doi:10.3772/j.issn.2095-915x.2018.01.004

人工智能糖网眼底图像识别在真实世界的应用

黄潇[1] 谷硕[2] 马晓晔[1] 梁文君[1] 张韵[1] 高连娣[1] 魏锐利[1]

1. 海军军医大学长征医院眼科 上海 200003;
2. 北京郁金香伙伴科技有限公司 北京 100091

摘 要 人工智能图像识别为医学影像识别带来革命性的发展。我们团队前期基于超过 18 万张来自 EyePACS 的眼底彩照构建了辅助诊断糖尿病视网膜病变（文中简称糖网）的人工智能深度学习算法模型。在实验室条件下，该模型灵敏度和特异度分别为 95.3% 和 79.5% 的高分。本研究以 Airdoc 眼底病智能识别平台用户上传的糖尿病患者眼底彩照 34100 张为测试数据集，验证该算法在真实世界中的准确性。结果真实应用中，模型检测糖网的敏感性 sensitivity（判断被测者不患病的准确度）和特异性 specificity（判断被测者患病的准确度）分别是 94.6%、78.4%。该人工智能模型诊断糖网的准确度已超过三甲医院眼科医师平均水平，基本达到眼底病医师水平，且工作效率远远高于人类医师，为大样本人群普查、筛查是否有糖网病提供了有力工具。

关键词： 深度学习算法；糖尿病视网膜病变；眼底彩照

中图分类号： G35

开放科学（资源服务）标识码（OSID）

Artificial Intelligence of Diabetic Retinopathy Image Recognition Used in the Real World

HUANG Xiao[1] GU Shuo[2] MA Xiaoye[1] LIANG Wenjun[1] ZHANG Yun[1] GAO Liandi[1] WEI Ruili[1]

1. Department of Ophthalmology, Shanghai Changzheng Hospital, Second Military Medical University, Shanghai 200003, China;
2. Beijing Tulip Partners Technology Co., Ltd, Beijing 100091, China

作者简介： 黄潇（1980–），博士，主治医师，讲师，研究方向：眼科学，Email：sophiahx@smmu.edu.cn；谷硕（1992–），硕士，工程师，研究方向：电子工程，共同第一作者；马晓晔（1973–），博士，副教授，副主任医师，研究方向：眼科学；梁文君（1985–），学士，医师，研究方向：眼科学。张韵（1980–），学士，医师，研究方向：眼科学；高连娣（1975–），硕士，主管护师，研究方向：护理学与眼科特殊检查；魏锐利（1963–），硕士，教授，博导，主任医师，研究方向：眼科学，Email：ruiliwei@smmu.edu.cn，共同通讯作者。

Abstract Artificial intelligence image recognition brings revolutionary development to medical image recognition. This study built a deep learning algorithm based on more than 180000 colorful photos of fundus from EyePACS for automated detection of diabetic retinopathy (DR). The test results of sensitivity and specificity achieved 95.3% and 79.5% respectively under laboratory conditions. This research tested the accuracy of the algorithm in the real world by using 34100 colorful photos of fundus uploaded by the users of Airdoc retinal disease intelligent recognition center from the internet. For detecting DR, the sensitivity and the specificity of the algorithm was 94.6% and 78.4%, respectively. The results indicated that the deep learning algorithm model had copied ophthalmologists' experience of diabetic retinopathy diagnose to an auxiliary diagnosis soft with higher efficiency similar accuracy as the human doctors. It is a potential helper for DR screening in huge people in the future.

Keywords: Deep learning; diabetic retinopathy; eye fudus color photos

1 引言

中国的糖尿病患者患病率为11.6%[1]，这1.14亿糖尿病患者每年每人都需要检查眼底以早期发现糖尿病视网膜病变[2]（下文简称糖网）。糖网早期治疗可保持有用视力。发现太晚会导致不可逆的视力丧失[3, 4]。糖网患者眼底筛查工作意义重大，但中国仅有三万五千名眼科医师，即使不做其他工作仅做眼底筛查，也不能完成这样巨大的工作量。而判读眼底照片是否发生了糖网，是一件专业性较强的任务，通常全科医师、内科医师甚至内分泌专业医师也难以胜任。

