

论著·临床研究

doi:10.3969/j.issn.1671-8348.2022.07.016

基于 BP 神经网络、随机森林和决策树建立早期慢性乙型病毒性肝炎肝硬化无创诊断模型*

唐艳芳,刘旭东[△],吕萍,林海,温映华

(广西中医药大学附属瑞康医院肝病科,南宁 530000)

[摘要] **目的** 探讨 BP 神经网络模型、随机森林模型和决策树 CHAID 算法模型对早期慢性乙型病毒性肝炎肝硬化的诊断价值。**方法** 收集 2018 年 1 月至 2019 年 8 月在该院住院且行肝组织活检的 106 例慢性乙型肝炎病毒(HBV)感染者的临床资料,包括人口学、中医四诊资料、生化指标、肝脏硬度值(LSM)及肝组织病理学结果等参数,采用 SPSS18.0 软件进行非参数检验和 χ^2 检验,经单因素分析,筛选出对早期肝硬化有统计学差异的指标作为自变量,以有无肝硬化作为因变量,分别建立 BP 神经网络、随机森林和决策树 CHAID 算法模型,通过计算 3 个模型的正确率、错误率、混淆矩阵、灵敏度、特异度、约登指数、阳性预测值、阴性预测值、受试者工作特征(ROC)曲线下面积(AUC)值等来评价和比较 3 种模型的优劣。**结果** 成功建立了早期慢性乙型病毒性肝炎肝硬化无创诊断模型,在训练组样本和测试组样本中,随机森林算法模型预测正确率(84%,100%)、灵敏度(0.813,1.000)、约登指数(0.701,1.000),ROC AUC(0.896,1.000)均高于 BP 神经网络算法模型和决策树 CHAID 算法模型。**结论** 基于随机森林算法的诊断模型建立可以准确判断早期慢性乙型肝炎肝硬化,其预测能力优于 BP 神经网络模型算法模型和决策树 CHAID 算法模型,具有良好的临床应用价值。

[关键词] 乙型肝炎肝硬化;BP 神经网络;随机森林;决策树 CHAID 算法;无创性诊断**[中图分类号]** R512.6**[文献标识码]** A**[文章编号]** 1671-8348(2022)07-1161-06

Establishment and comparison of non-invasive diagnostic models for early chronic viral hepatitis B cirrhosis based on BP neural network, random forest and decision tree*

TANG Yanfang, LIU Xudong[△], LYU Ping, LIN Hai, WEN Yinghua

(Department of Hepatology, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning, Guangxi 530000, China)

[Abstract] **Objective** To explore the diagnostic value of BP neural network model, random forest model and decision tree CHAID algorithm model for early stage chronic viral hepatitis type B (HBV) cirrhosis. **Methods** The clinical data of 106 chronic HBV-infected patients who were hospitalized in the hospital and had liver tissue biopsy from January 2018 to August 2019 were collected, including parameters such as demographic and TCM four-diagnostic data, biochemical indicators, liver stiffness (LSM) values and liver histopathological findings. Non-parametric tests and chi-square tests were performed by using SPSS18.0 software. After univariate analysis, the indicators that were statistically significant for early cirrhosis were screened as independent variables, and the presence or absence of cirrhosis was used as the dependent variable. BP neural network, random forest and decision tree CHAID algorithm diagnostic models were established respectively. The advantages and disadvantages of the three models were evaluated and compared by calculating the correct rate, error rate, confusion matrix, sensitivity, specificity, Youden index, positive predictive value, negative predictive value and area value receiver operating characteristic (ROC) curve (AUC) of the three models. **Results** A non-invasive diagnostic model for early stage chronic viral hepatitis type B cirrhosis was successfully established. The correct prediction rate (84%, 100%), sensitivity (0.813, 1.000), Youden index (0.701, 1.000) and AUC (0.896,

* 基金项目:2019 年中医药传承与创新人才培养平台建设项目(全国名老中医药专家周培郁传承工作室)(国中医药人教函[2019]41 号);广西中医药大学“歧黄工程”高层次人才团队培育项目(中西医结合防治肝病创新团队)(桂中医大人[2021]10 号);广西卫生厅自筹经费科研项目(Z20180885)。作者简介:唐艳芳(1984—),副主任医师,硕士,主要从事中西医结合防治慢性肝病研究。△ 通信作者, E-mail: lxdlxh@163.com。

1.000) of the random forest model were higher than those of the BP neural network algorithm model and the decision tree CHAID algorithm model in both the training group samples and the test group sample. **Conclusion** The random forest diagnostic model can accurately determine early chronic HBV cirrhosis. Its predictive ability is better than that of the BP neural network algorithm model and the decision tree CHAID algorithm model, and has good clinical application value.

