

专题——中医证候创新研究与多维度中药药效评价

冠心病患者痰热互结证 CHAID 决策树
识别模式的研究史琦¹ 陈建新² 赵慧辉² 王伟²

(1 中日友好医院中医肺病二部,国家中医药管理局重点研究室(肺病慢性咳嗽),中医药防治过敏性疾病北京市重点实验室(BZ0321),中日友好医院呼吸中心;国家呼吸疾病临床研究中心,北京,100029; 2 北京中医药大学,北京,100029)

摘要 目的:建立冠心病不稳定性心绞痛患者临床常规检测指标对痰热互结证的识别模式。方法:选取 2010 年 4 月至 2011 年 4 月多中心收集的冠心病不稳定性心绞痛患者 411 例的基本资料、中医四诊信息及临床常规检测指标进行归一化处理后,采用 CHAID 决策树方法从 90 个临床常规检测指标中自动提取痰热互结证的识别规律。对其中 212 例患者进行痰热互结证识别模式的外验证。结果:Cl 离子、缩短分数、RDW-CV、血常规 RBC、D-II 聚体、CK-MB、PTA 和 BUN 共 8 个属性指标经筛选后进入决策树识别模型。该模型对 411 例患者的测试结果显示:敏感度为 75.0%,特异度为 86.9%,检验准确率为 86.1%。外验证模型缺失 RDW-CV,模型识别准确率为 85.8%,敏感度为 89.5%,特异度为 85.5%。结论:临床常规检测指标经 CHAID 决策树方法筛选后,可以直观、清晰的进行冠心病不稳定性心绞痛患者痰热互结证的识别,自动归纳识别规律,在中医证型-生物学指标对应模式的数据挖掘中具备一定的优势。

关键词 识别模式;CHAID 决策树;痰热互结证;冠心病;不稳定性心绞痛

Study on Recognition Modes of Phlegm-Heat Stasis Syndrome Decision Tree of Patients with Coronary Heart DiseasesShi Qi¹, Zhao Huihui², Chen Jianxin², Wang Wei²

(1 The Second Department of Pulmonary Disease in Traditional Chinese Medicine, China-Japan Friendship Hospital, The Key Unit of SATCM Pneumonopathy Chronic Cough and Dyspnea, Beijing Key Laboratory of Prevention and Treatment of Allergic Diseases with TCM (No. BZ0321), Center of Respiratory Medicine, China-Japan Friendship Hospital, National Clinical Research Center for Respiratory Diseases, Beijing 100029, China; 2 Beijing University of Chinese Medicine, Beijing 100029, China)

Abstract Objective: To build a recognition mode of phlegm-heat stasis syndrome according to indexes of clinical routine examinations of patients with unstable angina (UA). **Methods:** From April 2010 to April 2011, the clinical basic data, TCM four-diagnosis information and clinical routine examination indexes of 411 patients with UA were collected and normalized, and the recognition rules of phlegm-heat stasis syndrome was automatically extracted from 90 clinical examination indexes by applying CHAID decision tree method. The verification of recognition mode of phlegm-heat stasis syndrome in 212 patients was conducted. **Results:** A total of 8 indicators, including Cl ion, shortening fraction, RDW-CV, blood RBC, D-II mer, CK-MB, PTA and BUN were indexed into the decision tree model. The test results of 411 patients of the model showed that the sensitivity was 75.0%, the specificity was 86.9%, and the correct rate was 86.1%. The verification of association patterns lacked RDW-CV, and its correct rate was 85.8%. The sensitivity was 89.5%, and the specificity was 85.5%. **Conclusion:** The CHAID decision tree model can clearly and directly conduct recognition of phlegm-heat stasis syndrome with the basis of clinical routine examination indexes, and automatically summarize recognition rules, which has certain advantages in data mining of syndrome-physiochemistry model.

Key Words Recognition mode; CHAID decision tree; Phlegm-heat stasis syndrome; Coronary heart disease; Unstable angina

中图分类号:R2-03 文献标识码:A doi:10.3969/j.issn.1673-7202.2018.09.001

冠状动脉粥样硬化性心脏病指冠状动脉(冠脉)发生粥样硬化引起管腔狭窄或闭塞,导致心肌缺

血、缺氧或坏死而引起的心脏病,简称冠心病(Coronary Heart Disease, CHD),又称缺血性心脏病(Ische-

基金项目:国家科技重大新药创制专项(2009ZX09502)

作者简介:史琦(1983.09—),女,博士研究生,主治医师,研究方向:中医药防治心系及肺系疾病的临床及基础研究, E-mail: shiqi19830910@163.com

