

基于贝叶斯网络与决策树的中药注射剂致药品不良反应的流行病学特点研究[△]

黎 鹞^{1*}, 覃慧莲², 甘键钟³, 邝俊健^{1,4#}, 梁健龙¹, 梁杰豪¹, 黄秋燕^{1,4}, 梁 波^{1,4}, 李 甜¹, 梁 俊¹(1. 梧州市中医医院药学部, 广西 梧州 543002; 2. 梧州市中医医院外三科, 广西 梧州 543002; 3. 梧州市中医医院下肢骨二科, 广西 梧州 543002; 4. 梧州市药事管理质量控制中心, 广西 梧州 543002)

中图分类号 R932;R969.3 文献标志码 A 文章编号 1672-2124(2021)01-0101-05

DOI 10.14009/j.issn.1672-2124.2021.01.025

摘要 目的:探讨广西壮族自治区梧州市(以下简称“梧州地区”)中药注射剂致药品不良反应(adverse drug reaction, ADR)的流行病学特点。方法:收集2011—2019年梧州地区中药注射剂致ADR报告,使用Microsoft ACCESS软件构建数据库,进而以SPSS Modeler 14.1数据挖掘平台,应用CART、CHAID决策树算法和贝叶斯网络等方法开展数据挖掘研究。结果:共收集中药注射剂致ADR详细案例1370例,其中一般的ADR为1239例(占90.44%),严重的ADR为131例(占9.56%)。贝叶斯网络分析结果显示,中药注射剂致ADR的临床表现与患者年龄等因素存在相关性,在>60岁年龄段人群中,皮肤损害所占比例逐渐降低,其他系统损害所占比例逐渐升高。决策树CART算法结果显示,ADR的临床表现主要为皮肤损害,各年龄段患者的ADR以皮肤损害为主。决策树CHAID算法结果显示,ADR的临床表现主要为皮肤损害,一般的ADR以皮肤损害为主,严重的ADR以全身系统损害、呼吸系统损害为主。结论:梧州地区中药注射剂致ADR的流行病学特点可能与患者年龄、ADR发生时间和ADR类型等因素相关,尚需更大样本量的数据分析验证。

关键词 中药注射剂;数据挖掘;药品不良反应;决策树;贝叶斯网络

Epidemiological Characteristics of Adverse Drug Reactions Induced by Traditional Chinese Medicine Injections Based on Bayesian Network and Decision Tree[△]

LI Li¹, QIN Huilian², GAN Jianzhong³, KUANG Junjian^{1,4}, LIANG JianLong¹, LIANG Jiehao¹, HUANG Qiuyan^{1,4}, LIANG Bo^{1,4}, LI Tian¹, LIANG Jun¹(1. Dept. of Pharmacy, Wuzhou Hospital of Traditional Chinese Medicine, Guangxi Wuzhou 543002, China; 2. Dept. of Surgery, Wuzhou Hospital of Traditional Chinese Medicine, Guangxi Wuzhou 543002, China; 3. Dept. of Orthopedics, Wuzhou Hospital of Traditional Chinese Medicine, Guangxi Wuzhou 543002, China; 4. Wuzhou Pharmaceutical Administration Quality Control Center, Guangxi Wuzhou 543002, China)

ABSTRACT OBJECTIVE: To probe into the epidemiological characteristics of adverse drug reactions (ADR) induced by traditional Chinese medicine injections in Wuzhou, Guangxi Zhuang Autonomous Region (Hereinafter referred to as “Wuzhou”). METHODS: ADR reports induced by traditional Chinese medicine injections in Wuzhou from 2011 to 2019 were collected. Microsoft ACCESS software was used to establish the database, and SPSS Modeler 14.1 data mining platform was used to carry out data mining research based on CART, CHAID decision tree algorithm and Bayesian network methods. RESULTS: A total of 1370 detailed cases of ADR induced by traditional Chinese medicine injections were collected, of which 1239 cases were general (90.44%) and 131 cases were severe (9.56%). The results of Bayesian network analysis showed that the clinical manifestations of ADR induced by traditional Chinese medicine injection were related to the age of patients. In the age group of >60 years old, the proportion of skin damage gradually decreased, and the proportion of other system damage gradually increased. The results of cart algorithm showed that the main clinical manifestation of ADR was skin damage, and skin damage was the main manifestation of ADR in all age groups. The results of decision tree CHAID algorithm show that the main clinical manifestation of ADR was skin damage, the general ADR was mainly skin damage, and the serious ADR were mainly

