

网络出版时间:

网络出版地址:

人工智能时代的病理组学

闫雯¹,李楠楠¹,张益肇²,来茂德³,许燕^{1,2}

摘要:在人工智能时代,计算机辅助的病理诊断为病理学的发展提供了新的发展空间,病理学的研究不再局限于传统人工病理分析。在人工智能的辅助下,计算机不但可以实现定量化病理诊断,还可完成疾病预后等病理学的相关研究。本文通过总结近十年来人工智能在病理学中的研究成果,提出病理组学的概念。病理组学的研究内容是指基于基于人工智能将病理图像转化为高保真度、高通量的可挖掘的数据,并用于定量化病理诊断和疾病预后,最后自动生成病理诊断报告。在人工智能技术的支撑下,病理组学的研究正向着更加自动化更加精准的方向发展,这也有益于充分利用现有医疗资源、节省研究成本、推动医疗发展。

关键词:人工智能;病理组学;病理诊断;预后

中图分类号:R-1 **文献标志码:**A

文章编号:1001-7399(2018)06-

doi:10.13315/j.cnki.cjcep.2018.06.

随着计算机技术的快速发展,人工智能已逐渐应用于各个行业。从智能问诊到治疗方案的建议,再到基于大数据的新药研发,人工智能技术正向医学各领域不断渗透。病理学作为研究人体疾病发生原因、发生机制、发展规律的重要学科,其发展与新兴技术的注入息息相关,定量化病理诊断便是人工智能与病理学有机结合的产物。

人工智能在病理学中的应用包括细胞学初筛、形态定量分析、组织病理诊断和辅助预后判断等方面^[1]。虽然人工智能不能取代病理医师的作用,但是随着人工智能在病理学研究中的应用越来越广泛,病理学的研究内容不断地被重新布

局与定义。在人工智能技术的推动下,病理组学便应运而生。“组学”一词,来源于英文中的-omics 后缀,意指某些种类个体的系统集成,现在主要指从整体的角度出发去研究人体的细胞组织结构、基因与疾病间的关联。类似于基因组学、蛋白质组学、以及影像组学等,病理组学(Pathomics)就是将病理图像转化为高保真度、高通量的可挖掘的数据集,这个数据集涵盖纹理特征、形态学特征、边缘梯度特征、生物学特性等定量特征,并将这些特征用于定量化病理诊断和疾病的预后。

1 病理组学的起源

病理切片是疾病诊断的金标准。随着病理学科的迅猛发展和诊断需求的增加,病理切片的数量呈现出爆炸式增长,特别是高水平三甲医院病理诊断例数和切片张数每年都在高速增长,临床上产生的大量切片给病理医师的工作带来了压力。海量的病理切片及其所包含巨大的信息量,使得人工分析过于耗时耗力,难以应对临床需求。而且人工诊断的准确性受到病理医师主观经验的影响,导致出自不同医师之间的诊断结果不可复现^[1]。

随着计算机技术和互联网技术的发展,病理切片数字化、“云病理平台”的建立以及基于人工智能的病理切片定量分析等给病理切片领域带来了重大革新^[2]。

数字病理^[3]是利用数字技术对病理图像进行数字化处理(摄取、拼接、压缩、储存等),保留高质量图像信息,并结合数据库技术形成数字病理切片系统。数字病理打破了传统病理学在存储、保真性和检索等方面的局限,通过图像的浏览分析来完成病理分析、疾病诊断、远程传输和病理教学等任务。

近年来,在互联网技术发展的大环境下,“互联网+”的形式越来越普遍,“云病理平台”便是在这种形势下产生的。云病理的出现使得远程病理诊断更加方便,使数字病理从“概念上”的远程医疗逐渐过渡到实际应用中,形成了云病理平台。云病理是指通过远程病理系统将数字病理切片扫描并传输到“云”上,通过常规组织病理诊断、远程专家病理会诊、远程病理人才培养、病理科研协作等流程,保障病理诊断质量、解决临床问题;打破地域医疗资源分布不均的问题。

目前,病理切片的分析还处于定性分析为主、定量分析为辅的阶段。虽然在研究领域,定量分析已经取得了长足的进步,但是现阶段的定量分析多是研究少数常见的特征,例如形态学特征^[4]、色彩和纹理特征^[5]、形状特征^[6]等,但是这种单一的特征分析并不能涵盖肿瘤的复杂度和多变性。

接受日期:2018-02-08

基金项目:深圳科创委技术攻关项(shenfagai2016627)、国家自然科学基金(81771910)、北京市自然科学基金(4152033)、国家科学技术重大专项(2017yf0110903)、北京航空航天大学软件开发环境国家重点实验室中央高校基础研究基金(sklsde-2017zx-08),中国的111项目(b13003)

