

· 临床研究 ·

# 深度学习中标准化人工标注在早产儿视网膜病变图像识别中的应用

汪信<sup>1</sup> 张贵华<sup>1</sup> 林建伟<sup>1</sup> 吉杰<sup>2</sup> 邱坤良<sup>1</sup> 张铭志<sup>1</sup><sup>1</sup>汕头大学·香港中文大学联合汕头国际眼科中心, 汕头 515041; <sup>2</sup>汕头大学网络与信息中心 515063

通信作者: 张铭志, Email: zms@jsiec.org

**【摘要】** 目的 评估深度学习中标准化人工标注在早产儿视网膜病变(ROP)图像识别中的应用价值。

**方法** 根据 ROP 国际分类标准,将 ROP 各期病变分为分期病变和附加病变。标注图片来源于汕头国际眼科中心,采用分层抽样法随机提取 2009 年 8 月至 2018 年 7 月 ROP 各期病变 1 464 帧图片,分为 3 个组:分期病变组(亚组 1:173 帧,亚组 2:117 帧)和附加病变组(亚组 1:163 帧,亚组 2:116 帧)为培训标注用图片;一致性标注图组(895 帧)为验证培训效果的图片组。标注人员包括 2 名资深专家、3 名高年资医生和 3 名实习医生。全体标注人员进行统一标注的图片标注培训。以专家组标注结果为参照,标注一致率 $\geq 90\%$ 作为合格标准。分组比较分析标注结果并与深度学习方法的结果进行比较。分析非专家医师组以及深度学习方法与专家标注结果比较的一致性。**结果** 首次培训后,高年资组和实习医师组整体一致率在 2 种病变标注中均低于 90%。经过 2~3 次标准化培训后,高年资组医师和实习医师组医师对分期病变标注的整体一致率分别为 98.99% (Kappa=0.979)和 99.22% (Kappa=0.984),对附加病变标注的整体一致率分别为 97.43% (Kappa=0.914)和 98.11% (Kappa=0.935),一致性好;人机结合基础上的深度学习对分期病变标注的整体一致率为 94.08% (Kappa=0.880),一致性较好。**结论** 标准化人工标注可提升深度学习对 ROP 图像识别的智能化效率,并能作为眼科医师同质化、规范化阅片培训的创新方法。

**【关键词】** 人工智能; 标准化标注; 深度学习; 新生儿视网膜病变; 分期病变; 附加病变**基金项目:** 广东省科技创新战略专项资金项目(汕府科[2018]157-46)

DOI:10.3760/cma.j.issn.2095-0160.2019.08.013

## Application of standardized manual labeling on identification of retinopathy of prematurity images in deep learning

Wang Ji<sup>1</sup>, Zhang Guihua<sup>1</sup>, Lin Jianwei<sup>1</sup>, Ji Jie<sup>2</sup>, Qiu Kunliang<sup>1</sup>, Zhang Mingzhi<sup>1</sup><sup>1</sup>Joint Shantou International Eye Center, Shantou University & the Chinese University of Hong Kong, Shantou 515041, China; <sup>2</sup>Network & Information Center, Shantou University, Shantou 515063, China

Corresponding author: Zhang Mingzhi, Email: zms@jsiec.org

**[Abstract]** **Objective** To evaluate the application of the standard manual labeling on identification of retinopathy of prematurity (ROP) images in deep learning. **Methods** According to the International Classification of ROP, different periods of ROP were classified into stage disease and plus disease in this study. From Joint Shantou International Eye Center from August 2009 to July 2018, a total of 1 464 labeled fundus retinal photographs were divided randomly by stratified sampling into 3 groups: stage disease group (subgroup 1:173, subgroup 2:117) was used to train for labeling stage disease, whereas plus disease group (subgroup 1:163, subgroup 2:116) was used to train for labeling plus disease, and consistent labels group consisted of 895 consistent labeled images on both disease. Graders consisted of senior experts, 3 senior ophthalmologists and 2 interns, and received training for classification and labeling on ROP fundus images. The results were compared among the doctors and doctors with deep learning, and the agreement between non-experts doctors and the reference standards, and deep learning and the reference standards were tested. **Results** After the first training, the overall agreement rate of the senior ophthalmologist group and the intern group were lower than 90% for both two disease labeling. After two to three times of training, in image of

consistent labels group, overall agreement rates of senior ophthalmologists and intern doctor's were 98.99% (Kappa = 0.979), 99.22% (Kappa = 0.984) on stage disease, and 97.43% (Kappa = 0.914), 98.11% (Kappa = 0.935) on plus disease, respectively. The agreement on stage disease using deep learning based on human-machine combination was 94.08%, Kappa value was 0.880, which achieved good degree. **Conclusions** Standardized manual labeling can improve the intelligentization of deep learning on identification of ROP images, and be considered as an innovative method of homogenization and standardized training for doctors in ophthalmology.

