

DOI:10.13288/j.11-2166/r.2024.02.007

“数据-知识”双驱动模式的辨证决策方法探索 ——以胃癌前状态为例

徐伟超^{1,2,5}, 杜艳茹^{2,5}, 郎晓猛^{2,5}, 姜莹莹^{2,5}, 贾文文^{2,5}, 康欣^{2,5},
郭硕³, 张坤⁴, 苏春芝², 田军彪², 魏晓娜², 杨倩^{1,2,5}✉

1. 河北中医药大学研究生学院, 河北省石家庄市桥西区新石南路326号, 050091; 2. 河北省中医院; 3. 河北医科大学第四医院; 4. 河北科技大学信息科学与工程学院; 5. 河北省浊毒证重点实验室

[摘要] 数据分析模型可辅助中医药经验的传承和临床诊疗, 以胃癌前状态为例探索“数据-知识”双驱动模式构建的可能性。数据驱动是围绕数据分析开展临床决策, 其辨证决策研究依托隐结构模型、部分可观察马尔可夫决策过程以确认疾病病因、证候要素、病机演变和辨证方案; 知识驱动是利用数据和信息推动决策和行动的过程, 其辨证决策研究依托卷积神经网络以提高局部辨病辨证准确性。“数据-知识”双驱动模式可弥补单一驱动数字模拟准确性的短板, 达到局部辨病与宏观辨证的平衡。在既往研究基础上探索胃癌前状态病证诊断辅助决策平台的构建方法, 认为可通过机器与算法辅助来延伸医生诊断和决策能力, 整合相关研究方法并获取胃癌前状态中医辨证和内镜病理诊断预测核心特征, 探索疾病中医辨证内镜病理识别要素, 为数据分析前沿技术在中医辨证智能化领域的深入研究和创新应用提供思路。

[关键词] 胃癌前状态; 数据分析模型; 数据驱动; 知识驱动; 辨证决策

近年来随着计算机工程技术与数据算法能力的不断发展, 数据分析模型广泛应用于中医经验传承领域^[1-2], 成为中医药与信息科学交叉融合的标志和中医药现代化的重要体现形式之一^[3]。数据分析模型可对较多数据进行分析与处理, 挖掘数据内的隐性规律和隐藏特征, 目前运用其探究临床观察性数据中的显性与隐性诊疗策略已成为中医经验传承与归纳的重要手段, 如在探究中医处方中的“症-证-方-药”潜在相关关系、中医临床实际诊疗路径中辨证论治规律等方面均有较好的应用^[4]。然而, 数据分析模型纷繁复杂, 精准性仍有不足, 无法满足现阶段中医临床诊疗对客观证据的更高要求, 亦不能满足中医临床持续创新和高质量发展需求。如何选择并改进合适的方法以满足相应知识挖掘的需要, 最终实现规律的准确发现、优化使用,

成为当前利用数据分析模型阐明中医学辨证决策路径的核心问题。

胃癌是我国第一大消化道恶性肿瘤^[5]。胃黏膜萎缩和肠上皮化生是胃癌发生的独立危险因素^[6], 并已被归类为胃癌前状态^[7], 其中慢性萎缩性胃炎是胃癌前状态的典型代表。目前, 延缓或阻滞疾病进展、降低癌变风险、改善临床症状已成为胃癌前状态的治疗目标。中医辨证论治可缓解胃癌前状态的消化不良症状, 有助于改善胃黏膜病理状况^[8]。而数据分析模型在中医辨证研究领域的广泛应用^[9], 有效推动了其在胃癌前状态相关疾病中的进一步深入应用。本文以胃癌前状态为例, 分别从数据驱动和知识驱动两方面探讨中医辨证决策, 进而探索“数据-知识”双驱动模式构建的可能性, 为中医数据分析模型的构建及辨证决策研究提供参考。

1 数据驱动辨证决策研究

中医临床诊疗形成了以个体化、区域化、流派化为主要代表特征的诸多有效辨证论治方案, 构建中医数据分析模型, 将名医经验以数据形式进行转

基金项目: 国家自然科学基金(82205314); 国家中医药管理局全国老中医药专家学术经验继承工作项目; 河北省重点研发计划(21377724D, 23377701D); 河北省“三三三人才工程”资助项目(C20231017)

