

· 智慧医疗 · doi:10.3969/j.issn.1671-8348.2022.16.030

网络首发 <https://kns.cnki.net/kcms/detail/50.1097.R.20220330.1638.004.html>(2022-03-31)

利用人工智能图像识别技术对染色体核型的分类研究*

丁晓燕,沈梦婕,李 岚,景鹏伟,黄国宁,叶 虹[△]

(人类胚胎工程重庆市重点实验室/重庆市生殖医学临床研究中心/重庆市妇幼保健院 400013)

[摘要] 目的 探讨利用人工智能图像识别技术对人类染色体核型识别的可行性。方法 将已标记的 1 652 例人类染色体图片分为 3 组,1 000 例训练深度神经网络,300 例测试神经网络的收敛效果,352 例验证训练后的神经网络的分类所需时间及其正确率。结果 人工智能深度神经网络对随机分布的染色体准确识别率达到 96.0%,识别时间短,与全自动染色体分析仪所需时间比较,差异有统计学意义($P < 0.05$),而准确率比较,差异无统计学意义($P > 0.05$)。结论 深度神经网络能够以较高准确率快速识别人类染色体核型。

[关键词] 染色体核型;人工智能;深度神经网络

[中图分类号] Q243;TP183

[文献标识码] A

[文章编号] 1671-8348(2022)16-2853-03

Research on classification of chromosome karyotypes by using artificial intelligence image recognition technology*

DING Xiaoyan, SHEN Mengjie, LI Lan, JING Pengwei, HUANG Guoning, YE Hong[△]

(Chongqing Municipal Key Laboratory of Human Embryo Engineering/Chongqing Municipal Clinical Research Center for Reproductive Medicine/Chongqing Municipal Maternal and Child Health Care Hospital, Chongqing 400013, China)

[Abstract] **Objective** To investigate the feasibility of using artificial intelligence image recognition technology to recognize the human chromosome karyotypes. **Methods** A total of 1 652 labeled human chromosome images were divided into three groups; 1 000 cases were used to train the deep neural network, 300 cases were used to test the convergence effect of the neural network, and 352 cases were used to verify the time and accuracy rate needed by classification of the trained neural network after training. **Results** The recognition rate of the artificial intelligence deep neural network on the randomly distributed chromosomes reached 96.0%, the recognition time was short, the needing time had statistical difference compared with the automatic chromosome analyzer ($P < 0.05$), while the accuracy rate had no statistical difference ($P > 0.05$). **Conclusion** The deep neural network can rapidly identify the human chromosome karyotypes with high accuracy rate.

[Key words] chromosome recognition; artificial intelligence; deep neural network

中国是出生缺陷高发国家,出生缺陷不但严重影响新生儿身心健康,也给家庭和社会带来沉重的负担^[1]。染色体异常是导致新生儿出生缺陷的常见原因,染色体结构或数目异常会引起一系列的临床疾病,是造成胎儿流产、新生儿畸形、智力低下、发育迟缓等疾病的根本原因^[2-4]。随着国家对出生缺陷的重视以及人们对生育质量要求的提高,产前诊断的需求量及工作量日益增加,羊水细胞培养及染色体核型分析是筛查胎儿先天疾病的首选方法^[5]。

随着辅助生殖技术(assisted reproductive technology, ART)的发展,越来越多的不孕患者选择通过 ART 解决生育问题。而随着“二孩”及“三孩”政策的

开放,高龄助孕的患者逐渐增多,助孕前的染色体核型检查及高龄孕妇的产前诊断已成为很多生殖中心的重要工作,需要安排大量的人力物力来完成,且需要等待较长的时间才能出具检测报告^[6]。

传统染色体核型分析主要依靠人工分析,经过专业训练的技术人员通过显微镜拍摄的单个细胞图片观察染色体,根据其形态结构判断染色体是否重复、缺失、易位或倒位。这一流程需要经验丰富的专业技术人员操作,虽然现有的专用显微镜可以自动或半自动进行染色体分类,但效果并非特别理想,未能分割的聚簇会导致系统误报,染色体碎片会增加系统的错误率^[6]。随着计算能力的飞速发展和大数据的应用,

* 基金项目:重庆市自然科学基金项目(cstc2018jcyjAX0710)。

作者简介:丁晓燕(1983—),主治医师,博士,主要从事生殖医学工作及研

究。△ 通信作者,E-mail:yehong1210@163.com。

人工智能(artificial intelligence, AI)图像识别技术深度神经网络成为新的图像识别热点。本研究以经过分类标记的染色体图库作为训练集,让神经网络学习人类染色体各个核型特征,实现可针对中期染色体图像自动分类过程,在保证分类准确率的同时降低对人工的依赖性。