**[Key words]** Artificial intelligence; Standardized label; Deep learning; Retinopathy of prematurity; Stage disease; Plus disease

**Fund Program:** Science and Technology Innovation Strategy Special Fund Project of Guangdong Province ([2018]157-46)

DOI:10.3760/cma.j.issn.2095-0160.2019.08.013

深度学习是人工智能 (artificial intelligence, AI) 领域的主要技术, 由于其对图像识别率高的特点而被应用于医学影像诊断, 目前已在眼科、放射科、皮肤科及病理学等方面广泛应用, 为临床诊断及远程医疗提供了新的诊疗模式<sup>[1-5]</sup>。深度学习不仅需要海量的人工标注的样本数据, 同时也要求训练集与外部应用集的高度一致, 这使得深度学习算法在医学领域的应用受到限制。深度学习模型的预测效果高度依赖训练样本的标注质量<sup>[6]</sup>。合格的样本既要符合内部既定标准, 又要能接受外部验证, 缺一不可, 样本标注的质量是保证 AI 一致性的关键。英国的糖尿病视网膜病变筛查项目中已将样本标注质量保证作为常规执行要求, 并建立了标准培训模式<sup>[7]</sup>, 但在眼科各类疾病应用标准化人工标注培训的结果与深度学习的比较研究未见报道。早产儿视网膜病变 (retinopathy of prematurity, ROP) 是世界上主要的儿童致盲眼病, 已列入新生儿常规筛查项目, 筛查所得的眼底照片中病变类型多样且变异大、图片质量不一, 影响阅片者内部或之间的诊断一致性, 因此, 有必要建立标准化人工标注培训方法。本研究中以 ROP 建立人机结合的深度学习方法, 探讨 AI 在 ROP 的临床监控和治疗中的应用价值, 为建立标准化的动态 AI 评价系统提供依据。

## 1 资料与方法

### 1.1 资料

**1.1.1 标注资料** 自 2009 年 8 月至 2018 年 7 月于汕头国际眼科中心进行 ROP 筛查并经 RetCam II 拍照的患儿眼底图中采用分层抽样法随机选取 1 464 张图片, 分为 3 个组: 分期病变组 (亚组 1, 173 张; 亚组 2, 117 张) 和附加病变组 (亚组 1, 163 张; 亚组 2, 116 张) 为培训用标注图片, 分别用于练习分期病变和附加病变的标注; 一致性标注图组 (895 张) 用于 2 种病变标注, 以验证培训效果。图像入选标准: (1) 图像中眼底

不能辨识部分少于 1/4 面积; (2) 病历资料完整者。图像排除标准: (1) 图像中超过 1/4 面积的眼底不能辨识者; (2) 已行 ROP 治疗者; (3) 病历资料缺失者。

**1.1.2 标注人员** 标注人员包括 2 名资深专家组 (ROP 筛查专家)、3 名高年资医师 (眼科从业 8 年以上) 和 3 名实习医师 (临床实习 1 年)。所有医师均接受标准化、规范化培训。

### 1.2 方法

**1.2.1 标注标准** 依据 ROP 国际分类标准<sup>[8]</sup>, 病变按病情严重程度分为 5 期: 1 期表现为分界线, 是眼底周边视网膜中位于血管区与无血管区之间的细小、低平但界限清楚的变化; 2 期表现为峭, 由分界线变高, 增宽逐渐发展而来; 3 期表现为嵴上纤维增生, 有嵴上新生血管形成伴纤维性增生组织; 4 期表现为不完全性视网膜脱离, 有颞侧纤维组织增生, 黄斑及血管弓受牵引。如脱离时未累及黄斑时为 4A 期, 否则为 4B 期; 5 期为完全性视网膜脱离。附加病变指后极部视网膜血管扩张迂曲, 至少达 2 个象限, 程度达到或严重于标准图片。前附加病变则指眼底后极部视网膜血管迂曲扩张, 但程度未达到附加病变; 或程度达到附加病变但只累及 1 个象限。本研究中将上述病变分为分期病变及附加病变两类进行标注: 分期病变类定义为视网膜上 1~5 期中的任意一期病变; 附加病变类定义为附加病变与前附加病变, 此外, 因 ROP 导致的任何部位的视网膜前出血及确定的玻璃体出血也列入附加病变中。当图片中显示分期病变时, 标注为 S1, 否则标注为 S0; 当病变为附加病变与前附加病变时, 分别标注为 P2 和 P1, 两者均无时, 标注为 P0。标注为 S1、P2 和 P1 者, 后续建议转诊, 转至眼底专科医师进一步诊治, 标为 S0 与 P0 者, 建议随访。