✉ 通讯作者: hbzyzz99@126.com

换录入、提取分析及智能学习是名医经验传承现代化的重要研究方向^[10]。在数据分析模型中模型选用和建立是数据分析的核心环节。

数据驱动是基于数据分析,通过数据采集和挖掘方法凝炼规律,围绕数据分析开展临床决策。隐结构模型(latent structure model, LSM)是数据分析的经典方法,可量化诠释隐性病机变量,其与序列决策的结合有望为胃癌前状态辨证模拟提供优化方案。LSM 早已广泛应用于中医辨证领域^[11-12],其有助于建立中医辨证客观统计基础,并可将名老中医的辨证思维与策略进行可视化呈现^[13]。本团队前期研究^[14]发现,京津冀地区慢性萎缩性胃炎患者的基本病机被 LSM 诠释为“浊毒内蕴兼伤阴”,通过综合聚类得出肝胃气滞证、脾胃虚弱证、胃阴不足证等 6 种常见证型。LSM 的综合聚类及后期诠释可量化复杂症状,为中医证候的客观判断提供一定参考。上述研究结果提示,LSM 能解决中医辨证论治萎缩性胃炎的症状-证候-药物聚类问题,还可量化中医隐性知识,分析不同变量在同一组数据中所处的主、次、兼地位。因此,LSM 有望在胃癌前状态的病机诠释中发挥中医辨证思维的客观数据量化优势,具有较强的主证决策提炼和兼证序列参考价值。

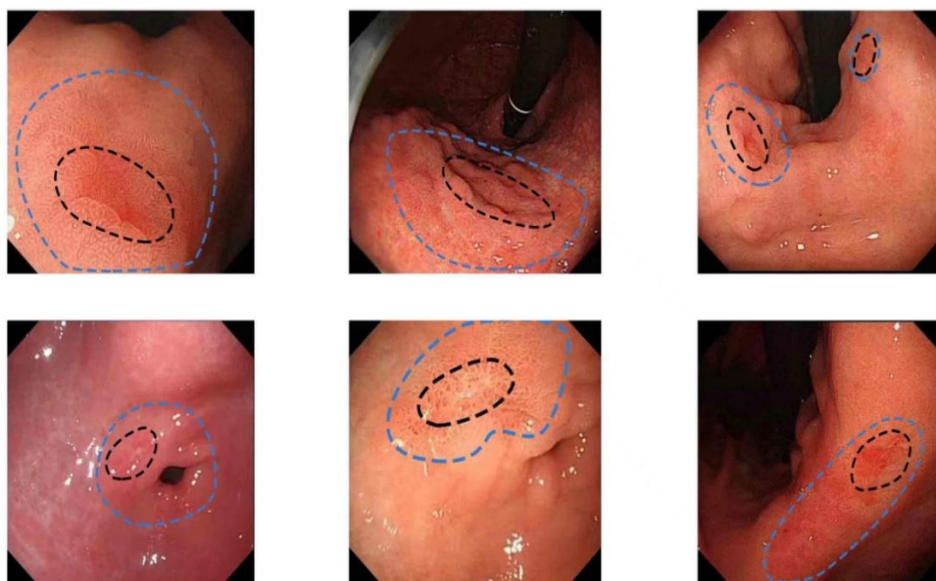
LSM 在具体应用时亦有局限性,即对于胃癌前状态为代表的这类病机复杂的疾病,辨证决策是受三因制宜影响的多次动态诊疗过程,而 LSM 只能提供一次决策^[15]。因此,结合其他决策方法,获取最优序列决策并开展方案优化是 LSM 在中医临床多变量研究应用中的重要科学问题。相关方法中,部分可观察马尔可夫决策过程(partially observable markov decision-making process, POMDP)是对系统状态观察不完全的马尔可夫决策模型,该模型可以精准地选取最优辨证方案,尤其适用于医疗决策序列问题^[16]。目前该模型已应用于慢性肾脏病^[17]、中风恢复期^[18]等中医领域研究中。有研究发现, POMDP 其他变型算法将有助于证候的动态性研究^[19]。本团队在长期实践中发现,通过 LSM 提取疾病-证候相互信息,对隐变量及显变量间关联度进行量化,结合 POMDP 挖掘和优化中医辨证方案,可为显化中医临床数据奠定部分可视化图像基础。因此, LSM 和 POMDP 相结合可更好地发现疾病病因、证候要素、病机演变规律和辨证方案,为尚无诊疗共识的病机辨证提供数据序列优化方案。