通信作者:王伟(1964.09—),男,博士研究生,教授,博士研究生导师,研究方向:中西医结合药理学, E-mail: wangwei@bucm.edu.cn; 赵慧辉(1977.05—),男,博士研究生,研究员,研究方向:中国传统医药防治心血管病的研究, E-mail: hh686@126.com

mic Heart Disease)。中医学注重整体观念和辨证论治,在冠心病的防御和治疗方面已逐步显示出其优势和特色。近年来,冠心病中医证候学的研究多采用冠脉造影金标准作为疾病诊断标准,在以往文献研究、回顾性研究基础上采用临床流行病学前瞻性研究,同时在系统论思想的指导下不断引入系统生物学方法,使得其研究在客观化、标准化方面取得了一定的成果和进展。中医“证”本质的研究在现代医学“实证主义”及“还原论”思想的引领下成为研究热点。但由于“证”不是对人体某个局部的反应,而是涉及到多个系统、多个器官、多个层次的整体反应,因此单系统、单层面的证候学研究显然难以全面揭示其科学内涵。

数据挖掘(Data Mining, DM)为解决“数据丰富,知识贫乏”状况而诞生,是数据库研究中极具应用前景的信息技术研究领域,融合了统计学、机器学习、人工智能、神经网络、信息检索、模式识别和数据可视化等多个领域的理论和技术^[1]。中医证候具有相对规律性、非单一趋向性改变的特点,其研究数据属于大量、随机、模糊性数据,涵盖了人体多方面、多层次、多指标的病理生理过程。数据挖掘技术正是一种从大量、有噪声、不完整的实际数据中,提取隐含规则、挖掘潜在信息的过程。决策树(Decision Tree, DT)是通过一系列规则对复杂数据进行分类的过程^[2]。中医四诊信息数量庞杂,在现代医学看来并非疾病的临床特异性表现,在西医临床诊断和疗效评价过程中的作用微乎其微。应用决策树的数据挖掘方法可以提取出与相应证型相关度最大的中医四诊信息要素,从而形成证候诊断模型。中医辨证过程以四诊信息为基础,临床常规生物学指标在辨证过程中常被忽视,然而我们前期的研究发现宏观层面的中医证候与微观层面的病理生理进程之间的互通可能正是以这些检测指标为基础^[3,4]。

本研究采用卡方自动交互检测(Chi-squared Automatic Interaction Detector, CHAID)决策树方法对冠心病不稳定性心绞痛(Unstable Angina, UA)痰热互结证患者的临床检验、检查指标群进行了分析处理,提取关联要素,旨在探讨临床常规生物学指标的组模式在 UA 患者中医证候/证候要素识别中的作用和意义。

1 资料与方法

1.1 一般资料 选取 2010 年 4 月至 2011 年 4 月中日友好医院、东直门医院、北京安贞医院、郑州市中医院、武汉市中医医院及湖北省中医院收治的冠

心病不稳定性心绞痛住院患者共计 411 例。

1.2 诊断标准 疾病诊断参照《慢性稳定型心绞痛诊疗指南》(美国心脏病学院/美国心脏学会/美国医师学院及美国内科学会联合议定,1999 年)^[5]及《不稳定性心绞痛诊断和治疗建议》(中华医学会心血管病学分会制订,2000 年)^[6]。中医证候、证型诊断参照《中医临床诊疗术语·证候部分》(国家技术监督管理局,1997 年)^[7]及《中药新药临床研究指导原则》(2002 年)^[8]。

1.3 纳入标准 1)符合冠心病不稳定性心绞痛诊断标准;2)年龄符合 35 ~ 75 岁之间;3)自愿签订知情同意书。

1.4 排除标准 1)急性心肌梗死、心包疾病、心肌疾病、心脏神经官能症、肋间神经痛、带状疱疹等疾病引起胸痛的患者;2)非冠心病引起心绞痛的患者,如风湿热、先天性冠脉畸形、主动脉瓣狭窄或关闭不全等;3)合并重度心律失常、脑中风、肾衰、肺部感染、肿瘤及其他严重原发性疾病患者;4)合并高血压患者血压未经控制或控制不佳者(收缩压 ≥ 180 mm-Hg 和/或舒张压 ≥ 110 mmHg);5)妊娠期、哺乳期患者;精神疾病患者;过敏体质患者。

1.5 研究方法

1.5.1 临床辨证及资料采集 入院 24 h 内对患者进行现病史、既往史、个人史、中医四诊信息等内容的全面采集和记录。疾病诊断及中医辨证由 3 名副主任医师及以上资格并具有相关临床工作经验 5 年及以上的临床医生完成。入院 1 周内完成血、尿、便常规、生化指标(肝功能、肾功能、血糖、血脂、心肌酶、尿酸、电解质等)、血清胰岛素、糖化血红蛋白、B 型尿钠肽等指标的检验及胸片、心电图、心脏彩超等的检查。