**1.2.2 培训流程** (1) 培训: 学习疾病分期与标注标准。(2) 标注训练: 2 名资深专家对各组图片进行预标注。定标一致的结果作为参考标准, 其图片用于另外

2 个组医师进行标注训练。各医师依据标注标准与参考图片独立标注。(3)质量监控:分析标注结果,查找差异原因并及时纠正。(4)再培训及标注:组内所有医师接受再培训及标注。但反复培训后,组内医师中有标注一致率不升反降,或个人一致率最低并影响整体一致率达标的医师需接受再培训。

**1.2.3 深度学习标注** 结合有专家标注的一致性标注图组图片,利用深度学习算法对分期病变进行识别训练,并输出自动学习的结果,内容包括提示病变位置的热力图 and 自动分类标注,后者以专家标注的结果为参考标准,计算一致率及 Kappa 值。

**1.2.4 评价指标** (1)一致率:一致率=(正确标出有病变的图像数+正确标出无病变的图像数)/所在图集数×100%。一致率分为个人一致率和整体一致率。整体一致率为组内 3 人标注结果均与参考结果相符的一致率,整体一致率≥90% 视为合格。(2)Kappa 值:Kappa = (Po-Pe)/(1-Pe),其中 Po 为实际一致率,Pe 为理论一致率。为非专家医师组整体标注结果与参考结果进行比较。Kappa 值分为 3 个评价等级:0~0.4:一致性较差;0.41~0.75:一致性中等;0.76~1.0:一致性好。

**1.3 统计学方法**

采用 SPSS 20.0 统计学软件进行统计分析。计数资料以例数和百分数表示,采用 McNemar 检验进行一致率的比较,P<0.05 为差异有统计学意义。采用 Kappa 检验进行一致性检验。

**2 结果**

**2.1 专家预标注**

对分期病变组标注,专家在亚组 1 和 2 中的一致性分别为 94.54% 和 97.50%;对附加病变组标注,专家在亚组 1 和 2 中的一致性分别为 89.07% 和 96.67% (表 1)。

**表 1 专家在培训图片组中标注一致性**

图片分组	亚组编号	原始图片数(n)	标注一致性[(n,%)]
分期病变组	1	183	173(94.54)
	2	120	117(97.50)
附加病变组	1	183	163(89.07)
	2	120	116(96.67)
合计		606	569

**2.2 分期病变组标注**

首次培训后,对分期病变亚组 1 图片标注,高年资组和实习医师组内个人一致率最高分别为 91.91% 和 64.74%,最低分别为 87.28% 和 59.54%;整体一致率

分别为 79.19% 和 48.55%,均明显低于 90% 的合格标准。第 2 次培训后对分期病变亚组 2 图片标注,高年资组与实习医师组内个人一致率最高分别达 98.29% 与 94.87%,最低分别为 80.34% 和 91.45%;整体一致率分别为 79.49% 和 81.20%。将 2 个组医师中一致率最低的高年资医师 3 与实习医师 3 再次培训,并重新标注分期病变亚组 2 图片,个人一致率分别提高至 98.29% 和 100.0%,高年资医师与实习医师组整体一致率分别提高至 94.87% 和 89.74%,高年资组和实习医师组医师对分期病变标注的整体一致率与专家组比较,差异均无统计学意义(P=0.687、0.388),对应 Kappa 值分别为 0.897 和 0.795,一致性好(表 2)。