[Key words] hepatitis B cirrhosis; BP neural network; random forest; decision tree CHAID algorithm; non-invasive diagnosis

乙型肝炎病毒(hepatitis B virus, HBV)感染仍是全球重大公共卫生问题。据世界卫生组织(WHO)报道,全球约有 2.57 亿慢性 HBV 感染者,非洲地区和西太平洋地区占 68%^[1]。全球每年约有 88.7 万人死于 HBV 感染相关疾病,其中肝硬化和原发性肝癌死亡分别占 52%和 38%^[2]。肝硬化是由不同病因引起的慢性、进行性、弥漫性肝病,是各种慢性肝病发展的晚期阶段^[3],早期积极治疗可有效控制疾病进展,延缓和阻止病情进展为肝硬化失代偿期、肝癌及其他并发症的发生,因此,早期诊断肝硬化对于患者的治疗决策、疗效评估及预后都有非常重要的临床意义。由于早期肝硬化患者往往缺乏典型的临床表现、体征,部分患者肝功能可一直正常,诊断较为困难,容易导致漏诊,延误了最佳治疗时机。目前早期肝硬化的诊断金标准仍然是肝组织活检,但由于获取活组织时的抽样误差及观察者的不同也可能出现分期不准确的情况^[4]。因此,临床上迫切需要寻找 1 种简单、易行、精确的非侵入性手段来替代肝组织活检。随着科学技术的不断发展,数据挖掘技术广泛应用于医学领域,在疾病诊断、预后判断、风险评估等方面具有良好的应用价值^[5-7]。本研究收集患者基本人口学资料、中医四诊资料、血清生化指标、肝脏硬度值(LSM)等参数,以病理学结果为金标准,分别采用 BP 神经网络、随机森林和决策树 CHAID 算法,构建 1 套早期慢性乙型肝炎肝硬化无创诊断模型,并比较 3 种模型的预测效果,以期达到高效、无创地诊断早期肝硬化的目的,为临床提供简便、可靠的无创肝硬化评估方法。

1 资料与方法

1.1 一般资料

选取 2018 年 1 月至 2019 年 8 月在本院住院的慢性乙型肝炎患者 106 例,年龄 18~59 岁,平均(39.87±10.36)岁,其中,男 83 例,女 23 例,病程 1~22 年。所有患者在知情同意的情况下均接受肝组织穿刺活检术,并进行病理学分期。按照病理诊断肝硬化的分期结果,将患者分为 2 组,分别为非肝硬化组(肝纤维化 S0~S3 期)和肝硬化组(肝纤维化 S4 期)。其中非肝硬化组 49 例(46.2%),年龄 20~53 岁,平均(36±9)岁;肝硬化组 57 例(53.8%),年龄 18~59

岁,平均(43±10)岁。2 组患者性别、年龄差异无统计学意义($P>0.05$),具有可比性。慢性乙型肝炎诊断参考中华医学会肝病学分会和中华医学会感染病学分会共同制订的《慢性乙型肝炎防治指南(2019 年更新版)》^[2]。排除标准:(1)拒绝加入本课题研究,未签署知情同意书的患者;(2)同时合并其他非嗜肝病毒感染者;(3)失代偿期肝硬化、肝肿瘤或合并其他脏器肿瘤者;(4)肝活检标本长度不足或临床资料不全的患者。

1.2 中医辨证分型

参照 2017 年中国中西医结合学会消化系统疾病专业委员会制订的《肝纤维化中西医结合诊疗共识意见》^[8],结合患者临床详细四诊资料,制订临床常见的 3 种肝纤维化主要证型——肝胆湿热证、肝郁脾虚证、痰瘀互结证。