1.5.2 指标预处理方法 1)本研究中临床常规理化指标具有如下特点:指标数量庞杂;2)计数资料和计量资料混杂;3)部分指标存在缺失值;4)各临床分中心的检测方法及其参考范围存在差异性。

1.5.3 数据预处理方法 1)缺省值 $> 30\%$ 的指标予以剔除;2)经临床医师判断无实际意义的指标予以剔除;3)与冠心病心绞痛病理生理过程明确不相关的指标予以剔除;4)采用归一化方法对最终入选的指标进行处理,最终用于 CHAID 决策树模型筛选的临床理化指标共计 90 项。

1.5.4 数据挖掘及模型验证 筛选过程为:将经过预处理的 90 项临床常规检测指标作为自变量,将是否具有“痰热互结证”作为因变量,其中“痰热互结

1.6 统计学方法 采用 SPSS 17.0 CHAID 决策树方法进行证候识别规律的挖掘。

1.6 统计学方法 采用 SPSS 17.0 CHAID 决策树方法进行证候识别规律的挖掘。

2.1 2组患者基线情况比较 411例UA患者中,男占60.3%,女占39.7%,平均年龄(60.93 ± 8.70)岁。经临床辨证为痰热互结证的患者28例(6.8%),非痰热互结证患者383例(93.2%)。2组患者在体重指数方面比较,差异有统计学意义($P < 0.05$),在吸烟史方面比较,差异有统计学意义($P < 0.01$)。在年龄、性别比例、饮酒史、冠心病病程、合并病以及用药方面比较,差异无统计学意义($P > 0.05$)。见表1。

项目	非痰热互结证 ($n=383$)	痰热互结证 ($n=28$)	Z/χ^2	P
年龄(岁)	61.15 ± 8.55	57.89 ± 10.26	-1.577	0.115
性别(男/女)	228/155	20/8	1.544	0.214
体重指数(kg/m ²)	25.30 ± 4.00 *	26.61 ± 3.88	-2.162	0.031
冠心病病程(年)	4.00 ± 6.79	4.25 ± 6.61	-0.825	0.409
吸烟史(例)	116(30.3%) [△]	18(64.3%)	13.726	0.000
饮酒史(例)	76(19.8%)	9(32.1%)	2.406	0.121
合并高血压(例)	269(70.2%)	22(78.6%)	0.877	0.349
合并糖尿病(例)	129(33.7%)	13(46.4%)	1.875	0.171
合并心律失常(例)	32(8.4%)	1(3.6%)	0.291	0.590
合并高脂血症(例)	84(21.9%)	9(32.1%)	1.554	0.213
合并陈旧脑梗(例)	31(8.1%)	1(3.6%)	0.247	0.619
合并陈旧心梗(例)	29(7.6%)	2(7.1%)	0.000	1.000
使用硝酸酯类药物(例)	199(52.0%)	13(46.4%)	0.319	0.572
使用抗血小板药物(例)	350(91.4%)	24(85.7%)	0.449	0.503
使用 ACEI/ARB 药物(例)	179(46.7%)	15(53.6%)	0.489	0.484
使用 β 受体阻滞剂药物(例)	228(59.5%)	15(53.6%)	0.383	0.536
使用钙离子拮抗剂药物(例)	96(25.1%)	7(25.0%)	0.000	0.994
使用他汀类药物(例)	241(62.9%)	21(75.0%)	1.646	0.199

原始五路

Node 0

Category	%	n
0	93.2	383
1	6.8	28
Total	100.0	411

真

≤ 0.30723

Node 1

Category	%	n
0	94.0	68
1	16.0	13
Total	16.7	81

縮短片數

≤ 0.40351; <missing>

Node 3

Category	%	n
0	94.7	54
1	5.3	3
Total	13.9	57

紅細胞分布寬度-CV值

> 0.40351

Node 4

Category	%	n
0	58.3	14
1	41.7	10
Total	5.8	24

血克及紅細胞

≤ 0.0180; <missing>

Node 5

Category	%	n
0	97.8	122
1	2.2	3
Total	30.4	125

新舊貧血同工資

(0.0180; 0.04808]

