



深度学习算法在角膜荧光染色分级评估中的应用*

赵雨暄¹, 张啸云², 杨必^{1,3,4△}, 刘陇黔^{1,3,4}

1. 四川大学华西临床医学院 眼视光学系(成都 610041); 2. 四川大学计算机学院 计算机科学与技术系(成都 610065);

3. 四川大学华西医院 眼视光学与视觉科学研究室(成都 610041); 4. 四川大学华西医院 眼科(成都 610041)

【摘要】 目的 探讨深度学习算法在角膜荧光染色分级评估中的应用价值。方法 采用横断面研究,共纳入四川大学华西医院隐形眼镜门诊就诊患者的角膜染色图片600张。其中500张用于构建算法,其余100张用于验证算法有效性并比较人工智能(artificial intelligence, AI)与视光初学者间评级准确性(accuracy, ACC)及诊断时长的差异。在完成初次评级1个月后,进行第二次评估,比较两次评估的一致性指数($kappa$ 值)。3位经验丰富的视光师综合分级的结果作为本研究的“金标准”。结果 通过全集、训练集、测试集交叉验证4种深度学习模型,ResNet34模型预测准确性最高。ResNet34深度学习模型在角膜染色分级上的准确性为93.0%,敏感性为89.5%,特异性为89.6%。AI与两位初学者比较,AI的准确性较高($ACC_{AI}=87.0\%$ 、 $ACC_{S1}=78.0\%$ 、 $ACC_{S2}=52.0\%$, $P_{ACC}=0.001$),同时AI的平均诊断时长短于视光初学者($t_{AI}=1.00$ s, $t_{S1}=11.86$ s, $t_{S2}=13.25$ s, $P_t=0.001$)。在两次评级的一致性比较中,AI($kappa_{AI}=0.658$, $P_{AI}=0.001$)的一致性程度高于视光初学者($kappa_{S1}=0.575$, $P_{S1}=0.001$; $kappa_{S2}=0.609$, $P_{S2}=0.001$)。结论 将深度学习算法应用于角膜染色分级评估具有一定的可行性及临床价值。在与初学者的比较中,AI表现出了较高的准确性和较好的一致性,能够指导视光师在临床工作中快速准确地评定角膜染色分级。

【关键词】 人工智能 接触镜 角膜染色

Application of Deep Learning Algorithm in the Grading Assessment of Corneal Fluorescein Staining ZHAO Yuxuan¹, ZHANG Xiaoyun², YANG Bi^{1,3,4△}, LIU Longqian^{1,3,4}. 1. Department of Optometry and Vision Sciences, West China School of Medicine, Sichuan University, Chengdu 610041, China; 2. College of Computer Science, Sichuan University, Chengdu 610065, China; 3. Laboratory of Optometry and Vision Sciences, West China Hospital, Sichuan University, Chengdu 610041, China; 4. Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu 610041, China

△ Corresponding author, E-mail: yangbi19830418@126.com

【Abstract】 Objective To explore the application value of applying deep learning (DL) algorithm in the grading assessment of corneal fluorescein staining. **Methods** A cross-sectional study was carried out, covering 600 corneal fluorescein staining photos acquired in the Contact Lens Clinic, West China Hospital, Sichuan University between 2020 and 2022. Out of the 600 photos, 500 were used to construct the algorithm and the remaining 100 were used for the validation of the algorithm and a comparative analysis of the difference in grading accuracy (ACC) and the length of diagnostic time between artificial intelligence (AI) and optometry students. One month after finishing the first grading analysis, assessment by AI and optometry students was conducted for a second time and results from the two rounds of assessment were compared to examine the intrarater agreement ($kappa$ value) of the two analyses. The grading analysis results of 3 experienced optometrists were used as the gold standard in the study. **Results** Findings of the cross validation with the complete dataset, the training dataset, and the test dataset showed that ResNet34 had the highest predictive accuracy among four DL models. ResNet34 DL model achieved an accuracy of 93.0%, sensitivity of 89.5%, and specificity of 89.6% in the grading of corneal staining. In the comparison of the grading accuracy of AI and two optometry students, AI showed better accuracy, with the respective grading accuracy being 87.0%, 78.0%, and 52.0% for AI, student 1, and student 2 ($P_{ACC}=0.001$). In addition, the average diagnostic time of AI was shorter than that of optometry students ($t_{AI}=1.00$ s, $t_{S1}=11.86$ s, $t_{S2}=13.25$ s, $P_t=0.001$). In the comparative analysis of the intrarater agreement between the two assessments, AI ($kappa_{AI}=0.658$, $P_{AI}=0.001$) achieved better consistency than the two optometry students did ($kappa_{S1}=0.575$, $P_{S1}=0.001$; $kappa_{S2}=0.609$, $P_{S2}=0.001$). **Conclusion** Applying deep learning algorithms in the grading assessment of corneal fluorescein staining has considerable feasibility and clinical value. In the performance comparison between AI and optometry students, AI achieved higher accuracy and better consistency, which indicates that AI has potential application value for assisting optometrists to make clinical decisions with speed and accuracy.