1.3 调查方法

在多次预调查的基础上,结合临床诊断自行制订《慢性乙型肝炎患者个案调查表》,详细记录患者的一般人口学资料、症状、体征、中医四诊资料、穴位信息、中医辨证分型、生化指标、病毒学标志物及肝脏硬度值(LSM)等数据。研究变量如下:(1)中医四诊资料;(2)症状和体征:疲倦乏力、纳差、肋肋疼痛、肝掌、蜘蛛痣、舌有瘀斑、肝俞压痛;(3)LSM 值;(4)生化指标及病毒学标志物:Ⅲ型前胶原(PCⅢ)、透明质酸酶(HA)、Ⅳ型胶原(CⅣ)、层粘连蛋白(LN)、血小板计数(PLT)、胆碱酯酶(CHE)、总胆红素(TBIL)、清蛋白(ALB)、球蛋白(GLOB)、丙氨酸氨基转移酶(ALT)、天门冬氨酸氨基转移酶(AST)、甲胎蛋白(AFP)、纤维蛋白原(FIB)、HBV DNA(HBV-DNA)水平、乙型肝炎 e 抗原定量(HBeAg)、HBV 表面抗原定量(HBsAg)。应用美国 RXL 全自动生化分析仪检测相关生化指标。检测项目均由本院检验科专业人员严格按照试剂盒说明书操作。

1.4 肝脏组织学病理检查

采用超声引导下经皮肝穿刺活检,肝组织学检查及病理阅片均由 1 名经验丰富的主治医师独立完成,肝组织长度大于或等于 1.5 cm,至少包括 6 个以上汇管区。依据 Scheuer 分期标准,纤维化程度分为:S0,无纤维化;S1,汇管区纤维化扩大,局限窦周及小叶内

纤维化;S2,汇管区周围纤维化,纤维间隔形成,小叶结构保留;S3,纤维间隔伴小叶结构紊乱,无肝硬化;S4,早期肝硬化。规定 S0~S3 为非肝硬化组,S4 为肝硬化组。

1.5 LSM 值测定

运用 FibroTouch(FT-C 型,无锡海斯凯尔医学技术有限公司)测定 LSM 值,患者取平卧位,双上肢置于头顶,双下肢向外侧平移,使身体成弓形,肋间隙增宽,选择右侧腋前线至腋中线第 7~9 肋间为检测点。使用 B 超探头二维成像,避开胆囊、大血管、胆管、囊肿等组织,通过专用探头固定于检测区域,发出固定低频率的剪切波,对肝脏组织实施瞬时主动激励,同时发出高频信号追踪剪切波在肝脏内的传播并计算传播速度,算出肝脏的 LSM 值,要求成功检测 10 次,四分位差小于 1/10 中位数,成功率为 100%,取中位数为检测结果,用弹性值(kPa)表示。上述操作均为同一位医师操作。

1.6 统计学处理

采用 Epidata3.0 软件建立数据库,双人双录入。应用 IBM SPSS18.0 软件进行统计分析。计数资料以频数表示,比较采用 χ^2 检验;不符合正态分布的连续性变量采用中位数和四分位间距 [$M(P_{25}, P_{75})$] 表示,比较采用非参数检验进行分析,采用单因素分析筛选出差异有统计学意义的变量分别建立 BP 神经网络算法模型,随机森林算法模型和决策树 CHAID 算法模型。通过随机样本分割节点,将患者分为训练集和测试集,按照机器学习的常规设置,训练集为 80%,测试集为 20%。模型具体设置:(1)BP 神经网络算法模型:采用多层感知器。隐含层设置为自动计算,最大训练时间为 15 min,最大训练周期数量为 250,最低准确性为 90%。先进行初始 BP 神经网络算法模型的建立。根据所收集的病历资料数据结构,确定 BP 神经网络算法模型基本框架,以经过自变量筛选得到的变量作为 BP 神经网络算法模型的输入变量,以有或无肝硬化作为输出变量。BP 神经网络算法模型的输入层层数与输入变量的个数相同;通过试凑法来确