作者单位:¹深圳北航新兴产业技术研究院/北京航空航天大学生物与医学工程系/北京航空航天大学软件开发环境国家重点实验室/生物医学工程高精尖创新中心,北京 100191

²微软亚洲研究院,北京 100080

³浙江大学医学院基础医学院,杭州 310058

作者简介:闫雯,女,硕士研究生。E-mail:evanna.yw@gmail.com
许燕,女,博士,副教授,通讯作者。E-mail:xuyan@buaa.edu.cn

病理组学需要更全面地提取图像的特征数据,人工提取特征显然不太现实,这就需要机器学习算法自动地、高通量地、全面地提取图像中的特征数据,这个过程称为特征学习^[7]。特征学习通过组合低层特征形成更加抽象的高层特征(或属性类别),特征表达更加全面客观。Ciresan 等^[8]使用卷积神经网络自动学习乳腺病理切片的特征并检测有丝分裂数目,证明了不需要人工设计特征,神经网络能更全面地学习到病理切片的特征,给病理切片分析带来了新的契机。

在数字病理技术、人工智能技术和互联网技术的支撑下,病理组学旨在深度挖掘病理图像的信息,自动学习切片量化病理特征(如有丝分裂数目、肿瘤的实质与间质的比例,黏液湖和癌细胞的比例等),并与基因组学、影像组学之间建立一定的联系,形成多组学交互式诊断体系。具体来讲,病理组学基于人工智能的数据特征提取算法,将病理数据转化为可挖掘的特征数据,通过分析特征数据中与其他组学和临床病理指标的相关性,最终实现组织学诊断、病理分型和分级评估预后等工作。

2 病理组学的常用研究手段

病理组学是将多种分析算法用于多种病理图像的诊断中,它将人工智能算法有机地融入到临床病理分析的各个环节,使多种病理分析任务之间具有递进地可承接性,实现从病理分析到生成病理报告这个高度一体化诊断体系。

2.1 病理切片的采集与重建 病理图像可分为组织学图像和细胞学图像,包括从基本病理诊断 HE 切片,到特殊染色、免疫组化病理图像,也有荧光原位杂交染色图像等。为了克服病理切片图像之间的异质性为定量分析带来的干扰,在病理分析中图像归一化的方法是很有必要的。

病理切片图像通常含有多达百万数量级的显微结构,信息量极大,其大小可以达到 50 G,检查一位患者可产生 1T 的数据。如此庞大的数据量十分不利于图像数据的存储、传输以及分析。为此,在保证图像质量的前提下,应尽量减少图像中的冗余信息,对图像进行重建与压缩,既保留用于后续诊断的必要信息,又节省存储空间与计算花费^[9]。

2.2 病理切片特征的自动学习 特征表达^[7,10]是指将原始数据输入转换为可以在机器学习任务中能够被高效利用、便于处理的表达形式。WSI 病理切片的信息含量通常是复杂、冗余的和高度可变的,因此从原始数据中发现有用的特征或表示是病理切片分析的前提。特征表达可分为人工特征提取和自动特征学习两种形式。

传统人工方法提取特征往往耗时耗力,而且依赖于专业知识。病理切片的复杂性和专业性导致人工特征难以全面地概括切片的有用信息。自动特征表达是指算法自动地从海量大数据中去学习特征,这种表达形式更加全面、客观、更便于处理。因此,自动特征学习常常被用来提取病理切片的特征^[7,10]。

特征学习分为有监督特征学习^[11]和无监督特征学习^[12]两种。有监督的特征学习从有标记的数据中学习特征

表达,无监督的特征学习是指从无标签的数据中学习特征。文献^[13]将多尺度卷积网络(MSCN)提取的特征与人工设计特征相结合,用于宫颈细胞质和细胞核的分割。文献^[14-15]使用 CNN 自动提取病理切片的特征,并结合切片图像的特点,选择特定的后期处理方法分割病理切片中独立的细胞核。Xu 等^[16]采用基于迁移学习的 CNN 提取脑病理切片的特征,用于脑肿瘤的分割和分类。

以上文献证明,基于神经网络的自动特征学习提取的特征更加全面、细致,更适用于病理分析。

2.3 病理切片的量化病理诊断 传统病理诊断依赖于病理医师的主观经验,通过其对于癌症的形态学特征的辨认来达到疾病诊断的目的。但是这样的定性诊断有一定的误诊率,其诊断结果有时难以重复,而且不足以得出精准的预后判断。

在精准医疗的大背景下,研究人员需要深层挖掘切片的信息,形成一系列用于病理诊断的量化指标,使病理诊断更加客观可靠。目前,定量分析已经成为临床病理诊断的一部分,但是由于数字病理图像的背景噪声、组织间的重叠和组织结构的不规则性等因素的影响,使得算法难以从以上干扰信息中提取出病理诊断的临床特征^[17]。