【Key words】 Artificial intelligence Contact lens Corneal staining

* 四川省科技计划项目(No. 2022YFS0368)资助

△ 通信作者, E-mail: yangbi19830418@126.com

我国近视患者数量逐年增加且发病年龄不断降低。近视已成为最常见的眼病之一,也对我国医疗卫生领域

产生了重大负担^[1]。随着近视度数的不断加深,眼轴长度不断增长,最终可能导致一系列严重的并发症^[2-3]。

接触镜是常用的近视矫正方法之一。虽然接触镜是一项安全的措施,但验配不当、摘戴和护理操作不当都可能会导致角膜损伤。观察和诊断角膜损伤最方便、实用和明确的手段是角膜荧光染色^[4]。一项针对配戴接触镜受试者的研究显示,54%的患者在配戴接触镜后出现了角膜染色^[5]。分级量表是记录和监测眼部组织临床变化的重要工具,其中Efron分级量表是常用于描述接触镜并发症的分级量表之一^[6]。

人工智能(artificial intelligence, AI)是计算机科学领域的一个重要分支,其通过模拟人类的学习行为,获取新的知识或技能^[7]。随着大数据科学技术的发展,AI已在角膜相关疾病的诊断中展现出巨大的应用前景,AI已成功应用于干眼的诊断、圆锥角膜早期的筛查和分级诊疗,算法均达到了较好的准确性、灵敏性和特异性,有望在临床中为医生提供客观的临床决策^[8-9]。

目前尚无研究报道AI在角膜荧光染色分级评估中的应用,因此本研究计划利用AI深度学习角膜染色,并在大量样本中进行验证,以保证其在临床分级评估中的准确性。

1 资料与方法

1.1 研究对象

本研究为横断面研究,纳入2020–2022年在四川大学华西医院隐形眼镜门诊按照标准流程验配接触镜的儿童青少年患者的角膜染色图片600张。角膜染色图片的纳入标准:8~20岁;无眼部器质性病变;无斜弱视;无眼部手术史;角膜染色图片涵盖整个角膜。排除标准:角膜染色图像像素点丢失或像素小于1920×1080。本研究通过四川大学华西医院生物医学伦理委员会批准(2022年审1592号)。

1.2 图像采集方法

在相同的标准流程下获取600张不同等级的角膜染色图片。用1~2滴无菌生理盐水(法适宝生理盐水,法国),将荧光素钠试纸(天津晶明新技术开发有限公司,中国)部分沾湿,将沾湿部分轻触患者睑结膜。在暗室内,使用钴蓝光滤光片在裂隙灯显微镜(TOPCON SL-1E,日本)的弥散光照明下观察角膜。使用佳能相机(CANON,日本)拍摄角膜染色图片,所有的角膜染色图片均使用同一台设备在相同的室内照明下拍摄。所有拍摄的图像都由3名经验丰富的视光师对其进行审核,确保图像具有临床意义。

1.3 评估方法

采用单纯随机抽样的方法抽取500张临床患者角膜染色图片,基于已发表并广泛应用的角膜染色分级方案(Efron分级标准)^[10-11],由3名经验丰富的视光师对其进行角膜染色等级的评估,确定角膜染色的分级,并以此作为本研究的“金标准”。结合视光师分析的结果,通过AI深度学习,构建算法。

1.4 算法设计

深度学习是一种基于多层神经网络的学习方法^[12]。卷积神经网络(convolutional neural network, CNN)是最常用的深度学习图像识别算法^[13]。CNN模型可以对角膜染色图像进行识别训练,在足够多样本的支持下,可以根据以往识别的图像,对新的角膜染色图像等级进行预测。

本研究中用于构建模型的服务器由四川大学机器智能实验室提供。从原始600张角膜染色图片中采取随机抽样的方式抽取其中的500张角膜染色图片用于构建算法。同时为提高学习效果,训练时对数据集中74%的图像进行水平镜像翻转,获得的增强数据集为原始图像的1.74倍,使模型在保证训练量的同时提高了鲁棒性和泛用性。最终使用870张角膜染色图片作为训练集和验证集用于构建及优化模型,其余100张图片作为测试集用于评估模型性能并对AI与视光初学者评级结果的准确性及诊断时长进行比较,具体流程如图1所示。

1.4.1 模型训练

在深度学习中,传统的CNN会随着网络层数的加深出现梯度消失或网络退化等问题^[14]。本研究使用的ResNet模型通过引入残差模块建立前层与后层之间的“短路连接”,整个网络只需要学习输入与输出差别的部分,在前向计算和反向传播过程中保留更多信息,从而适应深层CNN模型,大大提高了模型效果^[14]。DenseNet模型采用“密集连接”的机制,同时DenseNet模型的每个层都与前层通过特征在通道上的连接实现特征重用,这使得DenseNet在减少参数数量的同时实现更优的性能^[15]。