定隐含层节点个数,将 BP 神经网络的隐含层个数范围设置为(10,60),分别进行 50 次重复试验,记录每一次试验中的网络模型误差率,比较其中位数大小,根据隐含层节点数使 BP 神经网络算法模型的分类错误率误差最低时选择隐含层节点数。(2)CHAID 决策树算法模型:分割显著性水准 α 设置为 0.05,树的生长层数为 3 层,停止规则 $\alpha=0.05$,母、子母节点最小样本量分别设为 50、10,并产生 CHAID 决策树分析模型的乙型病毒性肝炎肝硬化预测概率。(3)随机森林算法模型:在 Rstudio 软件中,通过调用 randomforest 程序包来进行随机森林分类识别模型的构建。randomforest 函数在进行模型构建过程中主要包含 mtry(随机森林算法中用于设定决策树分支的变量个数)和 ntree(随机森林算法中决策树的数目)2 个模型参数。数据集中变量个数开平方根所得数值,通常作为参数 mtry 的取值方式,本研究随机森林算法模型的输入变量个数为 106,因此将参数 mtry 默认设定为 10;ntree 在通常情况下设定为 500,即随机森林算法中默认存在 500 棵树。根据上述默认参数,建立初始随机森林算法模型,在测试集中对分类模型效果进行评价,并选出最优参数建立最终模型。为了比较不同模型的诊断价值,通过计算 3 个模型的正确率、错误率、混淆矩阵、灵敏度、特异度、约登指数、阳性预测值、阴性预测值、受试者工作特征(receiver operating characteristic,ROC)曲线下面积(AUC)值等比较模型的优劣。以 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 血清学指标与肝硬化的单因素分析

由于 PCⅢ、CⅣ、LN、HA、HBV-DNA 等连续性变量不符合正态分布,采用非参数检验进行分析,结果显示:CⅣ、LN、HA、GLOB、PLT、FIB、ALT、AST、AFP、HBsAg、HBeAg 和 LSM 值与肝硬化有关,差异有统计学意义($P<0.05$),可进入之后的模型进行分析。PCⅢ、CHE、HBV-DNA、TBIL 和 ALB 在组间差异无统计学意义($P>0.05$),予以排除。见表 1。

表 1 106 例患者血清学指标的单因素分析 [$M(P_{25}, P_{75})$]

项目	肝硬化组 ($n=57$)	非肝硬化组 ($n=49$)	曼-惠特尼 U	P
PCⅢ (ng/mL)	15.0(10.1,37.1)	30.3(10.2,59.1)	1 244	0.365
CⅣ (ng/mL)	114.7(79.0,145.3)	54.3(42.0,61.6)	404	<0.001
LN(ng/mL)	124.9(98.7,173.7)	61.5(45.4,87.8)	472	<0.001
HA(ng/mL)	113.5(65.0,161.1)	56.1(42.3,71.9)	635	<0.001
HBV-DNA(IU/mL)	5 330.0(102.8,48 225.0)	6 470.0(10.0,4 360 000.0)	1 250	0.381
ALB(g/L)	43.3(39.4,45.5)	43.8(41.0,46.2)	1 312	0.636
GLOB(g/L)	29.8(27.4,32.8)	27.7(25.4,30.4)	951	0.006
PLT($\times 10^9/L$)	147.0(112.8,178.0)	218.0(172.0,245.0)	487	<0.001

续表 1 106 例患者血清学指标的单因素分析[$M(P_{25}, P_{75})$]

项目	肝硬化组($n=57$)	非肝硬化组($n=49$)	曼-惠特尼 U	P
FIB(g/L)	1.9(1.7,2.2)	2.2(2.0,2.8)	746	<0.001
ALT(U/L)	45.0(31.8,68.0)	25.0(20.0,46.0)	918	0.003
AST(U/L)	39.0(28.0,57.3)	24.0(18.0,33.0)	730	<0.001
TBIL(μ mol/L)	11.2(7.9,15.1)	11.9(9.1,14.0)	1 295	0.558
AFP(ng/mL)	7.1(3.7,19.1)	3.1(2.4,5.2)	754	<0.001
CHE(U/L)	7 499.5(5 617.8,8 802.3)	8 313.0(7 033.0,8 934.0)	1 094	0.063
HBsAg(COD)	6 239.5(5 482.0,6 924.0)	3 649.0(1 747.5,8 210.0)	1 019	0.019
HBeAg(COD)	0.1(0.1,0.9)	0.1(0,7.0)	988	0.011
LSM 值(kPa)	13.2(9.1,16.7)	7.8(6.8,9.3)	626	<0.001