△ 基金项目:广西壮族自治区中医药局自筹经费科研课题(No. GZZC2019276)

* 主管中药师。研究方向:中药临床药学。E-mail:422927606@qq.com

通信作者:副主任药师。研究方向:临床药学。E-mail:89570353@qq.com

systemic damage and respiratory damage. . CONCLUSIONS: The epidemiological characteristics of ADR induced by traditional Chinese medicine injections in Wuzhou may be related to the age of patients, occurrence time of ADR and types of ADR, which need to be analyzed and verified by data analysis of a larger sample size.

KEYWORDS Traditional Chinese medicine injections; Data mining; Adverse drug reactions; Decision tree; Bayesian network

药品不良反应(adverse drug reaction,ADR)信号筛选和发生特点(如 ADR 发生时间、类型与患者年龄、性别、联合用药及过敏史等的相关性)的挖掘是药品上市后安全性评价的重要研究内容之一,常用的研究方法主要有药物流行病学方法(如病例对照研究、队列研究)、贝叶斯网络方法和传统统计学方法等^[1]。目前,国内外使用数据挖掘决策树方法探索 ADR 数据平台中 ADR 与药品、患者年龄、性别等因素相关性的研究较少^[2]。本研究通过收集广西壮族自治区梧州市(以下简称“梧州地区”)ADR 监测中心自发呈报系统(spontaneous reporting system,SRS)上报的中药注射剂致 ADR 报告,借助 SPSS Modeler 14.1 数据挖掘平台软件,应用贝叶斯网络研究方法和决策树算法,对梧州地区中药注射剂致 ADR 信息进行挖掘,为梧州地区中药注射剂致 ADR 的发生规律和促进药物警戒研究新方法的探索提供参考。

1 资料与方法

1.1 资料来源

资料来源于 2011—2019 年梧州地区上报的有效 ADR 报告,共收集到 1 370 例中药注射剂致 ADR 案例,涉及 32 种该地区使用的中药注射剂。

1.2 数据库的建立

结合中药注射剂的特点和 ADR 报告的实际情况,应用 Microsoft ACCESS 软件建立数据库,字段涉及病案中的患者年龄、性别信息、ADR 发生时间、ADR 类型、药品通用名称、过敏史情况、原患疾病、关联性评价、联合用药以及预后转归等关键信息。

1.3 数据挖掘方法

应用 SPSS Modeler 14.1 数据挖掘系统中的贝叶斯网络方法,分析梧州地区中药注射剂致 ADR 的相关因素强度,再利用决策树算法(CART、CHAID),对主要影响因素与梧州地区中药注射剂致 ADR 之间的关系进行深入挖掘研究。决策树是数据挖掘中最为常用的分类预测模型之一,是能够使研究者直观地运用概率分析的一种图解法^[1,3-5]。决策树常见的算法有 CART、CHAID 等。其中,CART 算法的输入变量和输出变量可以是分类变量也可以是数值变量,只能是二叉树;CHAID 是卡方自动交互诊断器,输入变量和输出变量可以是分类变量也可以是数值变量,CHAID 为多叉树。CART 算法以 Gini 系数和方差为基础选择最佳变量和分割点,可依据检验样本集进行剪枝;CHAID 算法从统计显著性检验角度确定最佳分割点^[6]。故本研究选用 CART、CHAID 算法构建决策树,对梧州地区中药注射剂致 ADR 进行数据挖掘研究。

2 结果

2.1 基于数据可视化的信息分析

收集的梧州地区 1 370 例中药注射剂致 ADR 报告中,男

性患者 600 例(占 43.80%),女性患者 770 例(占 56.20%);患者年龄 2 个月至 97 岁,平均年龄为(52.06±24.05)岁,其中>60 岁患者 605 例(占 44.16%);根据患者的原发疾病分类,主要包括心血管疾病 285 例(占 20.80%)、上呼吸道感染 226 例(占 16.50%)、脑血管疾病 158 例(占 11.53%)、支气管肺炎 154 例(占 11.24%)、骨科疾病 152 例(占 11.09%)和神经系统疾病 82 例(占 5.99%)等,其中心血管疾病最多。