本研究使用了ResNet34、ResNet50、DenseNet121、DenseNet169四种深度学习算法对角膜染色图像进行训练。参考既往文献^[8],并依据经验和数据集的特征对部分参数进行了调整:将初始学习率设置为0.01,正则化权重(regularization weight)为0.0001,动量(momentum)为0.9。图像级别的分类阈值设定为0.50,模型在每次迭代中训练20个周期。采用模型训练准确性、验证准确性、测试准确性评估训练效果。

1.4.2 模型验证

在完成模型的训练后,通过验证集对模型调整超参数

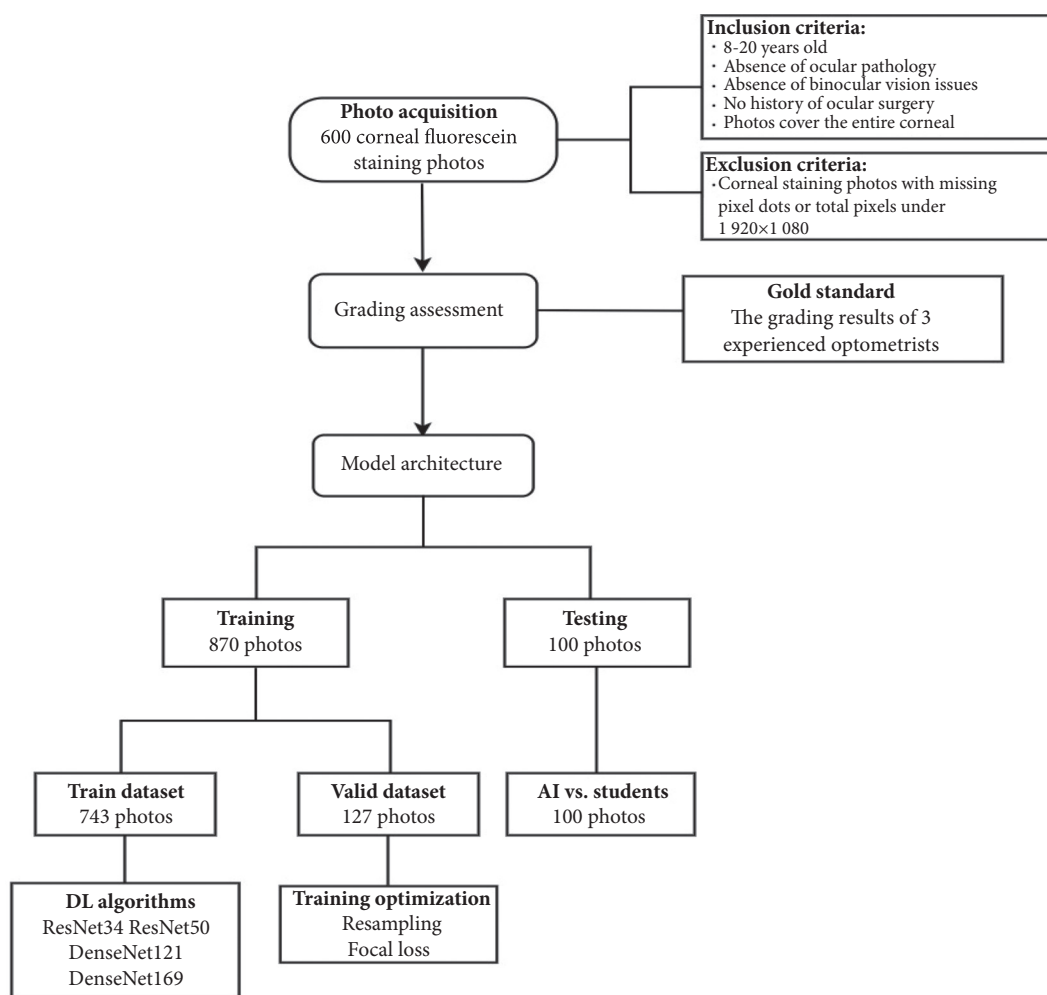


图 1 研究方法流程图

Fig 1 Flowchart of the research method

进行训练优化。在本研究中,主要使用了重采样(resampling)和Focal Loss的方法对训练优化。

本研究用于深度学习的角膜染色图片是典型的长尾分类数据,若直接利用此类数据集进行训练,往往会对头部数据过拟合,而忽略尾部数据的类别。重采样是解决长尾分类数据的方法之一,即对样本量较多的数据集(0级、1级)欠采样(under-sampling)和对样本量较少的数据集(2级及以上)过采样(over-sampling)。尽管重采样的方法可能会导致部分结果过拟合,但研究显示在提升CNN实际训练效果方面,重采样的方法常常能带来更好的结果^[16-17]。

Focal Loss主要用于解决图像领域数据不平衡、分类难度差异所造成的模型性能问题^[18]。Focal Loss通过控制正负样本的权重以及容易分类和难分类样本的权重提升模型的性能。为解决训练时前景和背景数量存在的不均衡,引入了权重因子 α 和调制因子 $(1-p_i)^\gamma$,这样保证正负样本以及难易分类样本的相对平衡。Focal Loss的计算

