·述　评·

可信病理人工智能：从理论到实践

周燕燕[1]，邓　杨[1]，包　骥[1]，步　宏[1,2]

四川大学华西医院 [1]临床病理研究所 [2]病理科，成都 610041

通信作者：包　骥，E-mail：baoji@scu.edu.cn

【摘要】人工智能正在融入病理学研究的各个领域，但在临床实践中却遇到了诸多挑战，包括因医疗数据隐私保护而形成“数据孤岛”，不利于人工智能模型的训练；现有人工智能模型缺乏可解释性，导致使用者无法理解而难以形成人机互动；人工智能模型对多模态数据利用不足，致使其预测效能难以进一步提升等。为解决上述难题，我们建议在现有病理人工智能研究中引入可信人工智能技术：（1）数据安全共享，在坚持数据保护的基础上打破数据孤岛，使用联邦学习的方法、仅调用数据训练的结果而不上传数据本身，在不影响数据安全的情况下大大增加可用于训练的数据量；（2）赋予人工智能可解释性，使用图神经网络技术模拟病理医生学习病理诊断的过程，使得模型本身具有可解释性；（3）多模态信息融合，使用知识图谱技术对更为多样和全面的数据来源进行整合并深入挖掘分析，获得更准确的模型。相信通过此三方面的工作，可信病理人工智能技术可使病理人工智能达到可控可靠和明确责任，从而促进病理人工智能的发展和临床应用。

【关键词】可信人工智能；病理；数据安全共享；可解释性；多模态信息融合

【中图分类号】R818.02；R36　　【文献标志码】A　　【文章编号】1674-9081(2022)04-0525-05

DOI：10.12290/xhyxzz.2022-0184

Trusted Artificial Intelligence for Pathology：From Theory to Practice

ZHOU Yanyan[1], DENG Yang[1], BAO Ji[1], BU Hong[1,2]

[1]Institute of Clinical Pathology, [2]Department of Pathology, West China Hospital, Sichuan University, Chengdu 610041, China

Corresponding author: BAO Ji, E-mail: baoji@scu.edu.cn

【**Abstract**】Artificial intelligence (AI) has gradually integrated into every aspect of pathology research. However, we also encounter some problems in the practical application of pathological artificial intelligence. 1. Research institutions attach importance to the protection of data privacy, which results in the emergence of data islands and is detrimental to our training of AI models. 2. The lack of interpretability of existing AI models leads to users' incomprehension and difficulty in human-computer interaction. 3. AI models make insufficient use of multi-modal data, making it difficult to further improve their predictive effectiveness. To address the above-mentioned challenges, we propose to introduce the latest technologies of trusted artificial intelligence (TAI) into existing research of pathological AI, which is embodied as the following: 1. Securely share data. We try to break data islands on the basis of adhering to data protection. We can use federated learning methods, only provide the results of data training without uploading the data itself, and greatly increase the amount of data that can be

基金项目：成都市新型产业技术研究院技术创新项目（2017-CY02-00026-GX）；四川大学华西医院临床研究孵化项目（20HXFH029）；四川大学华西医院学科卓越发展 1·3·5 工程项目（ZYGD18012）

引用本文：周燕燕，邓杨，包骥，等. 可信病理人工智能：从理论到实践［J］. 协和医学杂志，2022，13（4）：525-529. doi：10.12290/xhyxzz.2022-0184.

used for training without affecting the data security. 2. Give AI interpretability. The technology of graphic neural networks is used to simulate the process of pathologists' learning pathological diagnosis, making the model itself interpretable. 3. Fuse multimodal information. Use the technology of knowledge graph to integrate and deepen the analysis of more diverse and comprehensive data sources in order to derive more accurate models. Through the above three aspects, we can achieve reliable and controllable pathological AI and clear the responsibility through trusted pathological AI technology, so as to promote the development and clinical application of pathological AI.

【Key words】 trusted artificial intelligence; pathology; securely sharing data; interpretability; multi-mode information fusion

Funding: Technological Innovation Project of Chengdu New Industrial Technology Research Institute (2017-CY02-00026-GX); 1·3·5 Project for Disciplines of Excellence Clinical Research Incubation Project, West China Hospital, Sichuan University (20HXFH029); 1·3·5 Project for Disciplines of Excellence, West China Hospital (ZYGD18012)