1 370 例中药注射剂致 ADR 中,一般的 ADR 为 1 239 例(占 90.44%),严重的 ADR 为 131 例(占 9.56%);患者明确无过敏史的为 637 例(占 46.50%),患者明确有过敏史的为 71 例(占 5.18%),其余均为未写明具体的过敏药物类别或名称;1 073 例 ADR(占 78.32%)发生于输液开始后 60 min 内;314 例 ADR(占 22.92%)有联合用药记录,联合使用的注射剂包括中药和化学药品,其中有 101 例联合应用抗菌药物。

1 370 例中药注射剂致 ADR 累及人体多个器官和(或)系统,其中,皮肤损害 647 例(占 47.23%),循环系统损害 205 例(占 14.96%),全身系统损害 202 例(占 14.74%),呼吸系统损害 103 例(占 7.52%),消化系统损害 85 例(占 6.20%),过敏样反应 39 例(占 2.85%),神经系统损害 27 例(占 1.97%),神经系统损害 25 例(占 1.82%),其他 20 例(占 1.46%),过敏性休克 17 例(占 1.24%)。由此可知,中药注射剂致 ADR 的临床表现中,皮肤损害所占比例最高。

2.2 基于贝叶斯网络的梧州地区中药注射剂致 ADR 的发生概率分析

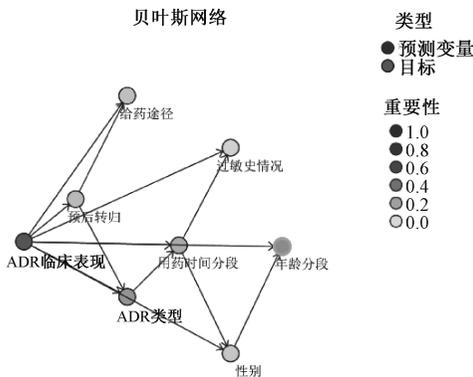
运用贝叶斯网络分析与梧州地区中药注射剂致 ADR 临床表现相关的因素,结果显示,患者年龄、ADR 类型和 ADR 发生时间与 ADR 临床表现的关系最紧密,见图 1、表 1—2。

2.3 基于决策树算法分析梧州地区中药注射剂致 ADR

2.3.1 运用决策树 CART 算法分析中药注射剂致 ADR 的临床表现:构建 CART 决策树,最小杂质改变设置为 0.000 1,最大代用项设置为 5,以灵活处理缺失值,尽量低于最优拆分引起的复杂度变量,同时对已生成的树按照一定的规则进行剪枝,以避免过度拟合,提高拟合效果。结果显示,决定决策树分类最重要的相关因素为 ADR 类型、患者年龄和用药时间,其次为预后转归情况等,与贝叶斯网络研究结果一致。决策树 CART 算法预测变量的重要性以及决策树模型图见图 2。

2.3.2 运用决策树 CHAID 算法分析不良反应临床表现:构建 CHAID 决策树,卡方类别目标设为 Pearson,最低单元格频率为 0.001,得到决策树 CHAID 算法的预测变量重要性及决策树 CHAID 模型图,见图 3。全部案例报告中,决策树 CHAID 分类最重要的相关因素为患者年龄,其次为 ADR 类型和 ADR 发生时间,与贝叶斯网络研究结果一致。

贝叶斯网络



预测变量重要性
目标:不良反应临床表现

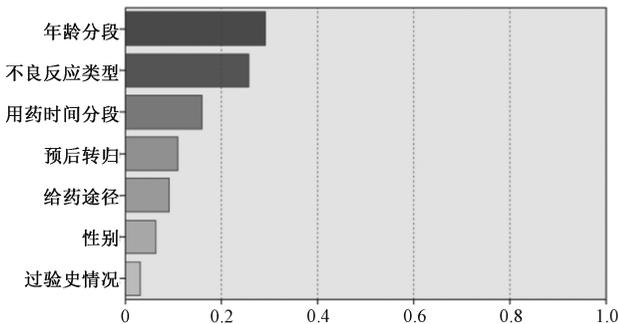


图1 基于贝叶斯网络的中药注射剂致ADR临床表现相关因素强度分析图

Fig 1 Intensity of clinical manifestations of ADR induced by traditional Chinese medicine injections based on Bayesian network