深度学习算法出现以来[5-7]，人工智能图像识别技术进步一日千里[8-9]，在解决特定问题上的表现已经超越了人类，带来医学影像识别领域许多革命性的进展[10]。国内已经有不少科研单位从事相关研究工作[11-16]，在糖尿病视网膜病变筛选领域也进行了一些尝试[17-19]，但均未在真实场景下使用过。依托国家科技信息资源综合利用与公共服务中心（STI）的支持，我们团队前期研究完成构建的人工智能糖网识别算法模型，在实验室条件下诊断糖网的准确率已经超过99%，远远超过人类眼科医师的平均水平，甚至超过眼底病专家的诊断水平。近来，我们利用一组真实世界的数据测试了该算法模型的灵敏度和特异性。与谷歌之前类似的工作相比[20]，我们的数据量更大，而且采用了Inception-v3与SVM结合的机器学习算法。尽管对于深度卷积神经网络算法，当图像数量大于60000张时，模型的性能已经达到饱和，但是我们发现Inception-v3与SVM结合的机器学习算法的性能优于深度卷积神经网络算法。

2 深度学习模型的构建

本研究建立的深度学习模型可以为医生提供诊断建议。本章介绍数据集，训练方法和模型评估方法。

2.1 数据集

模型的训练数据集来自Kaggle中的糖尿病性视网膜病变检测比赛[21]。数据集由来自

doi:10.3772/j.issn.2095-915x.2018.01.004

90056 病例的双眼眼底图构成，共计 180112 幅高分辨率图像。眼科医生根据国际糖网分期标准给每张图像分类。

0- 无糖尿病性视网膜病变

1- 轻度糖尿病性视网膜病变

2- 中度糖尿病性视网膜病变

3- 重度糖尿病性视网膜病变

4- 增殖性糖尿病性视网膜病变

数据集中 5 类图像的比例并不是均衡的，比例见表 1：

表1 原始数据集五分类图片比例

分类	0	1	2	3	4	Total
数量	137056	11263	23986	4172	3635	180112

这种数据不均衡的情况会导致图像数量占比大的分类准确率和召回率增强，而图像数量占比小的分类的图像特征学习不足。对于此数据集，相对于其他 4 类数据，0 类数据占绝对优势，如果我们在模型训练之前不添加预处理，模型对于 0 类将表现极高的特异性和敏感性。另一方面，其他类别分类准确性也会受到影响。因此，需要通过图像预处理来扩展数据集使得各个分类的数据比例均衡。

业内广泛应用的数据增强的方法如：裁剪，旋转，镜像变换和直方图均衡等。对于图 1 所示的眼底图像，由于眼底图像是圆形图像，旋转操作不会改变图像的特性。图像旋转用于提供额外的图像。

由于某些质量不够好（如曝光不足，漏光，镜片污渍），很多病理信息丢失。所以从数据集中删除低质量，甚至眼科专家无法根据眼底图像诊断糖网分级的图像。

去除掉低质量的图像后，通过图像旋转来增强图像不足的类别。由于类别 0 和类别 2 的数据充足，所以分别选择其中高质量的数据 15000 张和 15000 张。为了增加额外的图像，如图 1 类别 1 的数据旋转到 90 度，将类别 3 和类别 4 的图像旋转到 45 度，90 度，225 度和 270 度，增强后的图像数据量是原始数据集的 3 倍。预处理后的数据集中 5 类图像的比例如表 2 所示。