Node 6

Category	%	n
0	77.8	21
1	22.2	6
Total	0.0	27

鐵血黃蛋白結晶度

> 0.04808

Node 7

Category	%	n
0	96.6	172
1	3.4	8
Total	43.3	178

尿表數

≤ 0.30683

Node 8

Category	%	n
0	100.0	0
1	0.0	0
Total	11.4	47

紅細胞分布寬度-CV值

> 0.30683; <missing>

Node 9

Category	%	n
0	70.0	7
1	30.0	3
Total	2.4	10

血克及紅細胞

≤ 0.62100

Node 10

Category	%	n
0	64.0	11
1	15.4	2
Total	3.2	13

新舊貧血同工資

> 0.62100

Node 11

Category	%	n
0	27.3	3
1	72.7	8
Total	2.7	11

鐵血黃蛋白結晶度

≤ 0.05660

Node 12

Category	%	n
0	98.9	24
1	11.1	3
Total	6.5	27

尿表數

> 0.05660; <missing>

Node 13

Category	%	n
0	100.0	98
1	0.0	0
Total	23.8	98

紅細胞分布寬度-CV值

≤ 0.84482

Node 14

Category	%	n
0	94.1	10
1	5.9	1
Total	4.1	17

血克及紅細胞

> 0.84482

Node 15

Category	%	n
0	50.0	5
1	50.0	5
Total	2.4	10

新舊貧血同工資

≤ 0.28107

Node 16

Category	%	n
0	91.7	60
1	8.3	0
Total	17.5	72

鐵血黃蛋白結晶度

> 0.28107

Node 17

Category	%	n
0	100.0	126
1	0.0	0
Total	25.8	126

尿表數

图 1 痰热互结证 CHAID 决策树模型

项临床常规检测指标建立的 CHAID 决策树识别模型结果显示:共筛选出 8 个用于决策树判断的属性指标,分别是氯离子(Cl)、缩短分数、D-II 聚体、红细胞分布宽度-CV 值(RDW-CV)、血常规红细胞(RBC)、肌酸激酶同工酶(CK-MB)、凝血酶原活动

度(PTA)和尿毒氮(BUN)。该模型包括了 18 个节点和 10 个终节点,形成了 10 条判断是否“痰热互结证”的识别路线。比如,Cl 离子 $\leq 0.367\ 23$,缩短分数 $>0.493\ 51$,RBC $>0.621\ 00$,则判断为痰热互结证的比例为 72.7%;非痰热互结证的比例为 27.3%,

表 2 痰热互结证决策树模型的 10 倍交叉验证结果

CHAID 决策树	TN	FP	FN	TP	敏感度(%)	特异度(%)	准确率(%)
痰热互结证	333	50	7	21	75.0%	86.9%	86.1%

注:敏感度 = TP/(TP + FN);特异度 = TN/(TN + FP);准确率 = (TP + TN)/(TP + FP + TN + FN)