表1 基于贝叶斯网络的患者年龄与ADR临床表现概率分析

Tab 1 Probability of age and clinical manifestations of ADR induced by traditional Chinese medicine injections based on Bayesian network

ADR 临床表现	不同年龄段人群发生 ADR 的概率(男性/女性)				
	≤20 岁	>20~40 岁	>40~60 岁	>60~80 岁	>80 岁
皮肤损害	0.29/0.14	0.18/0.18	0.25/0.34	0.22/0.28	0.05/0.05
循环系统损害	0.04/0.04	0.07/0.12	0.26/0.36	0.50/0.34	0.12/0.12
全身系统损害	0.11/0.03	0.06/0.10	0.19/0.20	0.48/0.50	0.16/0.17
呼吸系统损害	0.05/0.03	0.03/0.26	0.30/0.17	0.62/0.41	0/0.14
消化系统损害	0.19/0.08	0.03/0.16	0.11/0.35	0.53/0.35	0.14/0.06
过敏样反应	0.24/0.05	0/0.27	0.24/0.23	0.24/0.27	0.29/0.18
运动系统损害	0/0	0.25/0	0.08/0.27	0.67/0.53	0/0.20
神经系统损害	0.18/0.07	0.09/0.07	0.27/0.43	0.45/0.36	0/0.07
其他	0.27/0	0.09/0.44	0.27/0.22	0.27/0.33	0.09/0
过敏性休克	0/0	0.14/0.20	0.29/0.30	0.57/0.50	0/0

表2 基于贝叶斯网络的 ADR 发生时间与临床表现概率分析

Tab 2 Probability of occurrence time and clinical manifestations of ADR induced by traditional Chinese medicine injections based on Bayesian network

ADR 临床表现	输液开始后不同时间段发生 ADR 的概率(严重的 ADR/一般的 ADR)							
	0~5 min	>5~10 min	>10~20 min	>20~30 min	>30~40 min	>40~60 min	>1~2 h	>24 h
皮肤损害	0.10/0.06	0.15/0.12	0.10/0.11	0/0.07	0.01/0.01	0.10/0.30	0.10/0.09	0.45/0.24
循环系统损害	0.09/0.14	0/0.17	0/0.13	0/0.08	0.09/0.02	0.55/0.26	0/0.07	0.27/0.14
全身系统损害	0.08/0.03	0.08/0.17	0.15/0.15	0.08/0.15	0/0.04	0.23/0.27	0.08/0.05	0.31/0.13
呼吸系统损害	0.22/0.12	0.17/0.24	0.11/0.16	0.06/0.08	0/0.01	0.22/0.27	0/0.02	0.22/0.09
消化系统损害	0.12/0.16	0/0.06	0/0.16	0.12/0.04	0/0	0.38/0.36	0/0.09	0.38/0.13
过敏样反应	0.24/0.21	0/0.14	0.08/0.07	0.04/0.14	0/0	0.20/0.21	0.20/0	0.24/0.21
运动系统损害	0.67/0.12	0/0.12	0.33/0.12	0/0	0/0	0/0.29	0/0.12	0/0.17
神经系统损害	0/0.05	0.67/0.05	0/0.18	0/0.05	0/0	0/0.41	0/0.23	0.33/0.05
其他	0/0.10	0/0.20	0/0.10	0/0.05	0/0	0/0.35	0/0.05	0/0.15
过敏性休克	0.35/0	0.06/0	0/0	0.06/0	0.06/0	0.29/0	0.06/0	0.12/0

预测变量重要性
目标: ADR 临床表现

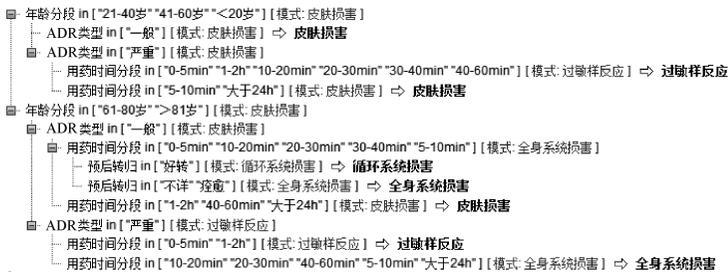
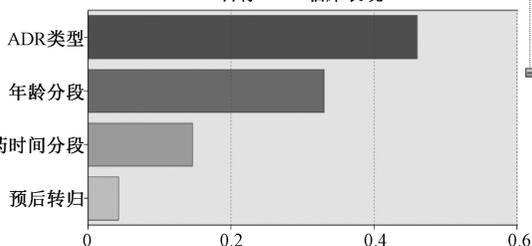