Med J PUMCH, 2022,13(4):525-529

人工智能（artificial intelligence，AI）被认为是推动科学发展的重要驱动力，正在融入社会生活的各个方面[1-2]。在与现实结合应用过程中，由于存在数据隐私泄露、不可解释、决策失误以及责任无法界定等情况，AI 面临不被信任的危机，阻碍了其在实践中的进一步发展。为促进 AI 的实践应用，华东师范大学软件学院院长何积丰院士于 2017 年 11 月首次提出了可信人工智能（trusted artificial intelligence，TAI）的概念。2020 年，欧盟发布了《人工智能白皮书》，提出 AI“可信生态系统”，旨在落实 AI 应用的管理框架，促进 AI 的使用[3]。中国信息通信研究院联合京东探索研究院于 2021 年 7 月发布了国内首本《可信人工智能白皮书》，提出 TAI 是从技术和工程实践的角度，落实伦理治理，实现创新发展和风险治理的有效平衡，其具备如下五要素：数据保护、透明可释、多元包容、可控可靠和明确责任[4]。

随着 AI 的发展以及全视野数字图像（whole slide image，WSI）的出现，使用计算机辅助病理诊断逐渐成为现实。病理 AI 经过近几年的发展，已能够有效识别病理切片上的组织生物学特征，在肿瘤区域识别、组织学分级、预测分子分型等任务中均取得了一定成效[5]。但病理 AI 广泛应用于临床诊断尚未实现，TAI 的提出为解决数据安全共享、AI 可解释性以及多模态信息融合问题提供了可行技术方案，将进一步促进 AI 在临床医疗及病理中的推广应用。

1 数据安全共享

目前，AI 已展示出在图像识别和大数据处理方面的优势，但 AI 技术尤其是深度学习技术对数据具有很强的依赖性，需要大量数据用于模型训练以得到高性能的系统。单一医疗机构的病理数据常常无法达到训练模型的数据要求，成立中心数据库、扩大数据量用于模型训练是理想的解决方案[6]。然而，由于隐私数据保护法规的颁布（如欧盟《通用数据保护条例》[7]及中国香港《人类数据隐私条例》[8]等）以及人们数据保护意识的提高，隐私保护逐渐受到重视，将不同机构间数据集合并成立数据中心的难度较大，“数据孤岛”现象不断浮现，导致 AI 在病理领域的应用受阻[9-10]。

为了在保护数据隐私安全的条件下解决“数据孤岛”问题，技术人员建议引入联邦学习（federated learning，FL）技术。FL 是一种多分布式联合学习技术，在数据有限共享的前提下，通过中心数据库传递系统参数，在多个数据库间开展学习，力求获得高精度的系统[11]。FL 并非全新的算法，其已广泛应用于放射学图像分析领域，解决影像数据分散的问题，并取得了不俗成效[12]。FL 在病理领域起步较放射领域晚，但也在逐步开展应用，Lu 等[13]使用 FL 算法成功训练了一套基于 WSI 预测生存周期的系统，与单一数据集训练系统相比，该系统具有更高的性能。

然而在实际应用中，由于各数据中心病理切片的试剂和制作工艺不同、数据标准不统一，导致切片质量存在较大差异，直接使用此类数据进行 FL 训练将会影响整个系统的性能，因此需采用标准化数据集进行训练。为获得标准化数据，在病理制片方面，建议通过医联体及医共体制订标准制片流程，以减小切片受试剂和染色步骤的影响；同时以机器自动化染色代替手工操作，从而减少手工染色误差。在计算机技术方面，可对数据集进行预处理，使数据在 FL 训练前达到较高的均一化，以进一步提高数据标准化率。总之，FL 在病理 AI 领域的应用仍有较大空间，未来将

对病理 AI 的发展提供极大帮助。

2 AI 模型可解释性

机器学习是常用的 AI 技术之一，但由于机器学习尤其是深度学习算法内部架构过于复杂，技术人员难以检测到模型内部的偏差，且系统决策难以追溯到输入特征，医生与 AI 缺乏有效交互，导致医生对 AI 并不信任，影响了其在医疗领域的应用，因此需增强 AI 模型的可解释性。深度学习解释的方法种类很多，可简单分为系统自带解释属性的事前解释和在系统决策后加入事后解释模型的事后解释 2 种方式[14]。

目前病理领域大多采用标注的数据直接训练算法模型，得到数字病理系统，然后置入可解释模型，解释决策的原因，属于事后解释。事后解释能够可视化输入数据特征与决策之间的关系，常用于标记 AI 决策依据的特征，帮助人类理解 AI 系统。通用的解释模型有反卷积网络（deconvolution）、积分梯度（integrated gradients）、梯度加权类激活映射（gradient-weighted class activation mapping，Grad-CAM）以及模型无关的局部可解析性算法（local interpretable model agnostic explanation，LIME）等，已在研究中广泛应用[15]。例如，Yu 等[16]使用卷积神经网络（convolutional neural network，CNN）训练系统识别肺鳞癌和腺癌，并使用 Grad-CAM 模型解释决策，根据显示区域重要性的热力图来看，AI 的决策特征来源于正确的鳞癌和腺癌组织区域。Sousa 等[17]使用 LIME 解释 CNN 模型如何从淋巴结图像中判断肿瘤细胞，发现 CNN 判断依据的图像特征与专家诊断依据的图像特征基本一致。