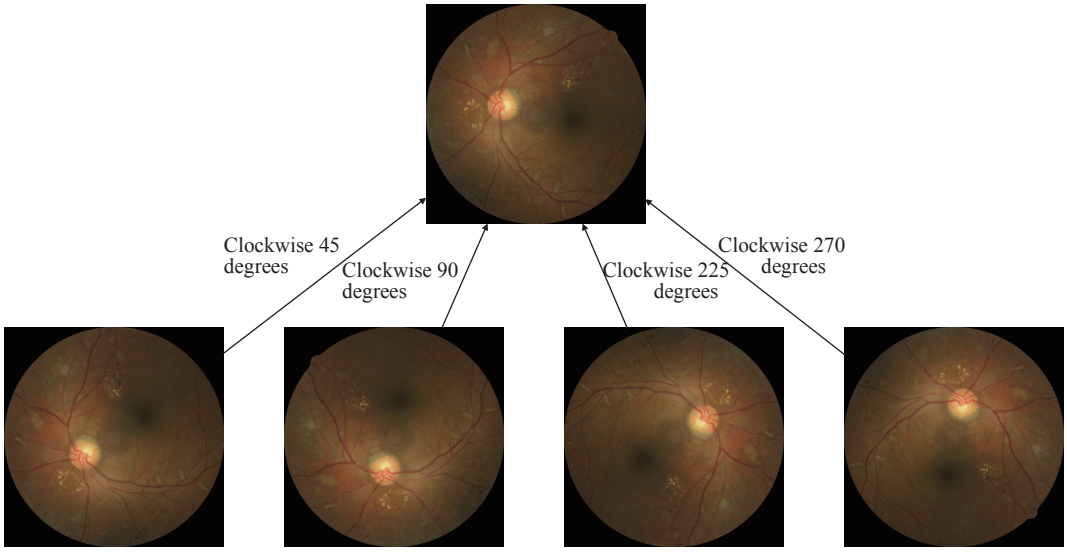

图1 图像增强处理

表2 增强数据集五分类图片比例

分类	0	1	2	3	4	Total
数量	15000	15000	15000	15000	15000	60000

2.2 训练

2.2.1 基础卷积神经网络训练

采用深度卷积神经网络 Convolutional Neural Network（CNN）ResNeXt 50 [22] 和 Inception-v3 [23] 来构建图像分类模型。ResNeXt 50 和 Inception-v3 的标准输入图像大小分别为 224×224 和 299×299。

然而，数据集中的眼底图像是 2300×2300。因此通过修改网络结构来接收大尺寸图像输入。同时，过大的图像会导致内存溢出。最终设定输入图像为 961×961。

为保持图像纵横比，图像的短边和长边比例应保持一致。调整图像大小分为以下 2 个步骤：将短边缩放至 961，使用相同的缩放因子缩放长边。然后，从图像的长边中心切割一个 961 大小的区域。

将 60000 张总数据集分为 3 组：训练组，验证组和测试组。每组比例分别为 80%，10%，10%。训练组和验证组用于训练网络。测试组只用于评估模型性能。经过 40000 次迭代，Inception v3 和 ResNeXt 50 模型得到收敛。

2.2.2 双眼融合 SVM

根据临床知识：患者的两只眼睛的糖网等级具有高度相关性。例如，如果一位患者的一只眼睛是糖网 4 级，则有较高的可能性他 / 她的另一只眼睛也是糖网 4 级。因此将此重要的先验知识用于糖网分类系统。

双眼融合 SVM[24] 是将训练好的深度卷积网络作为特征提取器。提取网络中最后一个池化层输出的（2048，1）维向量。然后，将每位患者的左右眼两个特征向量混合在一起，生成：(4096, 1) 维向量。混合方法保持每只眼睛特征中的内部顺序，并将两个特征联系在一起。如果第一个（2048，1）维特征向量来自左眼，他将代表左眼的特征，反之亦然。

使用随机图像增强方法来生成训练支持向量机需要的更多的眼睛特征。本研究通过旋转，翻转，亮度增强，对比度增强，锐度增强和噪声滤波器生成眼睛特征来对每位患者的眼底镜图像进行随机增强。而且因为算法对左右眼是区别对待的，所以应该保护两个眼底图像之间的差异。左右眼眼底镜图像之间的主要区别在于视盘的位置：左眼视盘（视网膜上视觉纤维汇集穿出眼球的部位）位于图像的左侧，右眼视盘位于图像的右侧。为了保持视盘的位置，图像旋转角度应小于 15 度，图像翻转采用垂直翻转。左右眼眼底图像在一个增强过程中使用相同的参数。

SVM 的内核是线性的，可以在 1000 次迭代后得到收敛。

2.2.3 评价标准

通过比较准确率，敏感性，特异性来判断模型的优劣。具体定义为：

$$\text{敏感性：}\frac{TP}{TP+FN}$$

$$\text{特异性：}\frac{TN}{TN+FP}$$

$$\text{准确度：}\frac{TP}{TP+FP}$$

TP：真阳性，FP：假阳性，
TN：真阴性，FN：假阴性。

doi:10.3772/j.issn.2095-915x.2018.01.004

2.3 实验室条件结果

使用测试组数据评估后，对于基本的 CNN 模型，Inception-v3 相对于 ResNeXt 50 的性能更好。ResNeXt 50，Inception-v3 和加入双眼融合 SVM 的 Inception-v3 的 5 分类正确率分别为 77.02%, 76.76% 和 82.54%。

如果将 0 级，1 级作为一类，2 级，3 级，4 级作为一类。Inception-v3 的二分类准确率为 90.22%，敏感性为 92.895%，特异性为 78.125%。ResNeXt 50 和 Inception-v3 之间的受试者工作特征曲线 Convolutional Neural Network（ROC）曲线比较如图 2 所示。