图2 CART 决策树算法预测变量的重要性及决策树模型

Fig 2 Importance of predictive variables of CART decision tree algorithm and decision tree model

3 讨论

3.1 梧州地区中药注射剂致 ADR 的流行病学特点

梧州地区中药注射剂致 ADR 发生情况复杂, ADR 的临床表现与患者年龄、性别、过敏史、原发疾病、联合用药情况、ADR 发生时间、给药途径以及预后转归情况等多种因素相关, 临床应引起重视。根据贝叶斯网络分析与决策树算法可以得到梧州地区中药注射剂致 ADR 的流行病学特点: 患者年龄、ADR 类型、ADR 发生时间与 ADR 临床表现的关系最紧密。

3.2 贝叶斯网络研究结果分析

由表 1 可知, ≤20 岁的患者中, 中药注射剂引起皮肤损害的发生概率最大(男性: 0.29; 女性: 0.14), 其次为消化系统损害(男性: 0.19; 女性: 0.08) 和全身系统损害(男性: 0.11; 女性: 0.03); >20 岁的患者中, 尤其是>60 岁的患者, 皮肤损害出

预测变量重要性
目标:ADR临床表现

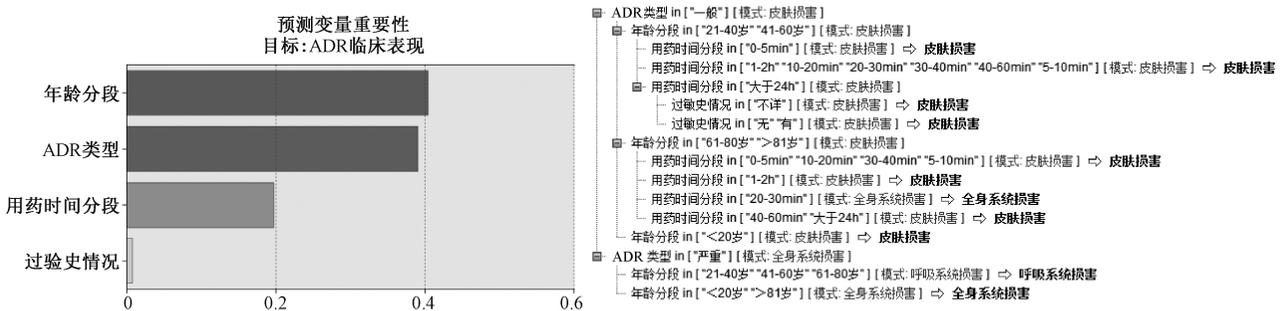


图3 CHAID 决策树算法预测变量的重要性以及决策树模型

Fig 3 Importance of predictive variables of CHAID decision tree algorithm and decision tree model

现的概率逐渐减少,循环系统损害、全身系统损害、消化系统损害、呼吸系统损害、运动系统损害和神经系统损害的发生概率逐渐增多。由表2可知,中药注射剂引起严重的/一般的皮肤损害在用药开始后>24 h的发生概率最大(严重的:0.45;一般的:0.24);严重的/一般的循环系统损害在用药开始后>40~60 min的发生概率最大(严重的:0.55;一般的:0.26);严重的过敏性休克在用药开始后0~5 min内的发生概率最大(为0.35);严重的/一般的过敏反应在用药开始后0~5 min内或用药开始后>24 h的发生概率最大(严重的:0.24;一般的:0.21);一般的全身系统损害在用药开始后>40~60 min的发生概率最大(为0.27);严重的/一般的呼吸系统损害在用药开始后>40~60 min的发生概率最大(严重的:0.22;一般的:0.27);严重的运动系统损害在用药开始后0~5 min内的发生概率最大(为0.67),一般的运动系统损害在用药开始后>40~60 min的发生概率最大(为0.29);严重的神经系统损害在用药开始后>5~10 min的发生概率最大(为0.67),一般的神经系统损害在用药开始后>40~60 min的发生概率最大(为0.41)。