局部辨病与宏观辨证如何有机结合,是中医临床研究持续聚焦的内在需求和核心科学问题之一。有研究基于慢性萎缩性胃炎中医辨证提出“舌镜互参”^[20],即将内镜作为中医临床望诊的延伸。另有研究发现,慢性萎缩性胃炎中医证型与胃镜象存在一定相关性^[21];慢性胃炎合并幽门螺杆菌感染患者的胃镜象中脾胃虚弱证多见黏膜白相,胃络瘀阻证黏膜多见颗粒样或结节样改变^[22]。就胃癌前状态而言,其诊断标准基于内镜和病理结果,但 LSM 和 POMDP 尚不具备图像处理能力,导致相关数字模拟的准确性不足;同时胃癌前状态作为病灶集中于局部的疾病阶段,需与整体证候结合,经提取优化主证信息才可辨证分类,导致病灶范围精准性和计算负担简化性成为亟待解决的技术矛盾。上述难点的解决需要借助智能计算与数理医学等学科的技术创新进行探索。

2 知识驱动辨证决策研究

知识驱动是利用数据和信息推动决策和行动的过程,涉及知识的获取、处理和应用,包含部分知识构建能力、知识联想能力、抽象分析能力等。卷积神经网络(convolutional neural network, CNN)作为知识驱动的典型应用场景和医学图像识别技术的前沿工具,在内镜和病理领域已为胃癌前状态辨证模拟提供了预测尝试^[23]。胃癌前状态部分患者内镜下的病理图像仅表现为细微形态学改变,其诊断受医师经验、设备条件、患者状态、样本误差等因素影响较大,故针对组织学标本开展病理学检查对消化系统疾病诊治至关重要。但病理检查结果也受标本来源特点、病理医师主观倾向等因素的制约^[24],从而影响诊断结论的一致性。本团队以同一病例为例,比较两个来自石家庄市三级甲等医院内镜主治医师的内镜下病理观察结果,发现两位医师所关注的病变范围存在差异,见图 1;两位医师针对同一个病理图像的诊断结论存在差异,见图 2。由此可见,医师的主观因素可能影响胃癌前状态的客观诊断。因此,借助基于机器学习的智能化图像识别技术,对胃癌前状态内镜和病理图像进行特征学习、分类和预测,具有可期的研究空间和实践前景。

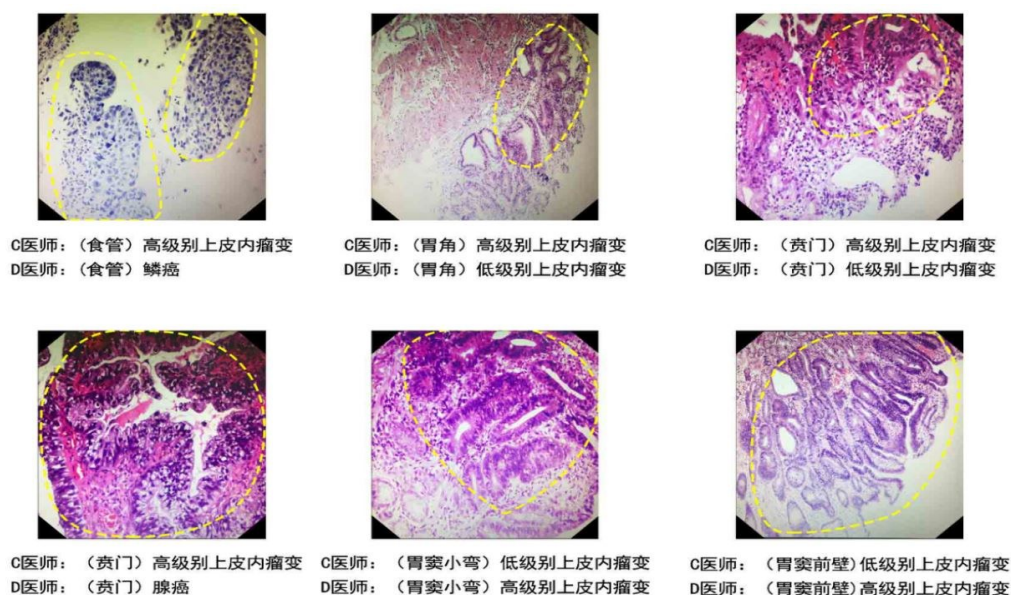
CNN 作为深层网络算法的典型代表,被认为是目前图像识别领域中性能最好的模型^[25],其与影像诊断的结合已成为智能医疗的研究热点。中医领域已开展借助 CNN 围绕脉搏波^[26]和舌苔特征^[27]