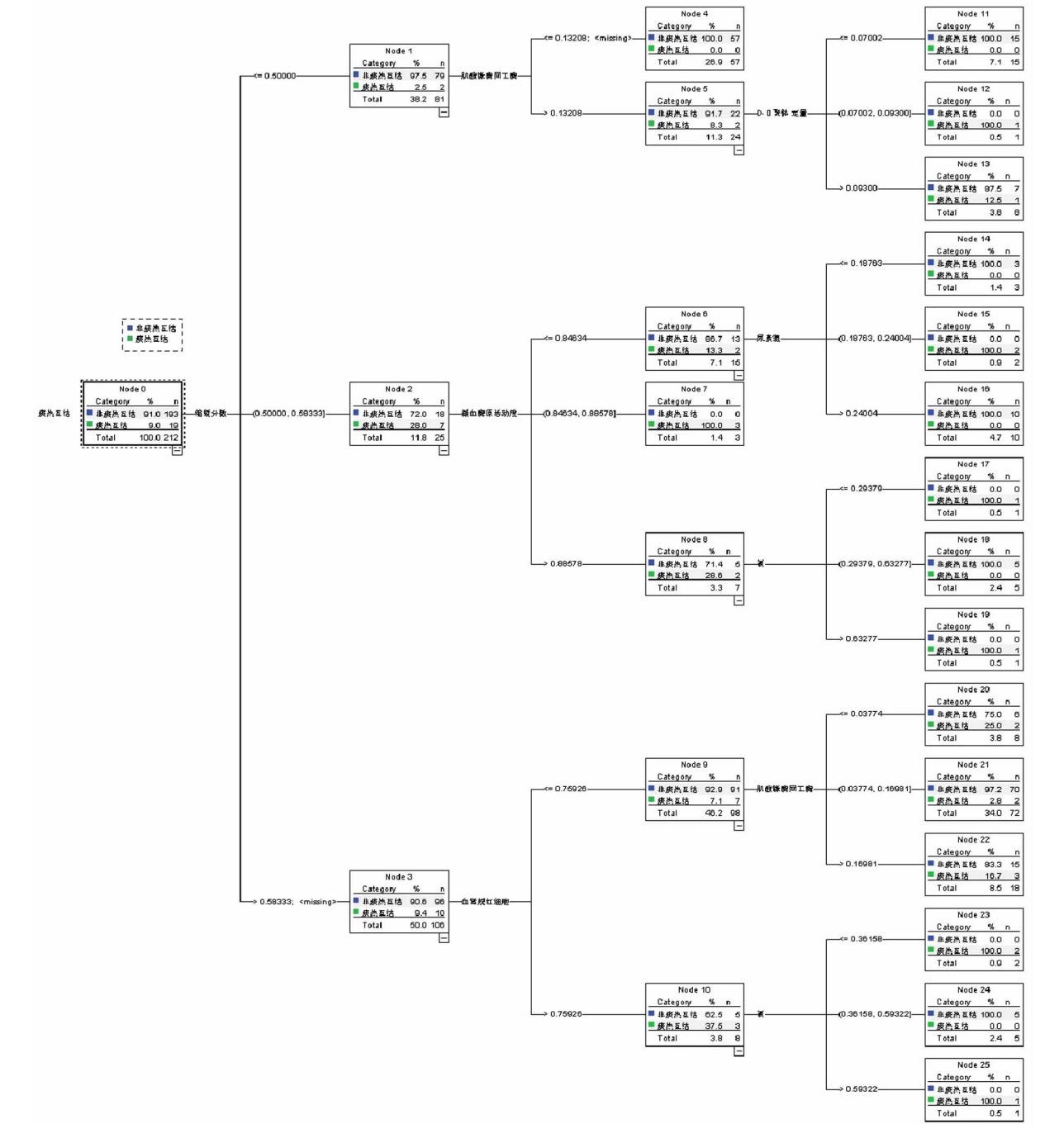


图 2 痰热互结证决策树外验证模型

表 3 痰热互结证外验证决策树模型的 10 倍交叉验证结果

CHAID 决策树	TN	FP	FN	TP	敏感度 (%)	特异度 (%)	准确率 (%)
痰热互结证	165	28	2	17	89.5%	85.5%	85.8%

注:敏感度 = TP/(TP + FN);特异度 = TN/(TN + FP);准确率 = (TP + TN)/(TP + FP + TN + FN)

见图 1。

2.3 决策树识别模型的验证 10 倍交叉验证结果显示:该决策树模型对是否“痰热互结证”的识别敏感度为 75.0%,特异度为 86.9%,准确率为 86.1%。见表 2。

2.4 痰热互结证决策树的外验证模型 痰热互结证决策树外验证模型包括 7 个属性指标:Cl 离子、缩短分数、RBC、D-Ⅱ聚体、CK-MB、PTA 及 BUN,原模型中的 RDW-CV 缺失。该模型形成了 17 条对判断是否“痰热互结证”的识别路径。见图 2。

2.5 痰热互结证决策树外验证模型的 10 倍交叉验证 采用 212 例 UA 患者建立的痰热互结证外验证模型经 10 倍交叉方法验证后结果显示:识别准确率为 85.8%,敏感度为 89.5%,特异度为 85.5%。见表 3。

3 讨论

痰浊和热毒是冠心病心绞痛患者较为常见的病因病机。“痰浊”和“热毒”导致冠心病发生发展的理论和临床研究经历了 3 个时期。萌芽期以《黄帝内经》和《金匱要略》为代表。《灵枢·邪客》:“心者,五脏之大主,精神之所舍也”。心神对机体津液代谢具有重要的调节作用。心能温肺助脾,如果脾为湿困,精微不运,凝结成痰,流于经脉,则阻滞于心脉;心阳可下暖肾水,化气行津。如心阳虚弱,阴阳气血不得接续,津液输布失度,日久则累聚成痰。张仲景在《金匱要略·胸痹心痛短气病脉证治》中提出“阳微阴弦”是胸痹心痛的主要病机,由此制定的瓜蒌薤白白酒汤、瓜蒌薤白半夏汤等方开创了豁痰宽胸、开痹散结,从痰浊论治冠心病心绞痛的大法。《素问·至真要大论》:“风、寒、湿、火、热诸邪,皆能致病心痛”,《素问·刺热篇》:“热争则卒心痛,烦闷善呕,头痛面赤,无汗”等论述初步提出了“热毒”致病理论。在发展期,诸多后世医家对“痰热”理论加以发挥。如《诸病源候论》:“……邪迫于阳气,不得宣畅,壅瘀生热,故心如悬而急烦懊痛也”。《古今医鉴》中指出心痛病由实热引发:“凡痛在心,连两胁至两乳下,牵引背板匙骨下而痛者,实热也”。张元素主张“久痛无寒而暴痛非热”是对“热厥心痛”的阐述。“有热厥心痛者,身热足寒,痛甚则烦躁而