量化病理诊断是指以特征学习为基础,从病理切片中提取特定组织结构并形成量化指标来评价病变的情况。其中,量化指标指有丝分裂的数目、肿瘤的实质与间质的比例,粘液湖和癌细胞的比例等。在实际应用中,往往根据病变发生部位的不同选择特定的指标完成病理诊断。在肠癌的检测中,腺体作为重要的诊断指标,往往需要将腺体从背景中分离出来以达到定量分析的目的^[18-19];在乳腺癌的诊断中,通常使用有丝分裂作为诊断指标,统计四个有丝分裂时期的不同形态有丝分裂像的数目作为病理诊断的依据^[8,20];在乳腺淋巴结转移的检测中,算法通过学习病变部位与正常组织的多种特征的差异,从而判断是否发生转移^[21-22]。

2.4 病理图像与癌症预后的相关性 病理诊断是癌症诊断的金标准,基于肿瘤组织学的病理检查(如肿瘤的大小、组织学的分级和管脉浸润等)、肿瘤的生物特征、以及患者本身的身体机能状态都是影响癌症的预后因素^[23]。例如,在胃癌患者的临床病理因素和预后分析中,文献^[24,25]通过免疫组化检测 p53 的表达,并分析临床病理因素和生存时间,可用于预后判断;在乳腺癌的预后中,淋巴结是否发生转移是重要的评价指标^[21-22]。但是目前的预后并没有给出一个可以量化的诊断指标,仍停留在定性层面,“预后较差”、“预后不良”等诊断并不能精确地描述患者的预后效果,因此如何量化这些因素给预后带来的影响是病理组学着重解决的问题。

在病理切片中,除了可描述性的特征之外,还有隐含在图像中可深度挖掘的信息对于疾病的预后有着重要作用。病理组学的目的是通过人工智能算法快速分析数量巨大的临床病理数据,以提取相关特征,自动检测病变区域,实现病

变区域快速量化(如面积、直径、灰度等),并确定病理特征与癌症预后之间的相关性,提高病理切片定量分析的效率、精度以及保证组织病理学分析的一致性。

2.5 基于病理图像分析结果生成病理诊断报告 病理报告^[26]通常分为快速冰冻病理报告和常规病理报告两种。前者通常是术中用于对肿瘤患者的肿块的性质做出判断,如果是癌症,就要对边缘是否含有肿瘤浸润进行确定,并要求30 min以内出结果。常规病理报告是指手术完成后,将从患者体内取下的组织固定、取材、脱水、包埋、切片等,然后根据疾病情况通过HE染色、特殊染色、免疫组织化学、荧光原位杂交等技术对疾病进行确诊,并为临床后续治疗提供信息的病理报告。

传统病理报告单一般包括大体所见、镜下所见、病理诊断三方面内容。其中,病理医师在显微镜下观察到的现象以及在分子水平上对活检组织进行的专业性描述,是非常重要的内容。类似地,电子病理诊断中,计算机使用分析算法对病理切片进行分析,并将分析结果用文字的方式记录下来,以便为后续研究提供参考。

因此,电子病理报告的建立就是使用人工智能算法构建一个既可以生成自然语言描述又可以分析病理图像的模式,该模型可以描述病理诊断的内容、分子检测的结果,也包括病理组学及其临床意义。

3 展望

综上所述,病理组学基于人工智能技术,将多种病理分析任务,如特征提取、量化病理诊断、疾病预后,有机地结合在一起,结合临床病理知识体系实现疾病诊疗、疾病预后以及生成病理报告的一体化病理分析。在病理组学的帮助下,病理形态学信息可以与基因分子信息、解剖影像和功能影像信息相互对比、相互融合,从而发现人眼不易察觉的细节和主观经验难以总结的规律,不断完善病理医师和数字病理诊断的知识体系,推动多组学交叉研究的产生,例如病理影像组学、病理基因组学,这样的交叉式组学更能发挥各方向内的优势,病理组学作为疾病诊断的金标准、影像组学的无创性使它应用更加广泛、基因组学用于分析分子分类的问题,这样的交叉研究不但能够提高疾病诊断的准确性、还能充分利用现有医疗资源、节省研究成本、推动医疗发展。

因此,人工智能时代的病理组学,不仅仅局限于病理切片的定量分析的基础性工作,而是在众多科学家的努力下正向更加自动化更加精准的方向发展。

参考文献:

[1] Gurcan M N, Boucheron L E, Can A, *et al.* Histopathological image analysis: a review [J]. *IEEE Rev Biomed Eng*, 2009,2:147 - 171.

[2] 刘炳宪, 谢菊元, 王焱辉, 等. 基于云计算的数字病理远程会诊及管理平台[J]. *中国卫生产业*, 2017,14(10):72 - 73.