但事后解释模型多基于输入及输出关系得出类似解析，虽可对 AI 系统的解释提供参考，但解释结果未必真实[18-19]，因此还需从技术上对模型进行完善。Li 等[20]设计了一种基于 Shapley Value 的特征重要性估算解释模型，在脑 CT 图像中用于确定自闭症分类模型中不同脑区的重要性。由于对于解释结果存疑，该团队继而基于 DeepSHAP 设计了一种 Dist DeepSHAP 解释方法，在生成重要性图像的同时生成对应的不确定图像，通过重要性图像确定模型决策的特征，再通过不确定图像排除不确定性高的区域，从而获得模型决策与图像特征的关联性[21]。

由于病理医生关注图像特征与决策之间的关系，根据图像特征构建具有可解释能力的系统亦是可靠的办法。研究者根据病理 AI 实际情况提出，可通过改善传统训练模式、开发 AI 与病理结合的新模式以及使用新的算法达到提升可解释性的目标。Sarder[22]在模型数据标注和训练中，从分割特征完整的信息单元提取定量特征以区分信息单元，再对整体进行信息聚合，得到了便于解释的模型。Hegde 等[23]开发了一种基于深度学习的组织病理学图像反向图像搜索工具 SMILY，对于输入的图像，模型输出相似的图像及信息，从而回答了何种图像特征决定模型决策的问题。

图神经网络（graph neural network，GNN）是一种用于处理图数据的神经网络结构，其特点是可以捕获实例之间的相互依赖关系并进行分析，故模型本身具有可解释性。对于医学图像而言，可将图像拆分成特征进行结构学习，通过面关联特征之间的关系，对模型作出解释。因其学习和建模过程类似于病理医生学习病理图像诊断的过程，是一种有潜力的可解释性算法，GNN 的预测过程如图 1 所示。GNN 能关联决策与图像特征之间的关系，与传统神经网络相比，具有更高的可解释性[24-25]。GNN 在病理领域的应用目前仍较少，本研究团队正在开展 GNN 方面的研究，提出以甲状腺细胞病理为基础，采用 GNN 技术进行特征提取。利用 GNN 能够可视化地提取局部节点和节点间的空间关系特征，解决当前 CNN 缺乏空间关系以及可解释性的问题。

3 多模态信息融合

病理诊断需基于临床资料、诊断意见等文本数据，病理、影像、超声等图像数据，分子检测等组学数据多种信息，而目前 AI 的预测往往仅基于病理图像，AI 模型对多模态数据利用不足，导致其预测效能难以进一步提升。结合多种整合信息设计的 AI 模型，单一特征失误对决策的影响更小，决策结果更加可靠，有利于 AI 在病理中的应用。

如何整合来自不同维度的信息呢？知识图谱（knowledge graph，KG）的提出成为解决这一难题的突破口。KG 本质上是一种语义网络，由节点（实体）和边（实体之间的关系）组成，在 KG 中，可以很好地处理各种维度的信息如图像、文本、诊断数据、描述信息等，并作出决策。若在病理 AI 中引入 KG，能有效整合病理诊断中不同来源的数据，结合多种信息作出决策，提升病理 AI 的效能[26]。近年来，研究者利用已有的临床知识（如医学教材、诊疗指南等）进行结构化表示构建 KG 系统，开发医疗