2.2 临床特征与肝硬化的单因素分析

χ^2 检验结果显示,中医四诊资料、肝掌、舌有瘀斑、疲倦乏力、纳差和肝俞压痛,差异有统计学意义($P<0.05$),可进入之后的模型进行分析,蜘蛛痣和肋肋疼痛差异无统计学意义($P>0.05$),予以排除,见表 2。

2.3 BP 神经网络、随机森林和决策树 CHAID 算法模型的混淆矩阵、正确率和错误率比较

本文采用混淆矩阵及整体正确率、错误率对模型进行评价,混淆矩阵结果见表 3。在训练集中,3 个模型正确率均较高,均在 90% 以上,随机森林算法模型更是达到了 100% 的正确率。进一步推广到测试集中,最优的模型是随机森林算法模型,正确率为 84%,其次为 BP 神经网络算法模型,正确率为 80%,再次为 CHAD 决策树算法模型,正确率为 72%。见表 4。

2.4 3 种预测模型性能比较

在同一测试集中,对比 BP 神经网络、随机森林和决策树 CHAID 3 种算法模型的分​​类效能,随机森林算法模型在真正患者识别早期肝硬化效果方面优于 BP 神经网络算法模型和决策树 CHAID 算法模型,随机森林模型的灵敏度(0.813)、特异度(0.889)、约登指数(0.701)、阳性预测值(0.929)、阴性预测值(0.727)及 AUC(0.896)均高于 BP 神经网络算法模型和决策树 CHAID 算法模型。由此得出,随机森林算法模型的预测效果最为理想,其诊断价值高于 BP 神经网络算法模型和决策树 CHAID 算法模型。见表 5。

表 2 106 例患者的临床特征比较(n)

项目	肝硬化组($n=57$)	非肝硬化组($n=49$)	χ^2	P
中医四诊			16.080	<0.001
肝郁脾虚证	15	27		
肝胆湿热证	19	15		
痰瘀互结证	25	5		
肝掌			9.048	0.003
有	23	6		
无	36	41		
蜘蛛痣			0.588	0.443
有	15	9		
无	44	38		
舌有瘀斑			26.635	<0.001
有	47	14		
无	12	33		
疲倦乏力			17.719	<0.001
有	43	15		
无	16	32		
纳差			15.865	<0.001
有	38	12		
无	21	35		
肋肋疼痛			2.809	0.094
有	36	21		
无	23	26		
肝俞压痛			7.941	0.005
有	40	19		
无	19	28		

表 3 3 种模型在训练集和测试集的混淆矩阵(行为实际值)

模型	BP 神经网络				随机森林				CHAD 决策树			
	训练集($n=81$)		测试集($n=25$)		训练集($n=81$)		测试集($n=25$)		训练集($n=81$)		测试集($n=25$)	
	无	有	无	有	无	有	无	有	无	有	无	有
无	35	1	8	3	36	0	8	3	36	0	8	3
有	6	39	2	12	0	45	1	13	2	43	4	10

无:非肝硬化组;有:肝硬化组。

表 4 3 种模型在训练集和测试集的整体正确率和错误率[n(%)]

模型	BP 神经网络		随机森林		CHAID 决策树	
	训练集	测试集	训练集	测试集	训练集	测试集
正确	74(91.36)	20(80.00)	81(100)	21(84.00)	79(97.53)	18(72.00)
错误	7(8.64)	5(20.00)	0	4(16.00)	2(2.47)	7(28.00)
合计	81(100)	25(100)	81(100)	25(100)	81(100)	25(100)

表 5 3 种预测模型评价指标体系比较

模型	BP 神经网络		随机森林		决策树 CHAID	
	训练集	测试集	训练集	测试集	训练集	测试集
灵敏度	0.975	0.800	1.000	0.813	1.000	0.769
特异度	0.875	0.800	1.000	0.889	0.947	0.667
约登指数	0.850	0.600	1.000	0.701	0.947	0.436
阳性预测值	0.886	0.857	1.000	0.929	0.956	0.714
阴性预测值	0.972	0.727	1.000	0.727	1.000	0.727
AUC	0.961	0.870	1.000	0.896	0.996	0.737
Gini	0.922	0.740	1.000	0.792	0.993	0.474