3.3 决策树结果的分析与比较

3.3.1 决策树 CART 算法结果:中药注射剂引起的全部 ADR 报告中,第一分组因素为 ADR 的临床表现,以皮肤损害最明显;第二分组因素为发生 ADR 患者的年龄分段,在各年龄段,均为皮肤损害为主;第三分组因素为 ADR 的类型,中药注射剂引起的严重的 ADR 主要发生于为用药后 0~5 min 或用药后 1~2 h,以过敏样反应为主。

3.3.2 决策树 CHAID 算法结果:中药注射剂引起的全部 ADR 报告中,第一分组因素为 ADR 的临床表现,以皮肤损害最明显;第二分组因素为 ADR 的类型,一般的 ADR 中各年龄段均以皮肤损害为主,严重的 ADR 以全身系统损害、呼吸系统损害为主,其中 ≤20 和 >80 岁 2 个年龄段以全身系统损害为主,>20~40、>40~60 和 >60~80 岁 3 个年龄段发生的严重的 ADR 主要集中于呼吸系统损害。

3.3.3 两种决策树算法结果比较:决策树算法有许多种,在选择决策树算法时主要需考虑因变量的类型,而变量的类型包括连续型变量、有序分类变量和无序分类变量。本研究的因变量选择为中药注射剂所引起的 ADR 临床表现,其为分类变量,故选择能够处理分类变量的决策树 CART、CHAID 算法^[7]。通过比较两种算法的分组结果可以发现,CART 算法与 CHAID 算

法选择的第 1 个分组因素均为 ADR 的临床表现,提示 CHAID 算法与 CART 算法选择最重要分组因素的能力相同,提示梧州地区中药注射剂引起的 ADR 主要表现为皮肤损害。CART 算法与 CHAID 算法中的分组数相同,某些分组均可细分为 4 个小组,例如,CART 算法中>60~80 岁、>80 岁组,一般的皮肤损害有 4 个分组;而在 CHAID 算法中,一般的皮肤损害组中,20~40 岁、>40~60 岁患者有 4 个分组。在采用相同的自变量、因变量和决策树停止生长条件下,两种决策树算法得到了两种不同的 ADR 临床表现模型,经比较后可发现,CART 算法与 CHAID 算法的细分能力均适合 ADR 建模研究分析。

3.4 患者年龄构成情况分析

纳入本研究的 1 370 例使用中药注射剂发生 ADR 的患者中,男性平均年龄为 (49.13 ± 26.53) 岁,女性平均年龄为 (54.34 ± 21.65) 岁;>40 岁患者共 999 例,其中>60~80 岁患者共 482 例(占 48.25%)。数据挖掘结果显示,梧州地区中药注射剂致 ADR 的临床表现与患者年龄的关系较紧密。既往已有专门针对 ADR 与患者年龄分布分析评价的研究,如孙言才等^[8]曾对 2005—2012 年某医院 12 249 例发生 ADR 患者的年龄情况进行分析,结果显示,41~50 岁、51~60 岁和 61~70 岁等 3 个年龄段发生 ADR 患者数构成比居前 3 位。梧州地区中药注射剂致 ADR 的数据挖掘提示:(1)皮肤损害在各年龄段均表现出较高的发生率,>60 岁人群中皮肤损害所占比例逐渐降低,其原因可能与青壮年的血清免疫球蛋白 G 水平较老年人高,更易发生由免疫球蛋白 G 介导的过敏反应从而引发皮肤损害^[9-11]。(2)ADR 病例数以及严重程度均随着年龄升高而出现增加的趋势,如皮肤系统以外的其他系统损害比较集中在 20~40、>40~60 和 >60~80 岁 3 个年龄段,与国内相关研究数据一致^[11-12]。其原因为老年人随着年龄增长,肝肾功能均有不同程度的减退,影响了药物的代谢与排泄,易发生药物蓄积,导致药效过强而发生 ADR^[13]。同时,年龄的增长使得机体对药物的耐受能力降低,老年人较其他年龄段人群更容易出现 ADR^[14]。因此,本研究结果说明梧州地区上报的中药注射剂的 ADR 具有警戒信号提取意义,与>60 岁年龄段密切相关,提示临床应关注该年龄段患者的 ADR 监测工作。

综上所述,本研究基于贝叶斯网络预测梧州地区中药注射剂致 ADR 的影响因素的同时,通过对比决策树算法预测结果,

(下转第 108 页)