	0.879	0.888	0.956
需转诊 DR	0.904	0.952	0.950	0.908
重度 NPDR 和 PDR	0.727	0.962	0.950	0.779
PDR	0.735	0.973	0.965	0.786

注: DR: 糖尿病视网膜病变; NPDR: 非增生性糖尿病视网膜病变; PDR: 增生性糖尿病视网膜病变

“嵩岳”系统中诊断是否罹患 DR 模型的 AUC 值为 0.920, 诊断是否需要转诊 DR 模型的 AUC 值为 0.925, 诊断是否为重度 NPDR 和 PDR 模型的 AUC 值为 0.845, 诊断是否为 PDR 模型的 AUC 值为 0.855 (表 2, 图 5~8)。

表 2 “嵩岳”机器人辅助诊断系统各模型对 DR 诊断的 AUC

模型	AUC	95% CI
是否罹患 DR	0.920	0.876~0.964
需转诊 DR	0.925	0.883~0.967
重度 NPDR 和 PDR	0.845	0.787~0.903
PDR	0.855	0.799~0.911

注: DR: 糖尿病视网膜病变; AUC: 曲线下面积; CI: 可信区间

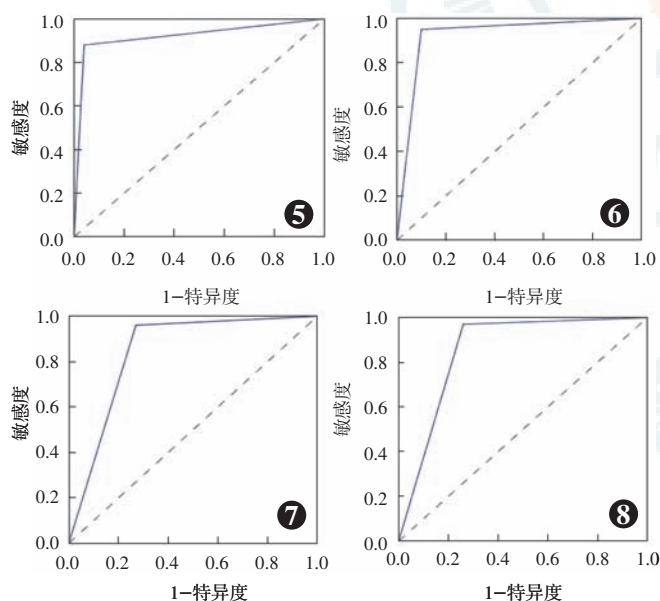


图 5 是否罹患 DR 模型 ROC 曲线 AUC 为 0.920 图 6 需转诊 DR 模型 ROC 曲线 AUC 为 0.925 图 7 重度 NPDR 和 PDR 模型 ROC 曲线 AUC 为 0.845 图 8 PDR 模型 ROC 曲线 AUC 为 0.855

3 讨论

DR 早期诊断和筛查的主要目的是为了检测出需要转诊到眼底病专科进行诊疗的患者, 便于早期发现、早期干预。英国一项研究显示, 2003—2016 年 DR 的

早期诊断计划使 DR 致盲率明显降低^[13], 因而开展 DR 诊断筛查服务, 避免致盲和劳动能力丧失对降低医疗压力和减少患者的负担具有重要的社会意义。然而, 鉴于 DR 检查的专业性和特殊性, 需要大量的人力和物力, 尤其中国糖尿病患者众多, 更增加其有效开展的难度^[14-17]。目前中国 87% 的糖尿病患者就诊于眼科医疗资源相对有限的县级及以下医疗机构, 70% 的糖尿病患者未接受规范的眼底检查, 中国 DR 的早期诊断和有效转诊面临巨大的挑战。