3 讨 论

在慢性肝病的进程中,肝硬化的早期诊断和评估,对慢性 HBV 肝炎患者选择合适的治疗策略和预测预后具有重要的临床意义^[9]。肝组织活检仍是评估肝纤维化程度的“金标准”,但其具有重复性差、侵入性、观察者内和观察者间的变异及易伴出血、疼痛等风险的缺点^[10-11],这使得早期肝硬化的诊断相对滞后。近年来,基于多项临床生物化学指标及影像学检查建立的肝纤维化无创性诊断模型受到越来越多的学者关注,这些模型对肝纤维化的诊断具有一定的临床运用价值,但以慢性 HBV 肝炎为研究对象所建立的模型较少且尚不成熟,国内学者提出的相关研究主要包括:上海肝纤维化组(SLFG)模型^[12]、Fibromodel^[13]、FibroIndex^[14]及 S 指数^[15-16]等。这些模型所纳入的部分参数在临床上不易获取,未经临床广泛验证,临床推广应用则受到一定的限制。随着临床医学研究数据量剧增,数据形式趋于复杂化、多样化,如何从海量的数据中获取更有价值和意义的信息,是当前医学研究发展的难点及热点问题。显然,传统的数据分析方法和工具已不能满足当前的需要,针对海量、复杂的医学数据,机器学习算法显示了明显的优势,比如:BP 神经网络算法、决策树 CHAID 算法可处理非线性关系的预测模型^[17-18],随机森林模型的稳健性较高,可以高效地处理大数据集^[19-20]。本研究以肝组织活检诊断结果为金标准,通过收集 106 例慢性 HBV 肝炎患者的临床症状、体征、中医四诊资料、生化指标、LSM 及穴位信息资料,分别建立 BP 神经网络、随机森林和决策树 CHAID 算法模型用于预测乙型肝炎病毒性肝炎早期肝硬化,尽可能选择最优的参数建

立模型,以更好地提升预测效果。通过对比研究,发现随机森林算法模型的预测性能高于 BP 神经网络算法和决策树 CHAID 算法模型,显示出了良好的优势,其中在训练集中的正确率为 100%,测试集中的正确率为 84.00%,AUC 为 0.896,这可能与随机森林算法的随机特征选择思想较其他分类器(如判别分析、支持向量机、神经网络等)表现得更为出色有关,并且它在处理过拟合问题上也非常得心应手^[21]。

本研究利用曼-惠特尼 U 检验和 χ^2 检验筛选出的有效变量有 C IV、LN、HA、GLOB、PLT、FIB、ALT、AST、AFP、HBsAg、HBeAg、LSM、中医四诊资料、肝掌、舌有瘀斑、疲倦乏力、纳差和肝俞压痛,这些危险因素与目前的临床、指南相符合,还提供一些对临床有重要意义的信息,值得进一步探究。另外,在本研究中医证型诊断也可以用于早期乙型肝炎病毒性肝炎肝硬化的预测模型中,这给临床提供了一个实用性的新思路。

综上所述,本研究采用的随机森林算法模型在判别早期乙型肝炎病毒性肝炎肝硬化研究中初步显示出较好的诊断效能和稳定性,在一定程度上减少了肝组织活检的必要性,对指导临床治疗和随访病情转归具有现实意义。相比于其他肝纤维化非创伤性诊断模型,本研究的预测因素更为全面,包括症状、体征、生化指标、LSM 值、中医证型、舌脉及穴位信息等,涉及范围大,可充分挖掘数据信息,模型更为完善,在这个数据量不断膨胀的时代,显示了其独特的优势。另外,研究采用多种模型对早期肝硬化进行预测,并比较各模型的预测效能,对临床的指导意义巨大。本研究也存在一些不足之处,首先本研究为单中心研究,所纳入

的患者数较少,可能存在一定的发表偏倚。其次,本研究并未对肝纤维化程度进行细分,只进行了明显肝纤维化和肝硬化的预测分析,属于 0-1 二分类数据集,这就使得本次研究对肝纤维化的具体分期预测不够明确,下一步将使用更大、更全面的样本集对本文中的模型进行验证和改进,从而建立更加完善的分类模型。