1 资料与方法

1.1 数据集

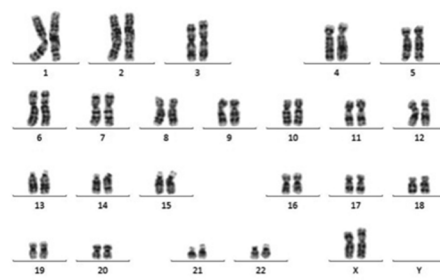
采用的所有图像均来自重庆市妇幼保健院生殖医学中心,包括外周血及羊水标本制备的染色体图像。图像库由 1 652 组不同中期染色体图片组成,每



图 1 中期染色体及其分析排版结果

组由两张图片组成,一张为经过专业医师分割的图片,一张为专业医师完成识别并分类排序的图片。见图 1。

随机将图像库中的 1 000 组归为训练组,用于神经网络参数的训练和修正。将 1 000 张已经标记的图片所包含的约 46 000 条染色体以 40 条一组重新随机分布到图片上,每条染色体需经过 0~360°的随机旋转,且互不重叠,最终生成 10 000 张图片组成训练组。剩余的 652 组图像分为测试组(300 例)与验证组(352 例),分别用于实时观察神经网络的收敛效果和检验神经网络的性能。



1.2 方法

1.2.1 软硬件平台

所有的网络均采用 Python3.6 语言编写的 TensorFlow1.2 框架下搭建,整个试验在 Ubuntu 16.04 系统下运行。计算机配置: Intel(R) i7-7820X CPU, 64G RAM, NVIDIA 2080TI 11GB GPU * 2。

1.2.2 不同染色体核型分类方法

一名长期进行染色体分类的专业技术人员利用普通显微镜(Olympus BX51)人工分析染色体图像(A组),利用全自动染色体显微图像扫描分析系统(ZEISS MetaSystems Imager Z2)对染色体图片进行分析(B组),与 AI 模型(AI 组)对染色体图片分析的准确率及分析时间进行对比($n=50$)。

1.3 统计学处理

采用 SPSS 20.0 进行统计分析,计量资料以 $\bar{x} \pm s$ 表示,计数资料以率表示,组间比较采用单因素方差分析,率的比较采用 χ^2 检验。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 AI 神经网络建立

采用一个通用图像分割网络 Mask-RCNN^[7],其是一种基于区域的神经网络架构。采用数据集预训练模型参数初始化对应的网络参数,其余参数进行随机初始化。验证组测试分类识别的准确率为 91.35%~98.61%,中位准确率为 94.38%。采用验证组对已收敛的深度网络进行测试,该网络能够较好地划出目

标区域并分类,见图 2。



图 2 经深度神经网络识别的染色体区域并分类

2.2 人机分类比较

长期进行染色体分类的专业技术人员利用普通显微镜进行一个染色体图像的分析时间大约为 10 min,而人工利用全自动染色体显微图像扫描分析系统对染色体图片进行分析,每个染色体图片大约需要 2~5 min,准确率均为 100%。

专业技术人员利用全自动染色体显微图像扫描分析系统分割粘连、重叠的染色体大约需要 1~2 min,然后将图片传递给深度神经网络,其能在几秒内完成染色体分类,正确率为 96.0%,因此在人工配合下可在

1~2 min 左右完成染色体的识别和分类。3 组的识别时间比较,差异有统计学意义($P < 0.05$);而准确率比较,差异无统计学意义,见表 1。

表 1 不同染色体分类方法耗时与准确率比较

组别	n	时间($\bar{x} \pm s, s$)	准确率(%)
A 组	50	620.3 \pm 144.4	100.0
B 组	50	184.1 \pm 60.9	100.0
AI	50	105.7 \pm 28.0 ^a	96.0

^a: $P < 0.05$,与 A、B 组比较。

3 讨论

以往的染色体核型分析工作几乎均由专业人员人工完成,包括培养、制片、镜检,其过程烦琐,虽然准确率有保障,但效率比较低,导致染色体检查结果等待时间较长。半自动或全自动染色体分析系统的出现使染色体核型分析的效率得到了很大的提高,然而这些系统只是在染色体图像的拍摄和选取方面做到了全自动,而在染色体图像分割和染色体核型分析方面还是需要专业技师的参与^[8],即仍然是通过人机交互的模式完成的,并没有真正做到“全自动”。