注：蓝色边框为内镜医师 A 标注的病变范围；黑色边框为内镜医师 B 标注的病变范围。

图 1 两位医师对 6 位胃部疾病患者内镜下病变范围判定结果比较

Figure 1 Comparison of the results of endoscopic determination of the extent of lesions in six patients with gastric diseases by two physicians



注：黄色边框为病变部位。

图 2 两位医师对 6 位食管或胃部疾病患者病理判定结果比较

Figure 2 Comparison of the pathological results of six patients with oesophageal or gastric diseases by two physicians

的相关研究。在消化系统疾病领域，有研究发现 CNN 可通过采集胃镜图像信息精准诊断食管静脉曲张和胃静脉曲张，准确率分别为 97.00% 和 92.00%，并可预测破裂风险^[28]。本团队构建 CNN 系统用于诊断胃癌，用时 42 s 分析 1091 张胃镜测试图像，结果准确率达 91.2%，显示出较强的临床诊断能力^[29]。本团队还借助 CNN 系统对萎缩性胃炎与胃癌的内镜图像进行研究^[30]，经网络搭建和

训练后 CNN 可将萎缩性胃炎与胃癌区别分类，准确率在第 29 轮时达到最大值 98%，且在 42 轮以后也基本稳定在 98%。需要注意的是，疾病的类型、浸润深度以及在胃中的位置是 CNN 分析与识别过程中的重要标记特征，影响着分类的速度与准确度。综上，可利用 CNN 技术开展胃癌前状态内镜病理图像与证候数据关联研究，提高辨病辨证准确性，这将具有创新意义和潜在临床价值。

3 “数据-知识”双驱动模式的探索

本团队以胃癌前状态的典型代表慢性萎缩性胃炎为例,运用“数据-知识”双驱动模式,尝试构建慢性萎缩性胃炎病证诊断辅助决策平台,通过机器与算法辅助,延伸临床医师诊断和决策能力,实现多维体征信息的获取、分析与整合。该平台一方面可辅助临床诊断,尤其是为诊疗水平相对落后地区提供决策支持;另一方面可为远程医疗提供方案。基于该平台,下一步拟实现患者健康自我检测、自我管理、实时监控与及时干预。与此同时,在平台模块下外接链接开展机器学习“萎缩性胃炎-非萎缩性胃炎,幽门螺杆菌感染-非幽门螺杆菌感染”内镜图片及病理图片的识别与预测方法。旨在回答:1)“数据-知识”双驱动能否实现既“算得快”又“算得准”;2)如果可以实现,后续研究又将如何用数据驱动实现规律发现,用知识驱动实现治疗决策。

基于上述思路,本团队在胃癌前状态患者临床信息库基础上,借助孔明灯隐结构分析软件、MATLAB 软件、CNN 系统等工具,结合贝叶斯信息准则评分、模拟临证动态建模等方法,运用因子分析、关联规则等算法选取显变量子集进行归纳提取和挖掘优化分析,拟构建掩膜区域卷积神经网络(mask regional convolutional neural network, Mask-RCNN)获取内镜病理图像高维特征,运用轻量级卷积神经网络(lightweight convolutional neural network, LCNN)优化网络参数,最终获取并确认胃癌前状态的病因病机、证候分布和内镜病理图像要素特征,以期系统揭示“数据-知识”双驱动模式在胃癌前状态辨证决策模拟中的诠释、优化及预测作用。

4 小结

综上,“数据-知识”双驱动模式的辨证决策方法探索立足于中医学、病理学与信息科学的多学科交叉,可为数据分析前沿技术在中医辨证智能化和名医经验传承现代化的深入研究与创新应用提供有益探索和尝试。然而,从理论层面,传统模型构建方法会产生大量冗余信息,难以获取精准目标(病灶)位置,从而影响病灶检测准确率;从操作层面,临床普遍缺乏专业计算设备,计算能力无法满足传统模型训练需求,算法执行效率低。未来如何更加精准地获取病灶范围,同时减轻计算负担,成为数据分析模型研究在医学领域应用中亟待解决的科学和现实问题,也是今后努力的方向。