**表 2 不同年资医师对分期病变组图片标注的一致率比较[(n,%)]**

医师年资	分期病变亚组 1 (N=173)		分期病变亚组 2 (N=117)	
	第 1 次 培训后标注	第 2 次 培训后标注	第 2 次 培训后标注	第 3 次 培训后标注
高年资医师 1	151(87.28)	115(98.29)	115(98.29)	115(98.29)
高年资医师 2	159(91.91)	113(96.58)	113(96.58)	113(96.58)
高年资医师 3	159(91.91)	94(80.34)	115(98.29)	115(98.29)
整体	137(79.19)	93(79.49)	111(94.87)	111(94.87)
实习医师 1	10(59.54)	111(94.87)	111(94.87)	111(94.87)
实习医师 2	11(64.74)	107(91.45)	107(91.45)	107(91.45)
实习医师 3	11(63.58)	107(91.45)	117(100.0)	117(100.0)
整体	84(48.55)	95(81.20)	105(89.74)	105(89.74)

**2.3 附加病变组标注**

首次培训后,对附加病变亚组 1 图片进行标注,高年资组和实习医师组内个人一致率最高分别为 80.37% 和 65.64%,整体一致率分别为 72.39% 和 39.88%,均低于 90% 的合格标准。第 2 次培训后对附加病变亚组 2 图片标注,高年资组与实习医师组内个人一致率最高均为 97.41%,最低分别为 68.10% 和 78.45%,整体一致率分别为 66.38% 和 74.14%,均低于 90% 的合格标准。将 2 个组医师中一致率最低的高年资医师 3 和实习医师 2 再次培训,并重新标注图集 2,个人一致率分别提高至 95.69% 和 99.14%,对应的整体一致率分别提高至 90.52% 和 96.55%,高年资组和实习医师组医师对附加病变标注的整体一致率与专家组比,差异均无统计学意义(P=0.053、0.261)。高年资组 Kappa 值为 0.627,一致性中等;实习医师组为 0.86,一致性好(表 3)。

**2.4 一致性标注图片组标注**

高年资组医师和实习医师组医师分期病变标注的整体一致率分别为 98.99% (Kappa = 0.979) 和

99.22% (Kappa=0.984), 对附加病变标注的整体一致率分别为 97.43% (Kappa=0.914) 和 98.11% (Kappa=0.935), 一致性好(表 4)。

**表 3 不同年资医师对附加病变组图片标注的一致率比较 [n(%)]**

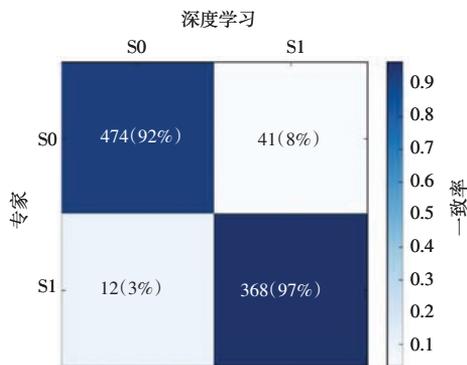
医师年资	附加病变亚组 1 (N=163)		附加病变亚组 2 (N=116)	
	第 1 次 培训后标注	第 2 次 培训后标注	第 3 次 培训后标注	第 3 次 培训后标注
	高年资医师 1	130(79.75)	110(94.82)	110(94.82)
高年资医师 2	131(80.37)	113(97.41)	113(97.41)	
高年资医师 3	124(76.07)	79(68.10)	111(95.69)	
整体	118(72.39)	77(66.38)	105(90.52)	
实习医师 1	107(65.64)	113(97.41)	113(97.41)	
实习医师 2	97(59.51)	91(78.45)	115(99.14)	
实习医师 3	91(55.83)	112(96.55)	112(96.55)	
整体	65(39.88)	86(74.14)	112(96.55)	

**表 4 不同年资医师对标注一致组图片标注的一致率比较 [n(%)]**

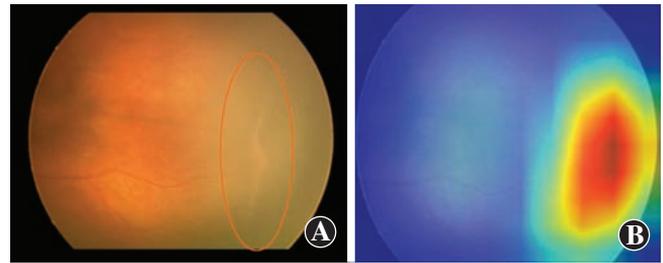
医师年资	分期病变 (N=895)	附加病变 (N=895)
高年资医师 1	887(99.10)	879(98.21)
高年资医师 2	888(99.22)	881(98.44)
高年资医师 3	894(99.88)	877(97.99)
整体	886(98.99)	872(97.43)
实习医师 1	894(99.88)	878(98.10)
实习医师 2	893(99.78)	881(98.44)
实习医师 3	890(99.44)	878(98.11)
整体	888(99.22)	878(98.11)