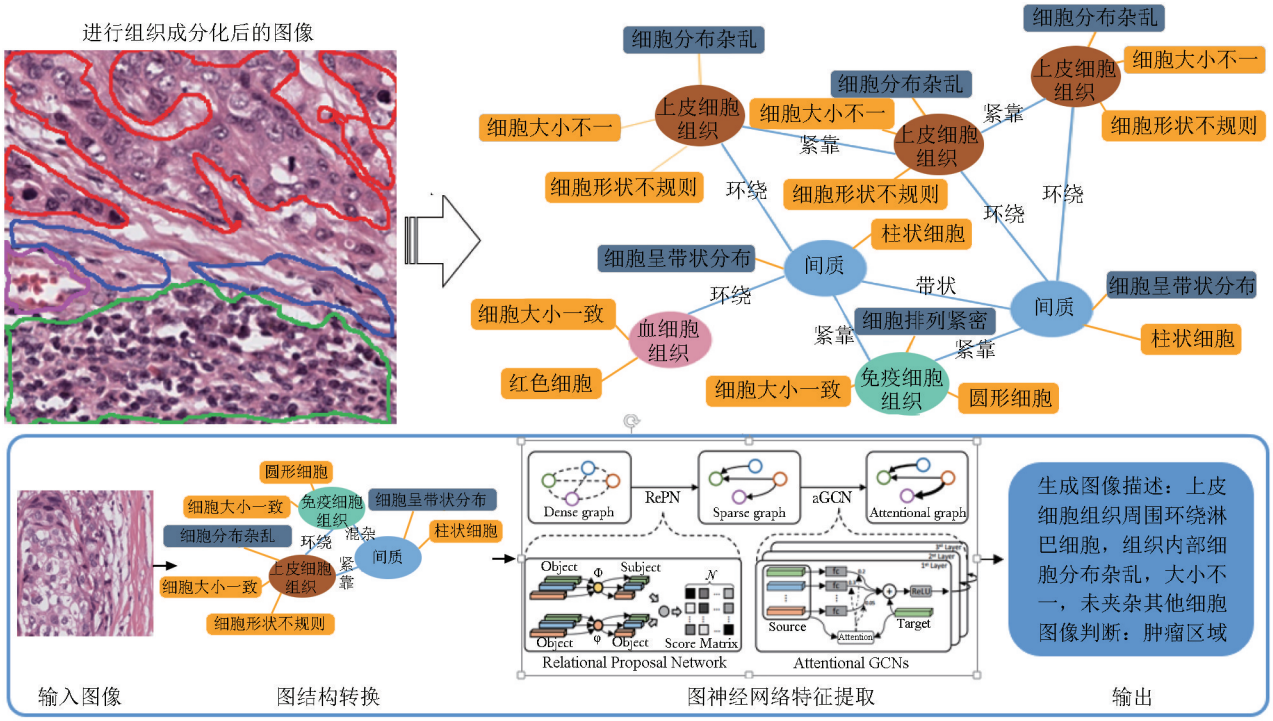

图 1 图神经网络预测过程

语义搜索引擎、医疗问答系统、医疗决策支持系统等，从而在临床环节有效进行辅助决策，例如国内百度的“灵医”、阿里巴巴的“Doctor You”、腾讯的“觅影”、国外的“沃森医生”等。在病理学领域，KG 系统也正在构建中，如对早期乳腺癌进行预后评估的“Adjuvant Online”系统等[27]。但目前此类系统主要基于文本信息构建，还需克服图像数据处理等困难，相信随着病理 AI 与 KG 的共同发展，多模态信息融合病理 AI 的辅助病理诊断将很快能够实现。

4 小结与展望

随着病理切片全数字化的实现、更多机器学习方法的出现以及计算机算力的提升，使用计算机辅助病理诊断将逐渐成为现实。但目前病理 AI 仍处于研究阶段，广泛应用于临床诊断尚未实现，未来可通过 TAI 等新技术手段提升病理 AI 的系统性能，促进其临床应用。现阶段，建议通过制订病理制片标准和规范以提高切片质量，并通过 FL 技术解决“数据孤岛”问题；使用各种解释方法以及 GNN 提升 AI 模型的可解释性；使用 KG 研发功能全面的 AI 系统，从技术上达到 TAI，配合诊断过程的可视化与交互性，使病理诊断结果更加可靠可控；使用 KG 以及机器学习模型搭建知识库，助力缺乏经验的病理医生快速成长。此外，在 AI 实践应用的过程中，仍需完善相关规范，从国家层面推进 AI 在病理中的应用。相信在不久的未来，TAI 将极大促进 AI 在病理领域的落地实践和技术推广。

作者贡献：周燕燕负责查阅文献、撰写论文；邓杨负责整理文献和论文修订；包骥、步宏负责论文构思及终稿审校。

利益冲突：所有作者均声明不存在利益冲突

参 考 文 献

[1] Li BH, Hou BC, Yu WT, et al. Applications of artificial intelligence in intelligent manufacturing: a review [J]. Front Inform Technol Electron Eng, 2017, 18: 86-96.

[2] Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine [J]. Nat Med, 2022, 28: 31-38.

[3] European Commission. White Paper On Artificial Intelligence-A European approach to excellence and trust [EB/OL]. (2020-02-19)[2022-04-05]. https://ec. europa. eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en. pdf.