吐,额汗出,知为热也”。《症因脉治·胸痛论》中“内伤胸痛之因,七情六欲,动其心火……;或怫郁气逆,伤其肺道,则痰凝气结”,体现了痰热互因。随着现代医学的不断发展,冠心病中医“痰”“热”研究日臻成熟。冠心病证候要素的相关研究发现,建立的 2 029 例冠心病心绞痛患者的证候数据库中痰浊偏热证在心绞痛证候分布特点上具有重要的位置^[9]。采用西雅图心绞痛量表(SAQ)评价了 233 例冠心病心绞痛患者的生命质量(躯体活动受限程度、心绞痛稳定状态、发作情况、治疗满意度及疾病认识程度 5 个维度)发现,其中对于生命质量的影响结果显示痰浊证呈现负相关,而热蕴证呈现正相关;热蕴证可以保护心绞痛发作并影响心绞痛的稳定状态,痰浊证对治疗满意度呈负向影响^[10]。

中医学的证候是在西医学疾病概念的基础上,对独立的病理、生理进程的整体概括和反应。若将现代医学中以“金标准”诊断疾病的思路方法套用至中医证候学研究中,则必将导致中医“证”的泛滥。生物学指标与“证”的相关性研究曾一度成为中医证候学研究的范式,但线性相关性结果难以揭示“证”的整体性。比如在冠心病中医证候与冠脉造影结果相关性研究中,有学者认为,冠心病实证(心血瘀阻型、痰浊内阻型)患者冠状动脉病变较虚证(心气虚弱型、心肾阴虚型)重,实证以多支病变为主,虚证以单支病变为主;冠脉病变积分由高到低依次为心血瘀阻型>痰浊内阻型>心气虚弱型>心肾阴虚型。另有学者通过比较分析中医证候与冠状动脉 Gensini 积分、病变支数的关系发现冠心病患者中医实证组以单支病变为主(41.3%),高于虚证组(25.0%)和虚实夹杂证组(23.6%);虚证组患者的冠状动脉病变累及支数大于虚实夹杂证组和实证组;虚证组的 Gensini 积分亦高于虚实夹杂证和实证组^[11]。上述相互矛盾的研究结果提示我们线性研究方法无法解决“证”的模糊性和不确定性,只有复杂性科学和非线性科学研究方法的恰当引入,才能实现“证”这一复杂系统的实质阐述。

决策树法是通过确定一系列的 if-then 逻辑(分枝)关系,形成分层规则,从而达到对研究数据进行精确预测或正确分类的目的。决策树学习属于机器

学习的范畴,是一种类似于判别分析的有监督的学习方法,其分类结果简单明确、结构直观,适用于较大规模的数据集研究^[12]。近年来,决策树在中医药研究中主要应用于证候特征及诊断的规范化、建立中医辨证识别模型、证候演变规律及其影响因素研究等方面^[13-14]。Kass 于 1980 年提出了 CHAID 即卡方自动交互检测应用于分类及等级数据的分析^[15]。其学习过程为选定结果变量即目标变量,分类指标与目标变量进行交叉分类,产生二维分类表;计算一系列二维分类表的卡方值,以 P 值最小者作为最佳初始分类表,在此基础上继续应用分类指标对目标变量进行分类,不断重复上述过程直至 P 值大于设定的有统计学意义的 α 值,分类停止,产生决策树模型。CHAID 的优势在于:分析过程能够显示出各个变量之间的相互作用关系;分析结果以树图形显示,分析过程及结果清晰直观;不会因自变量之间的共线性而遗漏有意义的变量^[16]。

本研究的痰热互结证决策树模型包括 Cl 离子、缩短分数、RDW-CV、血常规 RBC、D-Ⅱ聚体、CK-MB、PTA 及 BUN 共 8 个属性指标。CHD 患者特别是发生心力衰竭的患者可能出现低氯血症^[17],本研究痰热互结证患者的 Cl 离子水平显著低于非痰热互结证,说明痰热互结证患者的电解质紊乱情况更加明显。冠心病心力衰竭阳虚水泛证、心阳虚脱证患者与心气阴虚证、心气阳虚证、气虚血瘀证患者比较,其左室短轴缩短率(FS)有降低趋势^[18]。本研究中,痰热互结患者的 FS 水平无明显降低,说明痰热互结证患者的心室功能尚可。痰热互结证患者与非痰热互结证患者的 RDW 水平并无显著的统计学线性差异,但在决策树模型中,RDW 水平升高则痰热互结证患者的比例上升,说明 CHD 痰热证患者中 RDW 的升高趋势明显。研究显示不稳定性心绞痛患者气虚血瘀证与外周血 RBC、MCV、MCHC、DBIL 等相关;Logistic 回归结果显示:气虚血瘀证影响因素依次为胆红素、RBC、MCH、HCT 等^[19]。本研究中,痰热互结证患者 RBC 计数增高提示了其血液流变学的异常情况。D-Ⅱ聚体是交联纤维蛋白的降解产物,同时反应体内的纤溶活性和凝血活动,D-Ⅱ聚体浓度升高是血栓形成和继发性纤溶的重要标记物。实验证实血浆 D-Ⅱ聚体和 AS 密切相关^[20]。研究发现 AMI 和 UAP 患者的血浆 D-Ⅱ聚体水平明显高于 SAP 患者,且 AMI 患者升高更加明显,同时 D-Ⅱ聚体还可能与冠脉病变的严重程度相关^[21]。本次研究的识别模型及验证模型中,随着 D-Ⅱ聚体