参考文献

- [1]王子旭,王永炎.基于叙事医学视角的名老中医临床经验研究的思考[J].中医杂志,2022,63(22):2118-2121.
- [2]于林童,曲文白,余新波,等.数据挖掘方法在名老中医用药规律研究中的应用现状[J].中医杂志,2017,58(10):886-888,900.
- [3]程海波,张磊,付勇,等.2023 年度中医药重大科学问题、工程技术难题和产业技术问题[J].中医杂志,2023,64(14):1405-1421.
- [4]谷晓红,于河,卜德超,等.应用混合方法构建名老中医道术传承研究与推广范式[J].中医杂志,2022,63(15):1406-1411.
- [5]何敏均,季璐均,连力,等.2000-2019 年中国早发型胃癌和晚发型胃癌流行病学趋势分析[J].中华流行病学杂志,2023,44(8):1198-1202.
- [6]PIMENTEL-NUNES P, LIBANIO D, MARCOS-PINTO R, et al. Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade Portuguesa de Endoscopia Digestiva (SPED) guideline update 2019[J]. Endoscopy, 2019, 51(4): 365-388.
- [7]国家消化系统疾病临床医学研究中心(上海),国家消化道早癌防治中心联盟,中华医学会消化病学分会幽门螺杆菌学组,等.中国胃黏膜癌前状态和癌前病变的处理策略专家共识(2020 年)[J].中华消化杂志,2020,40(11):731-741.
- [8]唐旭东,吴皓萌.基于消化系统疾病现代研究的脾胃理论传承与创新[J].中医杂志,2022,63(2):101-105.
- [9]徐佳君,罗志明,赵文,等.基于人工智能算法的中医状态辨识规则[J].中医杂志,2020,61(3):204-208.
- [10]于明坤,于河,王迪,等.目标值法在名老中医经验传承临床研究中的方法学价值和设计要点[J].中医杂志,2020,61(1):36-41.
- [11]王天芳,张连文,赵燕,等.隐结构模型及其在中医证候研究中的应用[J].北京中医药大学学报,2009,32(8):519-526.
- [12]翟优,段飞,许玉龙,等.约束隐结构研究冠心病痰湿证的量化辨证规则[J].世界科学技术-中医药现代化,2021,23(7):2247-2252.
- [13]张东,冯贞贞,王露,等.基于隐结构结合关联规则的中晚期肺癌方药规律分析[J].中国中药杂志,2022,47(1):235-243.
- [14]徐伟超,许玉龙,张增博,等.基于隐结构模型的名老中医药专家辨证论治京津冀地区慢性萎缩性胃炎用药规律研究[J].中国中药杂志,2020,45(7):1691-1697.
- [15]许玉龙,王忠义,吕雅丽,等.中医药数据隐结构分析平台:孔明灯网络版[J].世界科学技术-中医药现代