公式为: $FL(p_i) = -\alpha_i(1-p_i)^\gamma \log(p_i)$ 。

1.5 算法性能评估

1.5.1 全集中算法评估

本研究选取870张不同等级的角膜染色图片作为训练集和验证集用于训练模型和优化模型,100张不同等级的角膜染色图片作为测试集评估模型有效性。通过全集中模型评估结果与3名经验丰富的视光师综合评估结果进行比较,以模型准确性(accuracy, ACC)、灵敏性、特异性作为评估参数。

1.5.2 AI与初学者准确性及诊断时长比较

由3名经验丰富的视光师首先对其余100张图片进行角膜染色分级的评估,并以此为参考,比较2名来自四川大学华西临床医学院眼视光学专业的初学者以及AI算法的评级准确性及诊断时长的差异。

1.5.3 AI和初学者前后两次评级的一致性比较

在完成初次评级的1个月之后,将图片调整为水平镜像图,并在Excel表格中利用随机数打乱图片顺序,在相同

的条件下再次完成同样的分级评估,且事先没有告知受试者第二次评估的内容和方法。将所有受试者两次评估的结果进行比较,以评估AI和初学者前后两次评估的一致性程度。一致性程度采用kappa值,计算公式为 $k = \frac{p_0 - p_e}{1 - p_e}$ [19]。其中 p_0 为观察符合率, p_e 为机遇符合率。

1.6 统计学方法

采用K-S检验分析数据分布正态性,满足正态分布的计量资料用 $\bar{x} \pm s$ 表示,非正态分布的计量资料用中位数 (P_{25}, P_{75}) 表示,分类变量用频数表示。采用McNemar-Bowker检验分别比较AI与两名初学者的角膜染色分级的准确性。基于数据分布的特征,采用成组t检验或秩和检验比较AI与两名初学者的平均诊断时长。采用Bonferroni法进行校正, $P < 0.017$ 为差异有统计学意义。AI和初学者前后两次评级的一致性采用kappa值分析。当kappa值位于0.0~0.20表示极低的一致性;位于0.21~0.40表示一般的一致性;位于0.41~0.60表示中等的一致性;位于0.61~0.80表示高度的一致性;位于0.81~1.0表示几乎完全一致[20]。

2 结果

2.1 全集中算法评估结果

算法构建共纳入有效图片870张,其中原图500张,370张水平镜面翻转图。采用4种深度学习模型构建,分别为ResNet34、ResNet50、DenseNet121以及DenseNet169,并采用重采样和Focal Loss的方法对每种模型进行优化。最终在训练集和验证集中,ResNet34模型表现出了最好的准确性,其训练的损失曲线和准确率曲线变化如图2所示。

通过全集、训练集、测试集的交叉验证,模型准确率见表1。ResNet34模型在角膜染色分级评估中的准确性为93.0%,敏感性为89.5%,特异性为89.6%,该结果在临床实践中被认为是可靠的。

2.2 AI与初学者比较

2.2.1 AI与初学者准确性及诊断时长比较

AI与两名视光初学者对染色图片分级的准确性及诊断时长的比较如表2所示。3组间评估准确性差异有统计学意义($P = 0.001$)。进一步两两比较,AI与初学者2之间差异有统计学意义($P = 0.001$),AI与初学者1之间差异无统计学意义($P = 0.094$)。3组间诊断时长差异均有统计学意义($P = 0.001$),AI诊断时长短于两位初学者。

2.2.2 AI和初学者前后两次评级的一致性比较

在完成初次评级之后的第二次评级准确性及诊断时长的比较如表3所示,其结果与初次评级结果相似,3组间

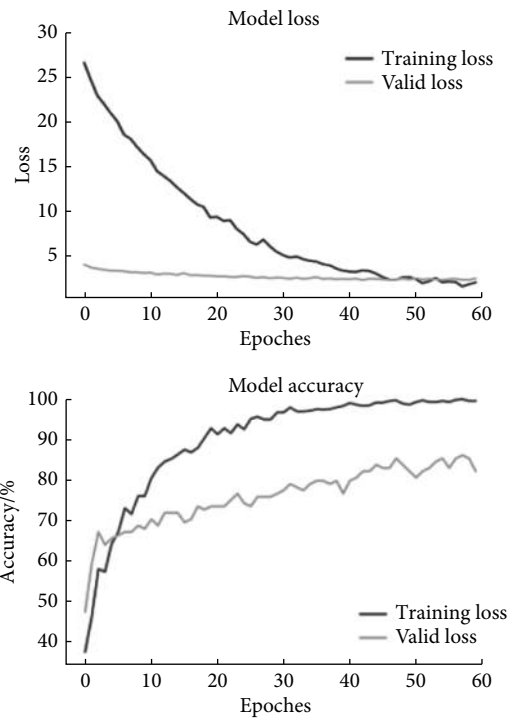


图 2 ResNet34模型损失和准确率变化图

Fig 2 ResNet34 model loss and accuracy graph for the training and valid dataset

表 1 ResNet34深度学习模型预测分级的数据分布及准确性

Table 1 Data distribution and accuracy in ResNet34 deep learning model for predicted grading results

Dataset	Number	Zero deviation of grading from estimated true Efron scale	Accuracy/%
Train dataset	870	815	93.7
Test dataset	100	87	87.0
Complete dataset	970	902	93.0

表 2 AI与初学者在分级准确性和诊断时长上的比较

Table 2 Comparison of grading accuracy and diagnostic time among AI and students

Variable	AI	Student 1	Student 2
Accuracy/%	87.0	78.0	52.0 [△]
t/s	1.00	11.86 [△]	13.25 [△]

[△] $P < 0.017$, vs. AI.