如图 3 所示，添加双眼融合后的 Inception v3 新模型，准确率为 92.62%，敏感性为 95.3%，特异性为 79.5%。

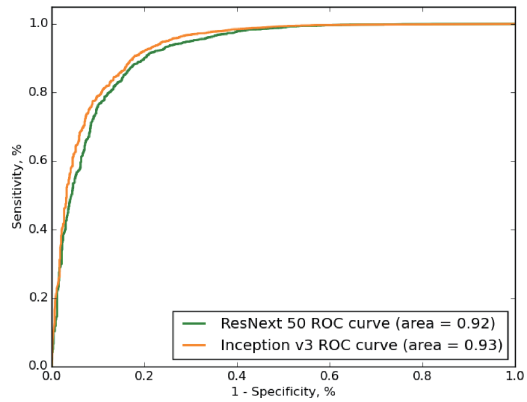

图2 ResNeXt 50 和Inception v3受试者工作特征曲线ROC曲线

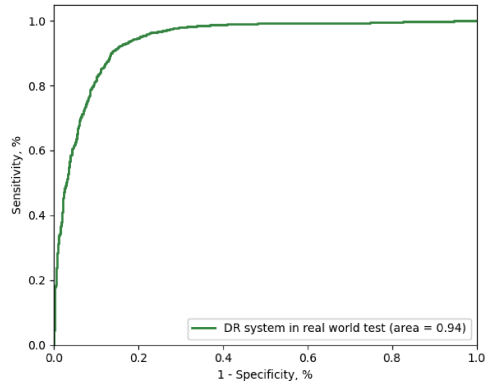

图3 Inception v3 DR分类模型受试者工作特征曲线ROC曲线

3 深度学习糖网智能识别模型在真实世界的应用

3.1 数据集

我们将上述深度学习糖网智能识别模型构建完成后成立了基于互联网的 Airdoc 眼底病智能识别平台，接受来自医院、社区和患者个人用户上传的糖尿病患者真实眼底照片共计 34100 张。以此为真实世界测试集。剔除了其中过曝，模糊和有污点的图像，但不做任何图像增强处理。

3.2 测试方法

3.2.1 算法模型选择

采用添加双眼融合后的 Inception-v3 新模型测试真实世界眼底照片数据集。因为在实验室条件下我们发现，采用采用添加双眼融合后的 Inception-v3 模型有最佳的准确度，敏感性和特异性。添加双眼融合后的 Inception-v3 模型本质上综合了深度神经网络和 Inception-v3 两种算法。分类的准确度更高。

3.2.2 具体流程

以二分类模型为例，五分类流程类似。图片作为二维的数组输入到模型中，模型输出一个 0 到 1 的非整数。1 为糖网，0 为非糖网，则输出的标量为输入眼底图片是糖网的几率。因为要将结果分为两类，所以需要设定一个阈值来区分糖网和非糖网。每设定一个阈值都可以通过对比人工和机器的判断结果计算出特异性和敏感性，对应为受试者工作特征曲线上的一个点。通过改变阈值从 0 到 1 即可得到受试者工作特征曲线。相应的对应关系为，阈值越大，

敏感性越高。

3.2.3 测试结果验证

以五分类模型为例，由两位三甲医院眼科专业主任医师分别评估数据集眼底照片给出的会诊结果为准，将眼底照片按糖网的严重程度分 5 级。再用已有的训练好的模型对临床所得眼底照片分 5 级。然后以专业医师的分类为标准，评估算法模型的准确性。具体的评价标准为灵敏性，特异性，和受试者工作特征曲线。灵敏性越高说明判断被测者患病的准确度越高，特异性越高说明判断被测者不患病的准确性越高。之所以做这样的区分是因为任何一种医学检测手段都不能百分之百确定被测者患病或不患病，而是以几率的形式给出结果。

3.3 结果

结果真实应用中，模型检测糖网的敏感性和特异性分别是 94.6%、78.4%。见图 4。

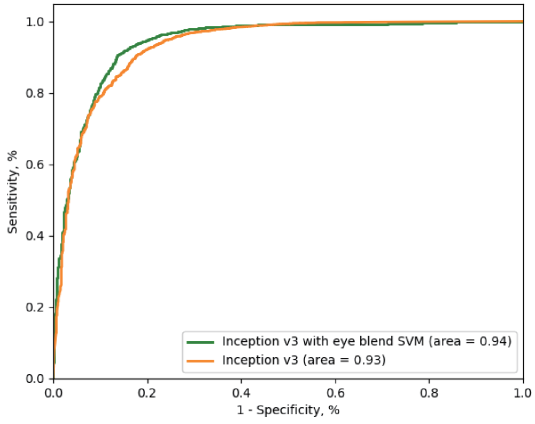

图4　真实世界 DR分类模型准确性ROC曲线

评估每张眼底照片，人工智能算法模型平均耗时 20 毫秒，单个人类医师则平均耗时 5 秒，是模型的 250 倍。换句话说，算法模型的效率是人类医师的 250 倍，而准确度达到人类专家会诊水平。

4 结论

该人工智能模型诊断糖网的准确度已超过三甲医院眼科医师平均水平，基本达到眼底病医师水平，且工作效率远远高于人类医师，为大样本人群普查、筛查是否有糖网病提供了有力工具 [25]。随着技术的改进，未来视网膜数字图像分析可能自动计算出更多临床有用数据，不仅对各类眼底疾病诊断提供支持，还可以为全身性疾病如高血压，脑卒中、阿尔海默病等提供辅助诊断信息 [26]。

参考文献

[1] Xu Y, Wang L, He J, et al. Prevalence and Control of Diabetes in Chinese Adults[J]. JAMA, 2013, 310(9):948-959.