[4] 中国信息通信研究院. 可信人工智能白皮书 [EB/OL]. (2021-07-09)[2022-04-05]. http://www. caict. ac. cn/k-

xyj/qwfb/bps/202107/P020210709319866413974. pdf.

[5] Parwani A. Whole Slide Imaging [M]. Switzerland: Springer Nature Switzerland AG, 2022: 223-236.

[6] Willemink MJ, Koszek WA, Hardell C, et al. Preparing Medical Imaging Data for Machine Learning [J]. Radiology, 2020, 295: 4-15.

[7] European Commission. General Data Protection Regulation [EB/OL]. (2016-04-27)[2022-04-05].https://gdpr. eu/article-1-subject-matter-and-objectives-overview/.

[8] Law Reform Commission. Hong Kong Person Date Privacy Ordinance [EB/OL]. (2012-10-01)[2022-04-05]. https://www. pcpd. org. hk/english/data_privacy_law/ordinance_at_a_Glance/ordinance.html.

[9] Kaissis GA, Makowski MR, Rückert D, et al. Secure, privacy-preserving and federated machine learning in medical imaging [J]. Nat Machine Intel, 2020, 2: 305-311.

[10] Zhou SK, Greenspan H, Davatzikos C, et al. A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises [J]. Proc IEEE Inst Electr Electron Eng, 2021, 109: 820-838.

[11] 刘再毅, 石镇维, 梁长虹. 推进联邦学习技术在医学影像人工智能中的应用 [J]. 中华医学杂志, 2022, 102: 318-320.
Liu ZY, Shi ZW, Liang CH. Promoting the application of federated learning in medical imaging artificial intelligence [J]. Zhonghua Yixue Zazhi, 2022, 102: 318-320.

[12] Yang D, Xu Z, Li WQ, et al. Federated Semi-Supervised Learning for COVID Region Segmentation in Chest CT using Multi-National Data from China, Italy, Japan [J]. Med Image Anal, 2021, 70: e101992.

[13] Lu MY, Chen RJ, Kong DH, et al. Federated learning for computational pathology on gigapixel whole slide images [J]. Med Image Anal, 2022, 76: e102298.

[14] Du MN, Liu NH, Hu X. Techniques for Interpretable Machine Learning [J]. Commun ACM, 2020, 63: 68-77.

[15] Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization [J]. Int J Comput Vis, 2020, 128: 336-359.

[16] Yu K, Wang F, Berry GJ, et al. Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks [J]. J Am Med Inform Assoc, 2020, 27: 757-769.

[17] Sousa I, Vellasco M, Silva E. Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases [J]. Sensors, 2019, 19: 1-18.

[18] Saporta A, Gui XT, Agrawal A, et al. Deep learning saliency maps do not accurately highlight diagnostically relevant regions for medical image interpretation [J]. MedRxiv, 2021. https://doi. org/10. 1101/2021. 02. 28. 21252634.

[19] Ehsan U, Passi S, Liao QV, et al. The Who in Explainable AI: How AI Background Shapes Perceptions of AI Explanations [J]. Arxiv, 2021. https://arxiv. org/abs/2107. 13509.

[20] Li X, Dvornek NC, Zhou Y, et al. Efficient Interpretation of Deep Learning Models Using Graph Structure and Cooperative Game Theory: Application to ASD Biomarker Discovery [J]. Inf Process Med Imaging, 2019, 11492: 718-730.

[21] Li X, Zhou Y, Dvornek NC, et al. Efficient Shapley Explanation for Features Importance Estimation Under Uncertainty [J]. Med Image Comput Assist Interv, 2020, 12261: 792-801.

[22] Sarder SP. From What to Why, the Growing Need for a Focus Shift Toward Explainability of AI in Digital Pathology [J]. Front Physiol, 2022, 12: e821217.

[23] Hegde N, Hipp JD, Liu Y, et al. Similar Image Search for Histopathology: SMILY [J]. NPJ Digit Med, 2019, 2: 56-65.

[24] Li X, Duncan J. BrainGNN: Interpretable Brain Graph Neural Network for fMRI Analysis [J]. Med Image Anal, 2021, 74, e102233.

[25] Li X, Zhou Y, Dvornek NC, et al. Pooling Regularized Graph Neural Network for fMRI Biomarker Analysis [J]. Med Image Comput Assist Interv, 2020, 12267: 625-635.

[26] Amit S. Introducing the knowledge graph [R]. America: Official Blog of Google, 2012.

[27] 崔洁. 面向乳腺肿瘤诊断的知识图谱及辅助决策研究 [D]. 上海: 东华大学, 2018.

（收稿：2022-04-06 录用：2022-05-26 在线：2022-06-10）
（本文编辑：李　娜）