- 化, 2020, 22(7): 2224-2232.
- [16] 张恒艳, 叶建州, 周也, 等. 灰色马尔可夫模型在中医证候预测中的应用进展[J]. 辽宁中医杂志, 2019, 46(1): 6-9.
- [17] 杨敏杰. 基于灰色马尔可夫的气阴两虚型慢性肾脏病中医疗效预测模型的设计与构建[D]. 北京: 北京中医药大学, 2018.
- [18] 杨慧婷, 蒋金兰, 金林珍, 等. 针灸治疗中风恢复期临床预测模型的建立和分析[J]. 中国针灸, 2021, 41(8): 855-860.
- [19] 杨敏杰, 周晟, 孟庆刚. 基于马尔可夫链的中医药疗效预测模型构建与探索[J]. 中华中医药学刊, 2018, 36(7): 1594-1597.
- [20] 王丽华, 单兆伟. 单兆伟慢性萎缩性胃炎诊疗经验浅探[J]. 辽宁中医杂志, 2015, 42(6): 1194-1195.
- [21] 吴丹. 慢性萎缩性胃炎胃镜象与证素特点的相关性研究[D]. 福州: 福建中医药大学, 2018.
- [22] 李湘蓝. 慢性胃炎合并幽门螺杆菌感染胃镜象与中医证型分布的关系研究[D]. 南京: 南京中医药大学, 2021.
- [23] YACOB YM, ALQURAN H, MUSTAFA WA, et al. H. pylori related atrophic gastritis detection using enhanced convolution neural network (CNN) learner[J]. Diagnostics (Basel), 2023, 13(3): 336.
- [24] 汪鹏, 谢静, 王雷, 等. 中国消化内镜活检与病理学检查规范专家共识(草案)[J]. 胃肠病学, 2014, 19(9): 549-553.
- [25] ZHAO T, HU Y, VALSDOTTIR LR, et al. Identifying drug-target interactions based on graph convolutional network and deep neural network [J]. Brief Bioinform, 2021, 22(2): 2141-2150.
- [26] 张选, 胡晓娟. 基于 GoogLeNet 和 ResNet 的深度融合神经网络在脉搏波识别中的应用[J]. 计算机系统应用, 2019, 28(10): 15-26.
- [27] 肖庆新, 张菁, 张辉, 等. 基于轻型卷积神经网络的舌苔颜色分类方法[J]. 测控技术, 2019, 38(3): 26-31.
- [28] CHEN M, WANG J, XIAO Y, et al. Automated and real-time validation of gastroesophageal varices under esophagogastroduodenoscopy using a deep convolutional neural network: a multicenter retrospective study (with video) [J]. Gastrointest Endosc, 2021, 93(2): 422-432, e3.
- [29] ZHANG L, ZHANG Y, WANG L, et al. Diagnosis of gastric lesions through a deep convolutional neural network[J]. Dig Endosc, 2021, 33(5): 788-796.
- [30] 徐伟超, 李博林, 许亚培, 等. 基于快速区域卷积神经网络萎缩性胃炎-胃癌胃镜图像自动识别模型的建立及临床测试[J]. 世界科学技术-中医药现代化, 2021, 23(9): 3274-3280.

Exploration of Decision-Making Methods Based on Syndrome Differentiation by “Data-Knowledge” Dual-Driven Models: A Case Study of Gastric Precancerous State

XU Weichao^{1,2,5}, DU Yanru^{2,5}, LANG Xiaomeng^{2,5}, LOU Yingying^{2,5}, JIA Wenwen^{2,5}, KANG Xin^{2,5}, GUO Shuo³, ZHANG Kun⁴, SU Chunzhi², TIAN Junbiao², WEI Xiaona², YANG Qian^{1,2,5}

1. Graduate School, Hebei University of Chinese Medicine, Shijiazhuang, 050091; 2. Hebei Provincial Hospital of Traditional Chinese Medicine; 3. The Fourth Hospital of Hebei Medical University; 4. School of Information Science and Engineering, Hebei University of Science and Technology; 5. Key Laboratory of Turbidity Toxin Syndrome (Hebei)

ABSTRACT Data analysis models may assist the transmission of traditional Chinese medicine (TCM) experience and clinical diagnosis and treatment, and the possibility of constructing a “data-knowledge” dual-drive model was explored by taking gastric precancerous state as an example. Data-driven is to make clinical decisions around data analysis, and its syndrome-differentiation decision-making research relies on hidden structural models and partially observable Markov decision-making processes to identify the etiology of diseases, syndrome elements, evolution of pathogenesis, and syndrome differentiation protocols; knowledge-driven is to make use of data and information to promote decision-making and action processes, and its syndrome-differentiation decision-making research relies on convolutional neural networks to improve the accuracy of local disease identification and syndrome differentiation. The “data-knowledge” dual-driven model can make up for the shortcomings of single-drive numerical simulation accuracy, and achieve a balance between local disease identification and macroscopic syndrome differentiation. On the basis of previous research, we explored the construction method of diagnostic assisted decision-making platform for gastric precancerous state, and believed that the diagnostic and decision-making ability of doctors can be extended through the assistance of machines and algorithms. Meanwhile, the related research methods were integrated and the core features of gastric precancerous state based on TCM syndrome differentiation and endoscopic pathology diagnosis and prediction were obtained, and the elements of endoscopic pathology recognition based on TCM syndrome differentiation were explored, so as to provide ideas for the in-depth research and innovative application of cutting-edge data analysis technology in the field of intelligent TCM syndrome differentiation.

Keywords gastric precancerous state; data analysis model; data-driven; knowledge-driven; decision-making based on syndrome differentiation

(收稿日期: 2023-03-19; 修回日期: 2023-10-31)

[编辑: 焦 爽]