的升高,痰热互结证患者的比例增多。CK-MB 属于诊断急性心肌梗死的传统心肌酶学指标。心肌急性缺血缺氧导致坏死时,细胞膜通透性改变,心肌细胞内的 CK-MB 释放入血清。CK-MB 的增高程度能较准确地反映心肌梗死的范围,但敏感性不足,与骨骼肌有交叉^[22]。痰热互结证模型中,CK-MB 被筛选为属性指标,但其变化趋势与该证型的关系需要进一步探讨。本研究结果显示在决策树模型中,随着 PTA 水平的升高,诊断痰热互结证的比例也随之升高,提示痰热证患者的凝血倾向更加明显。研究认为 BUN 可能是高血压等传统因素之外的冠心病不可忽视的危险因素之一^[23]。本研究中,痰热互结证患者的 BUN 水平明显低于非痰热互阻证,提示 BUN 可能不是识别痰热互结证的主要属性指标。

中医证候学的研究,其重点在于辨证的客观化和标准化。目前对于冠心病中医证候学研究已经取得了较为丰硕的成果,亟待解决的问题是如何为中医辨证提供同时具备良好特异性及敏感性的依据,将辨证指标定量化、标准化,进而形成独特的辨证体系,实现现代医学病理、生理进程与中医证候学病因、病机之间的良好沟通,最终阐明冠心病中医“证”的实质和本质。方法学一直以来都是中医证候学研究的最大瓶颈,中医辨证过程具有模糊性、复杂性、非线性及非定量性等特点。分子生物学研究方法的介入使得中医的“证”向实质化迈进了一大步,但“证”的本质往往不是仅涉及到一种分子,而是由多种分子相互作用而引发的证候群。针对单一分子的研究只能反应辨证实质的一个侧面,缺乏特异性和敏感性。缺乏多层次、多角度及高通量信息特点的“证”的研究也无法对证候进行全面阐释,无法揭示人体作为一个有机整体的“证”变化规律。随着系统论思想和系统生物学方法的不断引入,中医证候学的研究已经进入了立足于复杂科学和理论方法的全新阶段。数据挖掘技术凭借其特点和优势,逐步实现了对庞杂的中医证候数据的综合考量。数据挖掘方法可以有效的对中医证候数据进行分类预判,获得证候相关预测因子或核心证候,实现“即时辨证”^[24]。中医药信息化的发展促进了多数研究采用中医药数据库系统,实现了中医数据的规范性和高效应用,但目前研究大多属于小样本范畴,并没有真正实现中医“大数据”概念,研究的真实性、可重复性差,众多研究结果之间缺乏可比性,甚至筛选出的“核心指标”出现相互矛盾,难以指导临床应用。临床情况的复杂性和多样性决定了数据挖掘方

法建立的证候识别模型实现“即时准确辨证”尚待深入研究。我们在前期对冠心病不稳定性心绞痛患者血瘀证、气虚证、痰瘀互阻证等多种证型决策树模型构建及验证的基础之上^[25-27],建立了冠心病患者痰热互结证决策树模型,识别模型的准确率、特异度及敏感度均达到 75% 以上,外验证模型的准确率、特异度及敏感度均达到了 85% 以上。但研究中部分临床常规检测指标存在缺失值,CHAID 算法不能在树的生长过程中排除缺失值;病例数特别是外验证病例数相对较少,不适宜采用分割验证法对该模型进行进一步的验证。后续研究中,我们将把应用决策树方法形成的“网络”判断模型与其他数据挖掘方法的识别模式进一步比较,从而筛选出用于冠心病不稳定性心绞痛患者证候判别的“核心因素”。