随着计算机计算能力的飞速发展和大数据的应用,神经网络成为新的图像识别热点。AI 图像识别技术用于医学影像的辅助诊断已有大量研究,包括胸部、心血管疾病、骨科疾病及神经系统疾病影像等,这些研究成果如果用于临床将大大减少临床医师的工作量及辅助提高诊断的准确率^[9-10]。关于染色体核型的 AI 分析也早有研究,但有很多技术难题没有克服,且准确率一直较低。WU 等^[11]采用 VGG-net-d 的网络架构,识别染色体的准确率达到 63.5%。GUPTA 等^[12]采用孪生双网络互纠正分类法,准确率达到 85.6%。LIN 等^[13]采用基于 ResNeXt 框架的迁移学习识别法,准确率达到 94.09%。然而这些研究的准确率还远远不足以用于临床。

染色体识别的两个难点在于分割和分类^[14]。由于显微镜下的染色体难免粘连或重叠,不成熟的分割算法导致后续的分类准确率降低^[15]。然而人工分割重叠的染色体要求的技术不高,时间较短,且后续神经网络识别率还能得到提高。而影响染色体分类的难点主要在于染色体弯曲折叠等不同的形态。有研究发现将弯曲折叠的染色体进行伸直处理,使染色体的识别率达到了很大的提升,然而这种方法并不能完全适用于所有染色体图像中的情况^[16]。本研究发现只需将各种形态的染色体随机旋转,反复训练神经网络便可以高效识别呈现各种形态的染色体。因此相较于传统的染色体核型分析流程,采用预先人工处理,再采用神经网络进行分类的方法,不但提高了效率,同时降低了对技术人员的要求。

本研究通过现有图像库训练出的神经网络,对染色体的识别率可高达 96.0%,所耗费的时间仅为传统

人工识别的五分之一,较全自动染色体分析系统的所需时间也显著降低,大大提高了染色体核型分析的效率,从而降低了患者等待检测结果的时间。由于培养一名专业的遗传学医师需要漫长的时间,其经验也需要在工作中不断积累,期间必定耗费大量的人力物力,需要大量的医务工作者为染色体核型分析付出大量的工作。本研究建立的 AI 神经网络可以高效、准确地识别出染色体核型,而且可以集成到现有的核型分析系统中或通过网络集成到云端。并且随着应用的推广,图像库的数据集会得到不断的补充,通过 AI 神经网络的自主学习及调整功能,识别率会得到进一步的提高,将大大减轻医护人员的工作负担,并减轻患者的经济负担及减少患者的等待时间。

参考文献

- [1] 封志纯,王艳. 我国出生缺陷防控研究与应用进展[J]. 中国儿童保健杂志, 2019, 27(8): 813-815, 819.
- [2] 唐芳,陶晓海,卢航,等. 染色体异常、染色体多态性与生殖异常的相关性分析[J]. 海南医学, 2015, 26(16): 2379-2381.
- [3] 朱军. 出生缺陷及其监测[J]. 中国实用妇科与产科杂志, 2002, 18(9): 513-514.
- [4] 朱元昌,朱正锴,袁琳,等. 早期自然流产胚胎染色体数目异常的临床分析[J]. 中国优生与遗传杂志, 2020, 28(7): 823-825, 852.
- [5] WILSON R D, BLIGHT C, LANGLOIS S. Diagnosing chromosomal abnormalities from “big” to “small” with molecular cytogenetic technology[J]. J Obstet Gynaecol Can, 2009, 31(5): 414-421.
- [6] 范继红. 人类染色体图像的分割及数目统计[D]. 哈尔滨: 哈尔滨工程大学, 2009.
- [7] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(2): 386-397.
- [8] 吕巍,王树玉. 应用染色体自动分析系统加强对技术人员染色体识别能力的培训[J]. 中国优生与遗传杂志, 2015, 23(10): 135-136.
- [9] 陆滢尘,胡屹玲. 医学影像人工智能的研发应用现状与挑战[J]. 人工智能, 2021(3): 11-19.
- [10] 孔鸣,何前锋,李兰娟. 人工智能辅助诊疗发展现状与战略研究[J]. 中国工程科学, 2018, 20(2): 86-91.
- [11] WU Y R, YUE Y, TAN X, et al. End-To-End chromosome karyotyping with data augmentation using GAN[C]//2018 25th IEEE International Conference on Image Processing (ICIP), 2018-10-07, 2018: 2456-2460. (下转第 2865 页)