表 3 AI与初学者在分级准确性和诊断时长上的比较(水平镜像图)

Table 3 Comparison of grading accuracy and diagnostic time among AI and students (horizontally-rotated mirror photos)

Variable	AI	Student 1	Student 2
Accuracy/%	76.0	70.0	40.0 [△]
t/s	1.00	9.91 [△]	10.99 [△]

[△] $P < 0.017$, vs. AI.

准确性及诊断时长差异均有统计学意义($P=0.001$),初学者2准确性低于AI, AI诊断时长短于两位初学者。

两次评级中不同评估者对染色图片的评估等级分布如图3、图4所示。两次评级一致性指数如表4所示。在两次评级中, AI表现出了高度的一致性($kappa_{AI}=0.658$, $P_{AI}=0.001$), 两位视光初学者一致性中等($kappa_{S1}=0.575$, $P_{S1}=0.001$; $kappa_{S2}=0.609$, $P_{S2}=0.001$)。

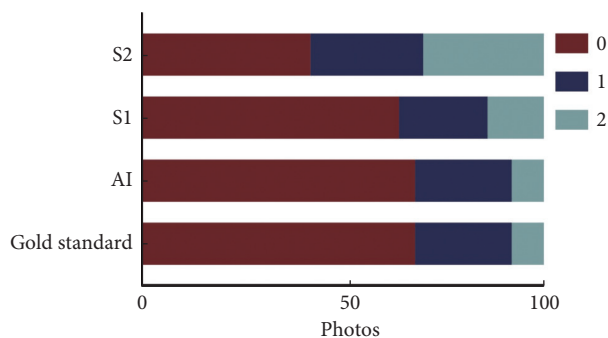


图3 AI与初学者在角膜染色评估的等级分布情况(原图)

Fig 3 Distribution of grading results for corneal staining assessment by AI and students (original photos)

AI: artificial intelligence; S1: student 1; S2: student 2.

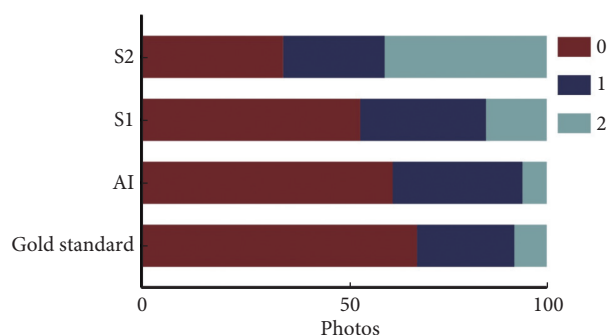


图4 AI与初学者在角膜染色评估的等级分布情况(水平镜像图)

Fig 4 Distribution of grading results for corneal staining assessment results by AI and students (horizontally-rotated mirror photos)

AI: artificial intelligence; S1: student 1; S2: student 2.

表4 AI与初学者在角膜染色评估一致性上的比较($kappa$ 值)

Table 4 Comparison of intra-rater agreement between AI and students ($kappa$ value)

Grader	First grading vs. standard	Second grading vs. standard	First grading vs. second grading
AI	0.726	0.517	0.658
Student 1	0.561	0.469	0.575
Student 2	0.230	0.110	0.609

2.3 AI误判图像

通过对测试集中AI未正确评估图像进行分析,总结了可能影响AI评估结果的4种典型图像。①荧光素钠染

料未均匀分布于患者角膜上(图5A)。②染料逐渐褪去,图像颜色较暗(图5B)。③拍摄过程中裂隙灯显微镜的亮度不完全一致,部分图片亮度过高(图5C)。④图片拍摄过程中,患者泪膜发生破裂(图5D)。

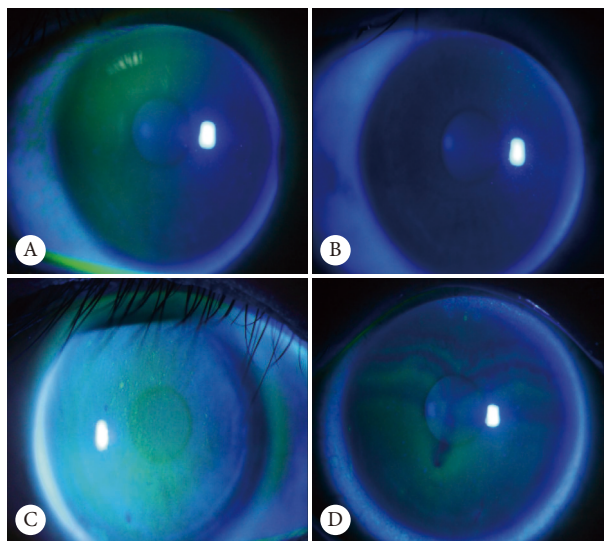


图5 AI未正确评级的图像示例

Fig 5 Examples of the photos that AI did not grade correctly

A, Uneven staining; B, staining fading; C, excessive slit lamp luminance; D, tear film breakup.