[2] Wu M. Prevalence of Diabetes among Men and Women in China[J]. New England Journal of Medicine, 2010, 362(12):1090.

[3] Liu L, Wu X, Liu L, et al. Prevalence of Diabetic Retinopathy in Mainland China: A Meta-analysis[J]. Plos One, 2012, 7(9):e45264.

[4] Chakrabarti R, Harper C A, Keeffe J E. Diabetic Retinopathy Management Guidelines[J]. Expert Review of Ophthalmology, 2012, 7(5):417-439.

[5] Budden D, Matveev A, Santurkar S, et al. Deep Tensor Convolution on Multicores[J]. 2017, arXiv:1611.06565v3.

[6] Cho M, Brand D. MEC: Memory-efficient Convolution for Deep Neural Network[J]. 2017, arXiv:1706.06873

[7] Ravanbakhsh S, Schneider J, Poczos B. Equivariance Through Parameter-Sharing[J]. 2017, arXiv:1702.08389.

doi:10.3772/j.issn.2095-915x.2018.01.004

[8] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2261-2269.

[9] Shrivastava A, Pfister T, Tuzel O, et al. Learning from Simulated and Unsupervised Images through Adversarial Training[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2242-2251.

[10] Retinopathy Group of Academy of Ophthalmology of Chinese Medical Association. Clinical Diagnosis and Treatment Guidelines of Diabetic Retinopathy in our Country(2014) [J]. Chinese Journal of Ophthalmology, 2014, 50(11):851-865.

[11] 高歌, 马帅, 王霄英. 计算机辅助诊断在医学影像诊断中的基本原理和应用进展[J]. 放射学实践, 2016, 31(12):1127-1129.

[12] 王媛媛, 周涛, 吴翠颖. 深度学习及其在医学图像分析中的应用研究[J]. 电视技术, 2016, 40(10): 118-126.

[13] 刘长征, 相文波. 基于改进卷积神经网络的肺炎影像判别[J]. 计算机测量与控制, 2017, 25(4): 185-188.

[14] 张娟, 蒋芸, 胡学伟,等. 基于卷积受限玻尔兹曼机的医学图像分类新方法[J]. 计算机工程与科学, 2017, 39(2):323-329.

[15] 郭树旭, 马树志, 李晶,等. 基于全卷积神经网络的肝脏CT影像分割研究[J]. 计算机工程与应用, 2017, 53(18):126-131.

[16] 苏宝星. 基于深度学习的医学图像模式分类研究[D]. 杭州: 浙江师范大学, 2016

[17] 梁平, 熊彪, 冯娟娟, 等. 糖尿病视网膜病变眼底图像分类方法[J]. 深圳大学学报(理工版), 2017, 34(3):290-299.

[18] 丁蓬莉, 李清勇, 张振, 等. 糖尿病性视网膜图像的深度神经网络分类方法[J]. 计算机应用, 2017, 37(3):699-704.

[19] 庞浩, 王枞, PANGHao, 等. 用于糖尿病视网膜病变检测的深度学习模型[J]. 软件学报, 2017(11).

[20] 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.

[21] Kaggle[EB/OL]. [2017-11-13]. https://www.kaggle.com/c/diabetic-retinopathy-detection.

[22] Xie S, Girshick R, Dollár P, et al. Aggregated Residual Transformations for Deep Neural Networks[J]. 2016.

[23] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision[J]. 2015:2818-2826.

[24] Suykens J A K, Vandewalle J. Least Squares Support Vector Machine Classifiers[J]. Neural Processing Letters, 1999, 9(3):293-300.

[25] 丁胜, 赵湜, 王红祥,等. 免散瞳眼底照相筛查糖尿病视网膜病变临床价值[J]. 中华实用诊断与治疗杂志, 2011, 25(11):1096-1097.

[26] 陶相宜, 董志军, 董微丽. 高血压性视网膜病变与高血压靶器官损害的相关关系分析[J]. 河北医药, 2017, 39(13):2004-2007.