“嵩岳”机器人是在 25 297 张 DR 眼底图像、64 803 标注张次、100 余万次有效病变标注基础上基于深度卷积神经网络建立的国际上首个 DR AI 机器人。与以往“机器学习”AI 技术中受特定规则和图像条件限制的不同, 其本质是通过大型人造神经网络(互连节点层)将自己重新排列为新信息, 从而让计算机获得自学能力, 而无需人工编程, 这种 AI 技术可以通过一种 AI 算法来分析眼底图像并以此识别 DR 特征, 具有较强的适应不同图像和数据的能力。目前, 现有的 DR 模型多数仅限于 Referable/No Referable 1 分类, 无法进行不同分级和模式 DR 的识别^[10, 16-17], 而“嵩岳”机器人是在 28 项眼底具体病变检测的基础上, 通过建立不同 DR 识别模式, 从而实现 DR 诊断分级的目的。从评估方式上看, “嵩岳”机器人包括了是否罹患 DR、是否需转诊 DR、是否为重度 NPDR 和 PDR 和是否为 PDR 4 种模式, 几乎包含了所有具有临床价值的诊断模式。从结果上看, 4 种识别模式的 AUC 值分别为 0.920、0.925、0.845 和 0.855, 均具有很高的诊断准确性, 其中临床上较为关注的转诊 DR 模式的敏感度和特异度均达到了 90% 以上, 与 Google^[10]和新加坡多病种研究^[18] DR 转诊模型诊断结果相近, 显示了良好的决策性能。

DR 分类识别模式是“嵩岳”机器人智能诊断的特色, 使得 DR 筛查进一步具体化和全面化, 更加符合临床实际, 适用于不同阶段 DR 人群的 AI 诊断。通过建立 28 项具体眼底病变检测模型, 实现了一个完全依靠病变检测且具备较强解释能力的不同 DR 筛查模型, 与 DR 的自动分级互相呼应、双重确认, 为 AI 结果的可解释性和合理性提供了重要依据。在图像技术处理方面, 与通常采用的压缩像素的图片处理方法不同, “嵩岳”机器人采用图像放大和子图分割的方法, 一方面可以精确识别细微病灶, 可达 20×20 像素, 另一方面保证了病变区域的像素信息不会丢失。此外, 由于受图像来源及拍摄实际的影响, 我们对每张图像配备清晰图, 对有争议的标注, 提交资深审阅小组共同讨

论,进行严格的质量控制,保证标注质量。

基于眼底图像 AI 辅助诊断系统是 DR 诊治工作的发展趋势,构建眼底图像资源库,将非结构化图像数据转换成结构化数据,并通过 AI 进行数字化分析,将极大推动和改善 DR 诊治工作滞后的窘境,“嵩岳”机器人显示了良好的诊断预测价值。然而 DR 眼底病变不是孤立存在的,可能合并黄斑变性、水肿、动静脉阻塞以及视盘病变等,因而一个全面的 DR 模型需要包含除识别 DR 外的其他眼底病变,下一步我们会进一步建立 DR 伴随其他病变的诊断模型,增加深度学习的图片样本量,使 DR 的 AI 筛查更加完善精确。