3 讨论

角膜上皮损伤是配戴接触镜中最常见的并发症之一^[5]。观察和诊断角膜上皮损伤最常用的方法是角膜染色^[4]。目前临床上主要依靠有经验的视光师对患者角膜染色的情况进行分级评估,这不仅耗费了大量的人力和时间,并且对于视光初学者来说评级会有一些的困难。

角膜染色在评估配戴接触镜后眼表完整性、比较不同接触镜治疗安全性以及观察治疗效果上起着重要的作用,但目前仍缺乏一种临床上可获得的、客观的评估方法。

近年来,深度学习算法已在角膜疾病领域如干眼病和圆锥角膜等疾病的诊断中取得了广泛进展。本研究使用ResNet34深度学习算法模型应用于角膜染色的分级评估,算法对870张角膜染色图片进行深度学习,结果显示模型在角膜染色分级诊疗上的准确性为93.0%,敏感性为89.5%,特异性为89.6%,这在临床上被认为是可靠的。除本研究所使用的深度学习模型,一些研究使用计算机图像处理的方式评估角膜染色分级^[21-23]。CHUN等^[21]将两位眼科医生与使用Microsoft Visual C++和Open CV® (Open Source Vision Library)编程算法的角膜染色评级结果分别进行比较,其结果显示出了较好的一致性($R=0.850$ 和 $R=0.903$),但与本研究所使用的深度学习算法模型相比,

其更加复杂且不易获取,因此临床上并未受到广泛应用。

在AI与视光初学者的比较中,AI表现出了较高的准确性和较短的用时。在准确性方面,AI的表现基本与模型训练的结果相似,其结果与初学者接近或高于初学者($ACC_{AI}=87.0\%$, $ACC_{S_1}=78.0\%$, $ACC_{S_2}=52.0\%$),且AI($t_{AI}=1.00\text{ s}$)在诊断时长上明显优于初学者($t_{S_1}=11.86\text{ s}$, $t_{S_2}=13.25\text{ s}$)。在第二次评估中,AI($ACC_{AI}=76.0\%$)准确性仍然与初学者接近或高于初学者($ACC_{S_1}=70.0\%$, $ACC_{S_2}=40.0\%$),但AI和初学者评估准确性和一致性均有所下降。由于DL是一种端到端的学习方法,即输入图像,机器直接输出分类结果,不需要人工进行编码,因此难以直接解释两次评级结果之间的差异,这种现象也被称为“黑箱现象”^[24]。两位初学者除第二次评估准确性和一致性下降外,两次评估间的一致性也一般,可能是由于第一次评级结果假阳性率较高,此外也可能是由于第二次评估时初学者遗忘了部分评估方法。

众所周知,临床经验对于眼科及视光从业者对疾病的评估准确性来说至关重要^[25]。KOURUKMAS等^[26]对角膜染色的重复性进行研究,发现研究人员对角膜染色的两次评估一致性一般($kappa=0.426$)。与本次研究相似,人类评估者的评估一致性不高($kappa_{S_1}=0.575$, $kappa_{S_2}=0.609$),而AI在与标准分级比较以及前后两次评级的比较中一致性程度均高于两位视光初学者。由此可见,AI对于图像评估的重复性优于人类评估者。由于接触镜需要长期的随访以确保镜片配戴的安全性,但每次复查很难保证患者将在同一家医院由同一位视光师为其进行复查,由于评估者以及量表使用的不同可能会导致评级结果产生一定的差异。因此,未来将AI应用于临床将有助于提高不同地区、不同评估者、不同评估时间点间的评估一致性。

视光初学者虽进行了3年的视光相关基础课程的学习,但对于角膜染色评级仍缺乏一定的经验。本研究发现,AI对于3种评级的分布基本接近于3位视光师的综合评估结果。两位视光初学者的评级结果中,2级及以上的评级比例较视光师更高,这可能是由于初学者缺乏一定的临床经验,容易将图片中由于染料分布不均的部位误判为角膜损伤区域,此外评级时仅参考Efron分级标准的手绘彩图进行评估而不参考患者其他配镜资料也容易将损伤情况严重化。因此从本研究结果来看,未来使用AI辅助初学者进行诊断能起到一定的帮助,有一定的教学指导意义。