参考文献

- [1] WHO. Global hepatitis report, 2017[EB/OL]. [2021-06-06]. <https://www.who.int/hepatitis/publications/global-hepatitis-report2017/en>.
- [2] 中华医学会肝病学会,中华医学会感染病学分会. 慢性乙型肝炎防治指南(2019 年版)[J]. 临床肝胆病杂志,2019,35(12):2648-2669.
- [3] FERNANDEZ J,BAUER T M,NAVASA M, et al. Diagnosis, treatment and prevention of spontaneous bacterial peritonitis[J]. *Baillieres Best Pract Res Clin Gastroenterol*, 2000, 14: 975-990.
- [4] CALDWELL S,LACKNER C. Perspectives on NASH histology: cellular ballooning[J]. *Ann Hepatol*, 2017, 16: 182-184.
- [5] 吴伟,郭军巧,安淑一,等. 使用思维进化算法优化的神经网络建立肾综合征出血热预测模型[J]. *中国卫生统计*, 2016, 33(1): 27-30.
- [6] YE H R,YANG Y,LIN X, et al. Application of BP neural network in the diagnosis of breast cancer with high frequency color Doppler ultrasound[J]. *Chin J Health Stat*, 2016, 33(1): 71-72.
- [7] TSENG W T,CHIANG W F,LIU S Y, et al. The application of data mining techniques to oral cancer prognosis[J]. *J Med Syst*, 2015, 39(5): 59.
- [8] 李军祥,陈諳,姚树坤. 肝纤维化中西医结合诊疗共识意见(2017 年)[J]. *中国中西医结合消化杂志*, 2017, 25(12): 895-900.
- [9] LEE S,KIM D Y. Non-invasive diagnosis of hepatitis B virus-related cirrhosis[J]. *World J Gastroenterol*, 2014, 20(2): 445-459.
- [10] Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group[J]. *Hepatology*, 1994, 20(1): 15-20.
- [11] 胡锡琪. 肝穿刺活组织检查是评估肝纤维化的金标准吗? [J]. *中华肝脏病杂志*, 2012, 20(8): 568-570.
- [12] ZENG M D,LU L G,MAO Y M, et al. Prediction of significant fibrosis in HBeAg-positive patients with chronic hepatitis B by a noninvasive model[J]. *Hepatology*, 2005, 42(6): 1437-1445.
- [13] HUI A Y,CHAN H L,WONG V W, et al. Identification of chronic hepatitis B patients without significant liver fibrosis by a simple noninvasive predictive model[J]. *Am J Gastroenterol*, 2005, 100(3): 616-623.
- [14] KODA M,MATUNAGA Y,KAWAKAMI M, et al. FibroIndex, a practical index for predicting significant fibrosis in patients with chronic hepatitis C[J]. *Hepatology*, 2007, 45(2): 297-306.
- [15] 周琨,郑瑞丹,咸建春,等. 从常规指标中建立肝纤维化非创伤性诊断模型[J]. *肝脏*, 2008, 13(5): 362-367.
- [16] 郑瑞丹,周琨,咸建春,等. 慢性乙型肝炎病毒感染患者肝纤维化非创伤性诊断模型的应用评价[J]. *肝脏*, 2008, 13(6): 451-455.
- [17] WANG C H,MO L R,LIN R C, et al. Artificial neural network model is superior to logistic regression model in predicting treatment outcomes of interferon-based combination therapy in patients with chronic hepatitis C[J]. *Intervirology*, 2008, 51(1): 14-20.
- [18] 宋健,吴学森,张杰,等. 三种统计学模型在糖尿病个体患病风险预测中的应用[J]. *中国卫生统计*, 2017, 34(2): 312-314.
- [19] 明均仁,肖凯. 基于 R 语言的面向需求预测的随机森林方法[J]. *统计与决策*, 2012, 28(9): 81-83.
- [20] 阚红星,张璐瑶,董昌武. 1 种 2 型糖尿病中医证型的舌图像识别方法[J]. *中国生物医学工程学报*, 2016, 35(6): 658-664.
- [21] YANHONG L,ZHI L,HUSHENG G, et al. Predicting congenital heart defects: a comparison of three data mining methods[J]. *PLoS One*, 2017, 12(5): e0177811.