**2.5 深度学习标注**

利用结合人工标注的 895 帧图片进行深度学习训练后,深度学习模型对分期病变自动分类的结果与专家标注结果见图 1,结果显示深度学习标注的一致率为 94.08%,Kappa 值 0.880,一致性好。热力图中标示出了深度学习自动识别出的病变位置(图 2)。



**图 1 深度学习与专家在标注一致性组中对分期病变标注一致性比较 [例数(一致率)]** 当图片中显示分期病变时,标注为 S1,否则标注为 S0



**图 2 原始眼底图与对应的热力图在病灶识别中的比较** A:原始眼底图 红圈区域显示视网膜上呈条形的分期病变 B:热力图 图中红色区域为 AI 识别的病灶位置

**3 讨论**

本研究中结果显示,3 个组医师间的标注结果存在差异。以专家组标注结果为标准,高年资医师组的个人及整体一致率在第 1 次培训后均明显高于实习医师组。经过多次标准化人工标注培训后,高年资医师组和实习医师组的对图片标注的整体一致率均明显提高,并在此过程中强调“标准化”和“重复性”,可使不同年资医师的标注能力提高且一致性符合预定标准。“标准化”包括疾病的分类标准化和人员培训流程标准化,有利于不同标注人员对疾病建立共识,减少主观偏倚。“重复性”则强调培训与标注练习反复循环进行,有利于知识在重复提取过程中增强医师们的理解记忆能力,并从理论顺利迁移至实践,提升学习效率<sup>[9]</sup>。

本研究中第 2 次培训后,部分医师标注一致率不升反降(如高年资医师 3),或是标注一致性偏低,影响整体一致性提高(如实习医师 2 和 3),分析原因可能是:(1)由于 ROP 漏诊可能造成患儿视力损害等严重后果,高年资医师出于漏诊的风险防控的心理,会在决策时对于非典型或临界的眼底改变倾向于认为有病变,即使该改变实际属于无病变。这种经验决策会增加假阳性以降低漏诊率,同时降低了标注一致率。(2)实习医师因为经验缺乏、对疾病的相关知识结构尚未完全建立,再加上标注时的专注力不够等原因,不能正确区分病变,造成一致率偏低,甚至下降。(3)不同图片组中的同类型病变存在变异,或有图片质量差异<sup>[10]</sup>,影响判断。因此,不同医师的内部知识构成、标注时的心理状态与精力、时间推移带来的记忆力和理解力改变等主观差异以及病变变异等客观因素或多或少影响标注结果,需要通过建立标准化、规范化流程来减少差异,由此提示标准化标注培训的重要性。然而一次性培训不能确保标注一致性明显提高或稳定,除需反复进行培训外,还必须借助质量监控环节<sup>[7,11]</sup>。

质量监控要求在标注后及时反馈结果并分析,找出与参考标注差异的真实原因,再进行有针对性的培训。该环节能帮助标注人员提高标注水平,减少错误发生。在一致性标注图片组图片中,高年资医师组和实习医师组医师的标注一致率显著提高,Kappa 值为 0.914 ~ 0.984,接近专家的结果,证实了良好的培训效果。因此,保证标注人员标注一致性提高需要进行标准化标注培训,并且需要将质量监控设置为必要环节。

深度学习算法具有自动特征提取功能,可基于有标注的样本自行学习并抓取图片中特征,最后输出自行分类结果<sup>[12]</sup>。保证深度学习准确度高的前提之一是样本在人工标注阶段被尽可能正确地标注分类。高质量样本标注可使深度学习算法自动特征提取的优势得到充分发挥,对病变识别与分类的准确度更高,与人工标注一致性更好。低质量样本中由于不同病变相互混杂且变异大,使得深度学习算法在学习后难以有效提取特征,分类预测的准确性大大降低。本研究中,深度学习算法在学习了附带专家标注的一致性标注图片组图片后,一致率在首次训练时即达到 94.08%。因此,深度学习算法结合自身的特征提取功能与人工高质量标注后,能得到理想的学习效果,体现了人机结合的优势。人机标注一致的结果在一定程度上可以印证人工标注的准确性,而不一致的结果能督促标注者进行核查。同时,热力图对病变范围的显示使深度学习模型的效果更直观,更利于分析人机间结果差异的原因。通过人工分析比较-查找解决问题-再培训-再训练的人机结合质量控制环节,能促进二者在病变认识上的共同提高。因此,在标准化标注培训的基础上,人机结合能相互促进并增强对临床知识的提取及学习能力。我们将进一步利用深度学习算法对附加病变的分类进行训练。