通过对测试集中AI未正确评估的图像进行分析,发现在本次研究中影响AI评估结果的原因主要为:①染料

未均匀分布于角膜上。②染色到拍摄时间间隔较长,部分染料褪去导致角膜区域几乎无明显的荧光素钠。③因弥散光的亮度不完全一致使点染部位与周围对比度降低。④泪膜破裂导致图像出现黑色斑块。以上4种原因均可能对AI寻找目标特征造成一定困难,导致评级结果产生一定的偏倚,但偏差程度基本未超过1级。虽然AI对以上情况的评估结果会产生一定的偏差,但视光师在对这类图评级时仍然能够准确评估,提示今后在对这类图片评估时应优先选择经验丰富的视光师对评级结果进行校正。同时,在今后的研究工作中,将进一步明确角膜染色图片拍摄时裂隙灯显微镜弥散光所使用照明的范围,染色后拍摄的时间范围等,从而提高AI在角膜染色分级评估时的准确性和可靠性。

本研究仍有一定的局限性,首先,由于我院隐形眼镜门诊的患者中,3级和4级的角膜染色情况较为少见,因此在算法设计时统一将2级、3级、4级的图片重新分类为2级及以上。但在临床中对于2级及以上的角膜染色基本处理原则相近,即停戴镜片,寻找病因,滴用促上皮生长因子等方式得到恢复,因此该分类方法并不会对临床的处理结果有显著影响。此外,由于目前的角膜染色分区标准并没有对中央区的大小和范围做明确的规定,因此使用AI直接对角膜染色分区进行判定有一定困难。最后,本研究所收集的图片均来自四川大学华西医院,样本数量较少。未来将收集更多来自不同地区和医院的染色图片,进一步扩大算法设计的样本量。

综上所述,本研究成功将深度学习算法应用于角膜荧光染色分级评估,该模型准确性符合临床预期。在与初学者的比较中,AI算法表现出了较高的准确性和较好的一致性,能够指导视光师在临床中快速准确地评定角膜染色分级。此外,图像拍摄亮度、时间等原因可能会使AI对角膜染色的评级结果产生一定偏倚,今后将进一步明确角膜染色拍摄的条件,减少因此类原因导致的AI评判失误。

* * *

作者贡献声明 杨必负责论文构思、数据审编、经费获取和审读与编辑写作,赵雨暄负责正式分析、调查研究、研究方法、可视化和初稿写作,刘陇黔负责研究项目管理、提供资源和监督指导,张啸云负责软件和验证。所有作者已经同意将文章提交给本刊,且对将要发表的版本进行最终定稿,并同意对工作的所有方面负责。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] DOLGIN E. The myopia boom. *Nature*. 2015, 519(7543): 276-278. doi: 10.1038/519276a.
- [2] VERKICARLA P K, OHNO-MATSUI K, SAW S M. Current and