- [25] LIAN N, TONG J, LI W, et al. Ginkgetin ameliorates experimental atherosclerosis in rats[J]. *Biomed Pharmacother*, 2018, 102:510-516.
- [26] 周红, 吴晓燕, 袁艺标, 等. 三种剂量维生素 D3 结合高脂饲料建立大鼠动脉粥样硬化模型的比较[J]. *中国动脉硬化杂志*, 2012, 20(11):995-998.
- [27] FU Y, SUN S, SUN H, et al. Scutellarin exerts protective effects against atherosclerosis in rats by regulating the Hippo-FOXO3A and PI3K/AKT signaling pathways[J]. *J Cell Physiol*, 2019, 234(10):18131-18145.
- [28] LÜ L, ZHANG D, SUN B, et al. Apocynum leaf extract inhibits the progress of atherosclerosis in rats via the AMPK/mTOR pathway[J]. *Pharmazie*, 2017, 72(1):41-48.
- [29] 何忠梅, 李成恩, 段翠翠, 等. 短梗五加果多酚预防大鼠动脉粥样硬化作用[J]. *食品科学*, 2018, 39(1):200-206.
- [30] 陶珍, 曲乐丰, 丁素菊, 等. 幽门螺杆菌感染致兔颈动脉粥样硬化模型的建立[J]. *第二军医大学学报*, 2010, 31(1):37-41.
- [31] 唐曦, 杨松, 苟博, 等. 高脂喂养联合免疫损伤建立兔动脉粥样硬化模型及评价[J]. *中国医药导报*, 2016, 13(18):21-24, 封 3.
- [32] 刘剑刚, 董国菊, 马鲁波, 等. 免疫损伤结合高脂饲料致兔动脉粥样硬化斑块形成的相关因素分析[J]. *实验动物科学*, 2013, 30(2):32-38.
- [33] YANG X, JIA J, YU Z, et al. Inhibition of JA K2/STAT3/SOCS3 signaling attenuates atherosclerosis in rabbit[J]. *BMC Cardiovasc Disord*, 2020, 20(1):133.
- [34] 张莉, 白林英, 母凯茜, 等. 灯盏乙素对小型猪动脉粥样硬化模型颈总动脉斑块组织中基质金属蛋白酶-1、2、9 表达的影响[J]. *中药药理与临床*, 2020, 36(6):91-97.
- [35] 王庆林, 陈文捷, 万佳, 等. 颈总动脉钳夹术加高脂饮食建立大鼠颈动脉粥样硬化模型[J]. *实用医学杂志*, 2010, 26(4):562-564.
- [36] 李土明, 徐丽英, 吴滢, 等. 补阳还五汤对颈总动脉粥样硬化模型大鼠炎症因子 MPO、Lp-PLA2 及 MMP-9 影响的实验研究[J]. *四川中医*, 2017, 35(10):45-47.
- [37] VON SCHEIDT M, ZHAO Y, KURT Z, et al. Applications and limitations of mouse models for understanding human atherosclerosis[J]. *Cell Metab*, 2017, 25(2):248-261.
- [38] ZHANG Y B, XU R, LI X J, et al. Effects of icariin on long noncoding RNA and mRNA expression profile in the aortas of apoE-deficient mice [J]. *Biosci Rep*, 2019, 39(7):BSR20190855.
- [39] JAWIEN J. The role of an experimental model of atherosclerosis: apoE-knockout mice in developing new drugs against atherogenesis[J]. *Curr Pharm Biotechnol*, 2012, 13(13):2435-2439.
- [40] HEWING B, FISHER E A. Preclinical mouse models and methods for the discovery of the causes and treatments of atherosclerosis[J]. *Expert Opin Drug Discov*, 2012, 7(3):207-216.
- [41] ZHAO Y, QU H, WANG Y, et al. Small rodent models of atherosclerosis[J]. *Biomed Pharmacother*, 2020, 129:110426.
- [42] FAN J, WANG Y, CHEN Y. Genetically modified rabbits for cardiovascular research [J]. *Front Genet*, 2021, 12:614379.

(收稿日期:2021-10-18 修回日期:2022-02-08)

(上接第 2855 页)

- [12] GUPTA G, YADAV M, SHARMA M, et al. Siamese networks for chromosome classification [C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2017:72-81.
- [13] LIN C, ZHAO G, YIN A, et al. A novel chromosome cluster types identification method using ResNeXt WSL model[J]. *Med Image Anal*, 2021, 69:101943.
- [14] 冯涛, 陈斌, 张跃飞. 基于改进的 Mask R-CNN 的染色体图像分割框架[J]. *计算机应用*, 2020, 40(11):3332-3339.
- [15] 邱俊玮, 孙频捷. 基于卷积神经网络的染色体核型分析方法研究进展[J]. *现代计算机*, 2021, 3:22-25.
- [16] JAVAN -ROSHTKHARI M, SETAREHDAN S K. A new approach to automatic classification of the curved chromosomes[C]//2007 5th International Symposium on Image and Signal Processing and Analysis, 2007:19-24.

(收稿日期:2021-10-18 修回日期:2022-02-11)