利益冲突 孙宇辉、丁大勇供职于北京致远慧图科技有限公司

参考文献

- [1] 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.
- [2] Guariguata L, Whiting DR, Hambleton I, et al. Global estimates of diabetes prevalence for 2013 and projections for 2035 [J]. Diabetes Res Clin Pract, 2014, 103(2): 137-149. DOI: 10.1016/j.diabres.2013.11.002.
- [3] Ponto KA, Koenig J, Peto T, et al. Prevalence of diabetic retinopathy in screening-detected diabetes mellitus; results from the Gutenberg Health Study (GHS) [J]. Diabetologia, 2016, 59(9): 1913-1919. DOI: 10.1007/s00125-016-4013-5.
- [4] Safi H, Safi S, Hafezi-Moghadam A, et al. Early detection of diabetic retinopathy [J]. Surv Ophthalmol, 2018, 63(5): 601-608. DOI: 10.1016/j.survophthal.2018.04.003.
- [5] Mookiah MR, Acharya UR, Chua CK, et al. Computer-aided diagnosis of diabetic retinopathy: a review [J]. Comput Biol Med, 2013, 43(12): 2136-2155. DOI: 10.1016/j.combiomed.2013.10.007.
- [6] Caixinha M, Nunes S. Machine learning techniques in clinical vision sciences [J]. Curr Eye Res, 2017, 42(1): 1-15. DOI: 10.1080/02713683.2016.1175019.
- [7] Quer G, Muse ED, Nikzad N, et al. Augmenting diagnostic vision with AI [J/OL]. Lancet, 2017, 390(10091): 221 [2019-01-13]. https://www.thelancet.com/pdfs/journals/lancet/P II S0140-6736(17)31764-6.pdf. DOI: 10.1016/S0140-6736(17)31764-6.
- [8] Wang M, Shen LQ, Pasquale LR, et al. An artificial intelligence approach to detect visual field progression in glaucoma based on spatial pattern analysis [J]. Invest Ophthalmol Vis Sci, 2019, 60(1): 365-375. DOI: 10.1167/iov.18-25568.
- [9] Rajalakshmi R, Subashini R, Anjana RM, et al. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence [J]. Eye (Lond), 2018, 32(6): 1138-1144. DOI: 10.1038/s41433-018-0064-9.
- [10] 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.
- [11] 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.
- [12] Abràmoff MD, Folk JC, Han DP, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy [J]. JAMA Ophthalmol, 2013, 131(3): 351-357. DOI: 10.1001/jamaophthalmol.2013.1743.
- [13] Scanlon PH. The English National Screening Programme for diabetic retinopathy 2003-2016 [J]. Acta Diabetol, 2017, 54(6): 515-525. DOI: 10.1007/s00592-017-0974-1.
- [14] 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.
- [15] Zhong Y, Wu J, Yue S, et al. Burden of diabetic retinopathy in mainland China: protocol for an updated systematic review and meta-analysis of prevalence and risk factors to identify prevention policies [J/OL]. Medicine (Baltimore), 2018, 97(50): e13678 [2019-03-01]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6320130/. DOI: 10.1097/MD.00000000000013678.
- [16] Yang QH, Zhang Y, Zhang XM, et al. Prevalence of diabetic retinopathy, proliferative diabetic retinopathy and non-proliferative diabetic retinopathy in Asian T2DM patients: a systematic review and Meta-analysis [J]. Int J Ophthalmol, 2019, 12(2): 302-311. DOI: 10.18240/ijo.2019.02.19.
- [17] Hu JY, Yan L, Chen YD, et al. Population-based survey of prevalence, causes, and risk factors for blindness and visual impairment in an aging Chinese metropolitan population [J]. Int J Ophthalmol, 2017, 10(1): 140-147. DOI: 10.18240/ijo.2017.01.23.
- [18] Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning [J]. Invest Ophthalmol Vis Sci, 2016, 57(13): 5200-5206. DOI: 10.1167/iov.16-19964.
- [19] Sánchez CI, Niemeijer M, Dumitrescu AV, et al. Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data [J]. Invest Ophthalmol Vis Sci, 2011, 52(7): 4866-4871. DOI: 10.1167/iov.10-6633.
- [20] Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes [J]. JAMA, 2017, 318(22): 2211-2223. DOI: 10.1001/jama.2017.18152.

(收稿日期:2019-03-26 修回日期:2019-06-30)

(本文编辑:刘艳)

广告目次

止血祛瘀明目片 陕西摩美得气血和制药有限公司……封二

同息通(曲安奈德注射液) 广东省医药进出口公司珠海公司……前插页

露达舒(氯替泼诺混悬滴眼液) 博士伦(上海)贸易有限公司……前插页

沃丽汀(卵磷脂络合碘片) 广东泰恩康医药股份有限公司……前插页

海德堡超清 OCTA+X 高视医疗……前插页

丽爱思(地夸磷索钠滴眼液) 参天制药(中国)有限公司……封三

迈达科技 天津迈达科技股份有限公司……封底