参考文献

- [1] 姚美村,袁月梅,艾路,等.数据挖掘及其在中医药现代化研究中的应用[J].北京中医药大学学报,2002,9(5):20-23.
- [2] 罗林英,陈金雄.基于决策树的外科手术信息挖掘与研究[J].医学信息,2008,21(11):1936-1939.
- [3] 吕爱平,李梢,王永炎.从主观症状的客观规律探索中医证候分类的科学基础[J].中医杂志,2005,46(1):5-7.
- [4] Zhao HH,Guo S,Chen J,et al.Characteristic Pattern Study of Coronary Heart Disease with Blood Stasis Syndrome Based on Decision Tree[J].International Conference on Bioinformatics&Biom,2010,4(4):1-3.
- [5] Gibbons RJ,Chatterjee K,Dale J,et al.ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina[J].J Am Coll Cardiol,1999,33(7):2092-2197.
- [6] 中华医学会心血管病学分会.不稳定性心绞痛诊断和治疗建议[J].中华心血管病杂志,2000,28(6):409-412.
- [7] 朱文峰,王永炎.中医临床诊疗术语-证候部分[M].北京:中国标准出版社,1997:17-20.
- [8] 国家食品药品监督管理局.中药新药临床研究指导原则[M].北京:中国医药科技出版社,2002:41.
- [9] 周景想,唐明,李洁,等.2029 例冠心病心绞痛中医证候特点及组合规律分析[J].中国中西医结合杂志,2011,31(6):753-755.
- [10] 王阶,汤艳莉,何庆勇.冠心病心绞痛患者证候要素与生存质量的相关性[J].中医杂志,2010,51(2):154-163.
- [11] 王振裕,时亮,陈嘉兴.冠心病患者中医证候与冠状动脉造影相关性研究[J].中华中医药杂志,2009,24(9):1194-1197.

- [12] 徐蕾,贺佳,孟虹,等.决策树技术及其在医学中的应用[J].数理医药学杂志,2004,17(2):161-164.
- [13] 李锋刚.数据挖掘技术在新安医学研究中的应用[J].安徽中医学院学报,2005,24(6):12-14.
- [14] 李梢,张宁波,李志红,等.慢性乙型肝炎患者肝胆湿热证和肝郁脾虚证的决策树诊断模型初探[J].中国中西医结合杂志,2009,29(11):993-996.
- [15] Kass GV. An exploratory technique for investigating large quantities of categorical data[J]. Journal of the Royal Statistical Society,1980,29(2):119-127.
- [16] 石玲,王燕.婴幼儿死亡危险因素的研究——兼论 CHAID 方法的原理及应用[J].中国卫生统计,2002,19(5):283-285.
- [17] 赵琛,王敏,范茜,等.老年冠心病慢性心力衰竭并发电解质紊乱的临床特点[J].齐鲁医学杂志,2008,23(5):398-402.
- [18] 杜柏,商秀洋,胡元会,等.冠心病心力衰竭患者中医证型与心率变异性、醛固酮及心功能关系的研究[J].中西医结合心脑血管病杂志,2010,8(8):906-908.
- [19] 其其格,李耀扬,张鹏,等.冠心病不稳定性心绞痛气虚血瘀证与血清理化指标间相关性研究[J].中华中医药杂志,2011,26(2):380-383.
- [20] Salomaa V,Stinson V,Kark JD,et al.Association of fibrinolytic parameters with early atherosclerosis The ARIC study. Thrombosis Risk in Communities Study[J]. Circulation,1995,91(2):284-290.
- [21] 董琦,牛凡,杨涓.D-二聚体、纤维蛋白原与冠状动脉病变程度及左心功能的关系[J].中国医疗前沿,2011,6(2):3-4.
- [22] 何德化,李卫华,谢强,等.冠心病患者分泌型磷脂酶 A2 与高敏 C 反应蛋白、肌酸激酶的检测及其临床价值[J].医学临床研究,2008,25(3):515-517.
- [23] 金春梅,金基永.血清白蛋白尿素氮肌酐水平与冠心病的相关性初探[J].现代医药卫生,2010,26(19):2912-2914.
- [24] 陈克龙,樊永平.数据挖掘中的分类算法及其在中医证候学中的应用[J].中华中医药杂志,2011,26(3):469-473.
- [25] 史琦,王伟,陈建新,等.基于决策树的冠心病痰瘀互阻证识别模式研究[J].中华中医药杂志,2013,28(12):3523-3526.
- [26] 史琦,陈建新,赵慧辉,等.冠心病病人血瘀证的决策树模型研究[J].中西医结合心脑血管病杂志,2013,11(8):897-900.
- [27] Shi Q,Zhao HH,Chen JX,et al.Study on Qi Deficiency Syndrome Identification Modes of Coronary Heart Disease Based on Metabolomic Biomarkers[J]. Evidence-Based Complementary and Alternative Medicine,2014,2014(3):281829.

(2018-04-24 收稿 责任编辑:张文婷)