综上所述,标准化人工标注培训能为 AI 深度学习提供高质量的人工标注数据,并借助数据在人机间的传输与结果反馈,促进两者对知识的共同提取,尤其是深度学习的智能化。眼科中用于检查诊断的影像图片种类繁多、数量巨大、病变变异大,致使年轻医师阅片标注的难度提升,学习曲线长。标准化人工标注培训有利于不同资质医师,尤其是年轻医师快速提取临床知识并掌握标注技能。此外,眼科临床中还存在大量病历信息资源,同样需要医师进行充分学习、整合分析及有效提取,但面对大数据的处理,人工显然不及 AI。相信在不久的将来,随着 AI 学习的不断深入,深度学习还能被用于各种宏观或微观的临床数据总结与挖掘,并对医师们进行反向输出与培训,帮助他们在临床

知识与技能方面的提高,达到同质化。因此,标准化人工标注可作为同质化且规范培训的创新方法。

**利益冲突** 所有作者均声明不存在利益冲突

**志谢** 感谢黄定国、公为芬、李太平医师,苏莹莹、陈铭雄、邝继发同学参与课题中的培训及标注,感谢黄艳虎同志给予统计学方法指导

## 参考文献

- [1] 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.
- [2] Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks [J]. *JAMA Ophthalmol*, 2018, 136 (7): 803-810. DOI: 10.1001/jamaophthalmol.2018.1934.
- [3] Zech JR, Badgeley MA, Liu M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study [J/OL]. *PLoS Med*, 2018, 15 (11): e1002683 [2019-02-10]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6219764/>. DOI: 10.1371/journal.pmed.1002683.
- [4] Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network [J/OL]. *PLoS One*, 2018, 13 (1): e0191493 [2019-01-12]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5774804/>. DOI: 10.1371/journal.pone.0191493.
- [5] Gecer B, Aksoy S, Mercan E, et al. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks [J]. *Pattern Recognit*, 2018, 84: 345-356. DOI: 10.1016/j.patcog.2018.07.022.
- [6] 林浩添, 吴晓航. 加快基于眼科图像数据库的眼病人工智能辅助诊断平台建设 [J]. *中华实验眼科杂志*, 2018, 36 (8): 577-580. DOI: 10.3760/cma.j.issn.2095-0160.2018.08.001.
- [7] Lin HT, Wu XH. Accelerating the construction of artificial intelligence diagnostic platform based on ophthalmic imaging database [J]. *Chin J Exp Ophthalmol*, 2018, 36 (8): 577-580. DOI: 10.3760/cma.j.issn.2095-0160.2018.08.001.
- [8] 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.
- [9] International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited [J]. *Arch Ophthalmol*, 2005, 123 (7): 991-999. DOI: 10.1001/archophth.123.7.991.
- [10] Karpicke JD, Blunt JR. Retrieval practice produces more learning than elaborative studying with concept mapping [J]. *Science*, 2011, 331 (6018): 772-775. DOI: 10.1126/science.1199327.
- [11] Ludwig CA, Greven MA, Moshfeghi DM. Predictors of treatment-warranted retinopathy of prematurity in the SUNDROP cohort: influence of photographic features [J]. *Graefes Arch Clin Exp Ophthalmol*, 2017, 255 (10): 1935-1946. DOI: 10.1007/s00417-017-3745-3.
- [12] Williams SL, Wang L, Kane SA, et al. Telemedical diagnosis of retinopathy of prematurity: accuracy of expert versus non-expert graders [J]. *Br J Ophthalmol*, 2010, 94 (3): 351-356. DOI: 10.1136/bjo.2009.166348.
- [13] Maier A, Syben C, Lasser T, et al. A gentle introduction to deep learning in medical image processing [J]. *Z Med Phys*, 2019, 29 (2): 86-101. DOI: 10.1016/j.zemedi.2018.12.003.

(收稿日期:2019-02-22 修回日期:2019-07-01)

(本文编辑:刘艳)