[3] Al-Janabi S, Huisman A, Van Diest P J. Digital pathology: current status and future perspectives[J]. *Histopathology*, 2012,61(1):1 - 9.

[4] Chang H, Borowsky A, Spellman P, Parvin B. Classification of Tumor Histology via Morphometric Context[J]. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, 2013:2203 - 2210. doi: 10.1109/CVPR.2013.286.

[5] Kong J, Sertel O, Shimada H, *et al.* Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation[J]. *Pattern Recognit*, 2009,42(6):1080 - 1092.

[6] Tosun A B, Gunduzdemir C. Graph run-length matrices for histopathological image segmentation[J]. *IEEE Trans Med Imaging*, 2011,30(3):721 - 732.

[7] Xu Y, Mo T, Feng Q, *et al.* Deep learning of feature representation with multiple instance learning for medical image analysis^[C]// *IEEE International Conference on Acoustics, Speech and Signal Processing*. Florence: IEEE, 2014:1626 - 1630.

[8] Giresan D C, Giusti A, Gambardella L M, *et al.* Mitosis detection in breast cancer histology images with deep neural networks^[C]// *International Conference on Medical Image Computing and Computer-assisted Intervention*. Berlin: Springer, 2013:411 - 418.

[9] Clincial H S. Histopathology specimens: clincial, pathological and laboratory aspects[M]. New York: Springer Verlag, 2013:519 - 611.

[10] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives[J]. *IEEE Trans Pattern Anal Mach Intell*, 2013,35(8):1798 - 1828.

[11] Mairal J, Ponce J, Sapiro G, *et al.* Supervised dictionary learning^[C]//*Advances in neural information processing systems*. 2009:1033 - 1040.

[12] Turian J, Ratino L, Bengio Y. Word representations: a simple and general method for semi-supervised learning^[C]// *Proceedings of the Meeting of the Association for Computational Linguistics*, 2010. Sweden: Springer, 2010:384 - 394.

[13] Xing F, Xie Y, Yang L. An automatic learning-based framework for robust nucleus segmentation[J]. *IEEE Trans Med Imaging*, 2016,35(2):550 - 566.

[14] Chen T, Chefd'Hotel C. Deep learning based automatic immune cell detection for immuno-histochemistry images^[C]// *International Workshop on Machine Learning in Medical Imaging*. Boston: Springer, 2014:17 - 24.

[15] Song Y, Zhang L, Chen S, *et al.* Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning[J]. *IEEE Trans Biomed Eng*, 2015,62(10):2421 - 2433.

[16] Xu Y, Jia Z, Ai Y, *et al.* Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation^[C]// *IEEE International Conference on Acoustics*. Brisbane: IEEE, 2015:947 - 951.

[17] Sirinukunwattana K, Pluim J P W, Chen H, *et al.* Gland segmentation in colon histology images: The glas challenge contest[J].

Med Image Anal, 2017,35:489 – 502.

- [18] Chen H, Qi X, Yu L, *et al.* Dcan: Deep contour-aware networks for accurate gland segmentation^[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Washington: IEEE, 2016;2487 – 2496.
- [19] Xu Y, Li Y, Liu M, *et al.* Gland instance segmentation by deep multichannel side supervision^[C]// InInternational Conference on Medical Image Computing and Computer-Assisted Intervention 2016. Cham: Springer, 2106;496 – 504.
- [20] Wu B, Fan B, Xiao Q, *et al.* Multi-scale deep neural network for mitosis detection in breast cancer histological images[J]. Preprints 2017, 2017110063 (doi:10.20944/preprints201711.0063.v1).
- [21] Wang D, Khosla A, Gargeya R, *et al.* Deep learning for identifying metastatic breast cancer [J]. arXiv preprint arXiv:1606.05718, 2016.
- [22] Liu Y, Gadepalli K, Norouzi M, *et al.* Detecting cancer metastases on gigapixel pathology images[J]. arXiv preprint arXiv:1703.02442, 2017.
- [23] 解建军, 张仁亚. 乳腺浸润性导管癌患者重要的组织学预后预测[J]. 临床与实验病理学杂志, 2012,28(1):100.
- [24] Qin L X, Tang Z Y, Ma Z C, *et al.* P53 immunohistochemical scoring: an independent prognostic marker for patients after hepatocellular carcinoma resection[J]. World J Gastroenterol, 2002,8(3):459 – 463.
- [25] Liu J, Ma Q, Zhang M, *et al.* Alterations of TP53 are associated with a poor outcome for patients with hepatocellular carcinoma: evidence from a systematic review and meta-analysis [J]. Eur J Cancer, 2012,48(15):2328 – 2338.
- [26] 张雨涛, 周洪园, 唐云, 等. 地市级医院 292 例远程病理会诊病例分析[J]. 临床与实验病理学杂志, 2016,32(1):64 – 67.