- predicted demographics of high myopia and an update of its associated pathological changes. *Ophthalmic Physiol Opt*, 2015, 35(5): 465–475. doi: [10.1111/opo.12238](https://doi.org/10.1111/opo.12238).
- [3] HOLDEN B A, FRICKE T R, WILSON D A, *et al*. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology*, 2016, 123(5): 1036–1042. doi: [10.1016/j.ophtha.2016.01.006](https://doi.org/10.1016/j.ophtha.2016.01.006).
- [4] BRON A J, ARGÜESO P, IRKEC M, *et al*. Clinical staining of the ocular surface: mechanisms and interpretations. *Prog Retin Eye Res*, 2015, 44: 36–61. doi: [10.1016/j.preteyeres.2014.10.001](https://doi.org/10.1016/j.preteyeres.2014.10.001).
- [5] NICHOLS J J, SINNOTT L T. Tear film, contact lens, and patient factors associated with corneal staining. *Invest Ophthalmol Vis Sci*, 2011, 52(2): 1127–1137. doi: [10.1167/iovs.10-5757](https://doi.org/10.1167/iovs.10-5757).
- [6] BEGLEY C, CAFFERY B, CHALMERS R, *et al*. Review and analysis of grading scales for ocular surface staining. *Ocul Surf*, 2019, 17(2): 208–220. doi: [10.1016/j.jtos.2019.01.004](https://doi.org/10.1016/j.jtos.2019.01.004).
- [7] RAHIMY E. Deep learning applications in ophthalmology. *Curr Opin Ophthalmol*, 2018, 29(3): 254–260. doi: [10.1097/ICU.0000000000000470](https://doi.org/10.1097/ICU.0000000000000470).
- [8] CHASE C, ELSAWY A, ELEIWA T, *et al*. Comparison of autonomous AS-OCT deep learning algorithm and clinical dry eye tests in diagnosis of dry eye disease. *Clin Ophthalmol*, 2021, 15: 4281–4289. doi: [10.2147/OPTH.S321764](https://doi.org/10.2147/OPTH.S321764).
- [9] LAVRIC A, VALENTIN P. KeratoDetect: Keratoconus detection algorithm using convolutional neural networks. *Comput Intell Neurosci*, 2019, 2019: 8162567. doi: [10.1155/2019/8162567](https://doi.org/10.1155/2019/8162567).
- [10] EFRON N. Grading scales for contact lens complications. *Ophthalmic Physiol Opt*, 1998, 18(2): 182–186. doi: [10.1016/s0275-5408\(97\)00066-5](https://doi.org/10.1016/s0275-5408(97)00066-5).
- [11] EFRON N, MORGAN P B, KATSARA S S. Validation of grading scales for contact lens complications. *Ophthalmic Physiol Opt*, 2001, 21(1): 17–29. doi: [10.1046/j.1475-1313.1999.00420.x-i1](https://doi.org/10.1046/j.1475-1313.1999.00420.x-i1).
- [12] GUO Y, LIU Y, OERLEMANS A, *et al*. Deep learning for visual understanding: a review. *Neurocomputing*, 2016, 187: 27–48. doi: [10.1016/j.neucom.2015.09.116](https://doi.org/10.1016/j.neucom.2015.09.116).
- [13] ANWAR S M, MAJID M, QAYYUM A, *et al*. Medical image analysis using convolutional neural networks: a review. *J Med Syst*, 2018, 42(11): 226. doi: [10.1007/s10916-018-1088-1](https://doi.org/10.1007/s10916-018-1088-1).
- [14] HE K, ZHANG X, REN S, *et al*. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: Las Vegas, NV, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [15] HUANG G, LIU Z, PLEISS G, *et al*. Convolutional networks with dense connectivity. *IEEE Trans Pattern Anal Mach Intell*, 2022, 44(12): 8704–8716. doi: [10.1109/TPAMI.2019.2918284](https://doi.org/10.1109/TPAMI.2019.2918284).
- [16] BUDA M, MAKI A, MAZUROWSKI M A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw*, 2018, 106: 249–259. doi: [10.1016/j.neunet.2018.07.011](https://doi.org/10.1016/j.neunet.2018.07.011).
- [17] HORN G V, PERONA P. The devil is in the tails: fine-grained classification in the wild. ArXiv, 2017. doi: [10.48550/arXiv.1709.01450](https://doi.org/10.48550/arXiv.1709.01450).
- [18] LIN T Y, GOYAL P, GIRSHICK R, *et al*. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell*, 2020, 42(2): 318–327. doi: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826).
- [19] MACLURE M, WILLETT W C. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol*, 1987, 126(2): 161–169. doi: [10.1093/aje/126.2.161](https://doi.org/10.1093/aje/126.2.161).
- [20] LANDIS J R, KOCH G G. The measurement of observer agreement for categorical data. *Biometrics*, 1977, 33(1): 159–174. doi: [10.2307/2529310](https://doi.org/10.2307/2529310).
- [21] CHUN Y S, YOON W B, KIM K G, *et al*. Objective assessment of corneal staining using digital image analysis. *Invest Ophthalmol Vis Sci*, 2014, 55(12): 7896–7903. doi: [10.1167/iovs.14-15618](https://doi.org/10.1167/iovs.14-15618).
- [22] RODRIGUEZ J D, LANE K J, OUSLER G W, *et al*. Automated grading system for evaluation of superficial punctate keratitis associated with dry eye. *Invest Ophthalmol Vis Sci*, 2015, 56(4): 2340–2347. doi: [10.1167/iovs.14-15318](https://doi.org/10.1167/iovs.14-15318).
- [23] AMPARO F, WANG H, YIN J, *et al*. Evaluating corneal fluorescein staining using a novel automated method. *Invest Ophthalmol Vis Sci*, 2017, 58(6): BIO168. doi: [10.1167/iovs.17-21831](https://doi.org/10.1167/iovs.17-21831).
- [24] STEAD W W. Clinical implications and challenges of artificial intelligence and deep learning. *JAMA*, 2018, 320(11): 1107–1108. doi: [10.1001/jama.2018.11029](https://doi.org/10.1001/jama.2018.11029).
- [25] TAN D N H, TAN M, LIEW H, *et al*. A data-driven approach to evaluate factors affecting resident performance in cataract surgery. *Int Ophthalmol*, 2023, 43(9): 3269–3277. doi: [10.1007/s10792-023-02730-1](https://doi.org/10.1007/s10792-023-02730-1).
- [26] KOURUKMAS R, ROTH M, GEERLING G. Automated vs. human evaluation of corneal staining. *Graefes Arch Clin Exp Ophthalmol*, 2022, 260(8): 2605–2612. doi: [10.1007/s00417-022-05574-0](https://doi.org/10.1007/s00417-022-05574-0).

(2023-03-14收稿, 2023-07-07修回)

编辑 余琳



开放获取 本文遵循知识共享署名—非商业性使用4.0国际许可协议(CC BY-NC 4.0), 允许第三方对本刊发表的论文自由共享(即在任何媒介以任何形式复制、发行原文)、演绎(即修改、转换或以原文为基础进行创作), 必须给出适当的署名, 提供指向本文许可协议的链接, 同时标明是否对原文作了修改; 不得将本文用于商业目的。CC BY-NC 4.0许可协议访问<https://creativecommons.org/licenses/by-nc/4.0/>。

© 2023 《四川大学学报(医学版)》编辑部 版权所有