

# 基于机器学习模型分析镇肝熄风汤治疗高血压病的配伍特征研究

郇家铭<sup>1</sup>, 陈晓晴<sup>2</sup>, 杨雯晴<sup>3</sup>, 李洁<sup>3</sup>, 滑振<sup>4</sup>, 王怡斐<sup>4</sup>, 李运伦<sup>1,4,5\*</sup>

1. 山东中医药大学第一临床医学院, 济南 250355
2. 济南市高新东区医院心内科, 济南 250101
3. 山东中医药大学中医药创新研究院, 济南 250355
4. 山东中医药大学附属医院心病科, 济南 250014
5. 心血管病中医精准诊疗山东省工程研究中心, 济南 250355

**摘要** 以2011—2019年山东中医药大学附属医院的高血压电子病历建立数据集,用Apriori算法量化镇肝熄风汤组成中药的配伍强度,卷积神经网络量化中药的剂量特征信息,蛋白质相互作用的拓扑特征分析生物特征信息,并将上述信息输入K最邻近分类算法、支持向量机、梯度递增决策树、贝叶斯网络和逻辑回归模型,评估各个模型效力,从多个维度体现镇肝熄风汤的组成配伍规律和作用机制。结果表明,K最邻近分类算法在5个模型中结果最优(AUC=93.5%),共获得了87组有效配伍,验证了镇肝熄风汤配伍调节细胞因子、减少炎症反应和代谢紊乱的作用机制。同时,外部测试表明此研究可外推至其他疾病,表明此模型可有效融合卷积神经网络数据和网络拓扑数据,将剂量信息和生物特征融入中药配伍挖掘,对传统的关联规则模型进行补充,具有良好的普遍适用性和外推性,为多模态的中药数据集挖掘提供了一种跨学科研究方法。

**关键词** 机器学习; 真实世界数据; 数据挖掘; 镇肝熄风汤; 高血压

高血压等慢性疾病每年导致全球近1000万人死亡,超过2亿人残疾,血压的异常升高导致血管受损,进而损害脑、心、肾脏等靶器官,患者常出现头痛、眩晕、心悸、乏力、水肿、夜尿增多等症状,且

收稿日期:2023-11-02;修回日期:2024-08-30

基金项目:山东省中西医结合高血压诊疗项目(2019-11);济南市“高校20条”项目(2020GXRC017);山东省自然科学基金重点项目(ZR2020KH034)

作者简介:郇家铭,博士研究生,研究方向为中医药治疗心血管疾病,电子信箱:huanjiaming001@126.com;李运伦(通信作者),教授,主任医师,研究方向为中医药治疗心血管疾病研究,电子信箱:li.yunlun@163.com

引用格式:郇家铭,陈晓晴,杨雯晴,等.基于机器学习模型分析镇肝熄风汤治疗高血压病的配伍特征研究[J].科技导报,2024,42(21):149-162;doi:10.3981/j.issn.1000-7857.2024.05.00473

逐渐加重,甚至诱发中风、心肌梗死、心力衰竭和终末期肾病等高危疾病<sup>[1-2]</sup>,严重降低生活质量。中医治疗强调整体观念,不仅能有效降低血压,还强调改善不适症状,副作用较少,长期服用较为安全,能够有效减少高剂量化学药物带来的副作用。此外,天然药物可显著减少患者的恐惧和抵触心理,提高依从性,其疗效已获得临床认可<sup>[3-7]</sup>。

镇肝熄风汤作为经典方剂,已广泛应用于高血压的临床治疗,有丰富的长期临床数据,其组成包括牛膝、赭石、龙骨、牡蛎、龟甲、白芍、玄参、天冬、川楝子、麦芽、茵陈和甘草。根据多项临床研究和系统评价报道<sup>[8-10]</sup>,在化学药物治疗的基础上联合镇肝熄风汤有助于加强降压效果,改善患者不适症状,提高患者依从性,实现降低心脏、脑、肾脏及血管并发症和死亡的总风险<sup>[11-14]</sup>。中医学是以“天人合一”为导向的医学模式,要求中药处方能够兼顾患者的不适症状,并强调“君、臣、佐、使”的剂量变化,具有繁杂的配伍规则<sup>[15-16]</sup>。以真实数据为基础充分进行数据挖掘,可贴合中医诊疗中的随证加减规律,充分体现诊疗中个体化和药物应用特征<sup>[17-19]</sup>。

数据挖掘是中医科研领域的有力工具之一,中医数据挖掘主要依靠临床医疗数据和典籍数据,围绕分析中药功效、挖掘“方-药-证”关系、总结核心药对/类方、创新方剂配伍等方面开展,以期挖掘中医临床个体诊疗的宏观规律、辅助临床诊疗和老中医经验传承。应用聚类分析、因子分析、复杂网络分析等无监督的方法可有效处理结构化的中药处方数据,匹配临床的方、药、证特征,但是中药配伍剂量的研究仍然不足,处方剂量依旧需要青年医师进行长年累月的临床摸索。在如今生命科学与中医学广泛交融的时代,组学技术日益推广,有监督学习或半监督结合规则的研究方式逐渐成为主流,中药疗效的机制研究逐渐深入<sup>[20-21]</sup>,已有研究对中药的调控机制与临床症状进行初步映射<sup>[22-23]</sup>。进一步开发模型框架,构建多尺度、多空间模型,将微观机制与临床配伍数据从多个维度结合,挖掘隐藏知识和规则,解释中医理论指导下药物配伍和剂量加减的合理性,可为临床决策提供多层次参考,增加中医药的临床应用价值<sup>[24]</sup>。

机器学习模型可以整合中药的多维度信息,将输入特征与期望输出进行匹配,形成学习函数。机器学习中常用有监督分类器模型包括K最邻近(K-nearest neighbor, KNN)分类算法、支持向量机(support vector machine, SVM)、梯度递增决策树(gradient boosting decision tree, GBDT)、贝叶斯网络(bayesian network, BN)、逻辑回归(logistic regression, LR)。KNN对多维信息的局部特征更敏感,而SVM等则更能反映其全局特征,在中医药大数据挖掘和生物信息分析中有广泛应用<sup>[25]</sup>。卷积神经网络(convolutional neural network, CNN)作为一种人工神经网络,可以响应一定范围内的周围神经链接,对于输入数据的二维结构处理具有出色表现,同时相较于其他前馈神经网络需要考量的参数更少,更加适合中医处方特征处理<sup>[22]</sup>。

因此,本研究基于电子病历数据建立镇肝熄风汤配伍药物加权网络,使用CNN获取药物的配伍剂量梯度;基于生物信息数据建立蛋白质相互作用网络反映生物效应特征;使用机器学习算法搭建模型,输入镇肝熄风汤的临床与生物特征数据,挖掘其核心配伍,从多个角度讨论镇肝熄风汤在真实世界中应用规律。

## 1 材料和方法

### 1.1 数据准备

本研究数据来源于2011年7月11日至2019年10月22日山东中医药大学附属医院心内科住院和门诊电子病历,其中包括17800份高血压患者电子病历及处方(图1)。临床数据的运用获得山东中医药大学附属医院伦理委员会批准,批件号:(2023)伦审第(024)号-KY,并接受该委员会监督。

#### 1.1.1 纳入标准

(1)符合原发性高血压诊断标准;(2)年龄在18~85岁。

#### 1.1.2 排除标准

(1)妊娠和哺乳期妇女;(2)合并严重心脑血管疾病、恶性肿瘤、严重造血系统、呼吸系统、消化系统和感染性疾病患者;(3)病历资料不完整。

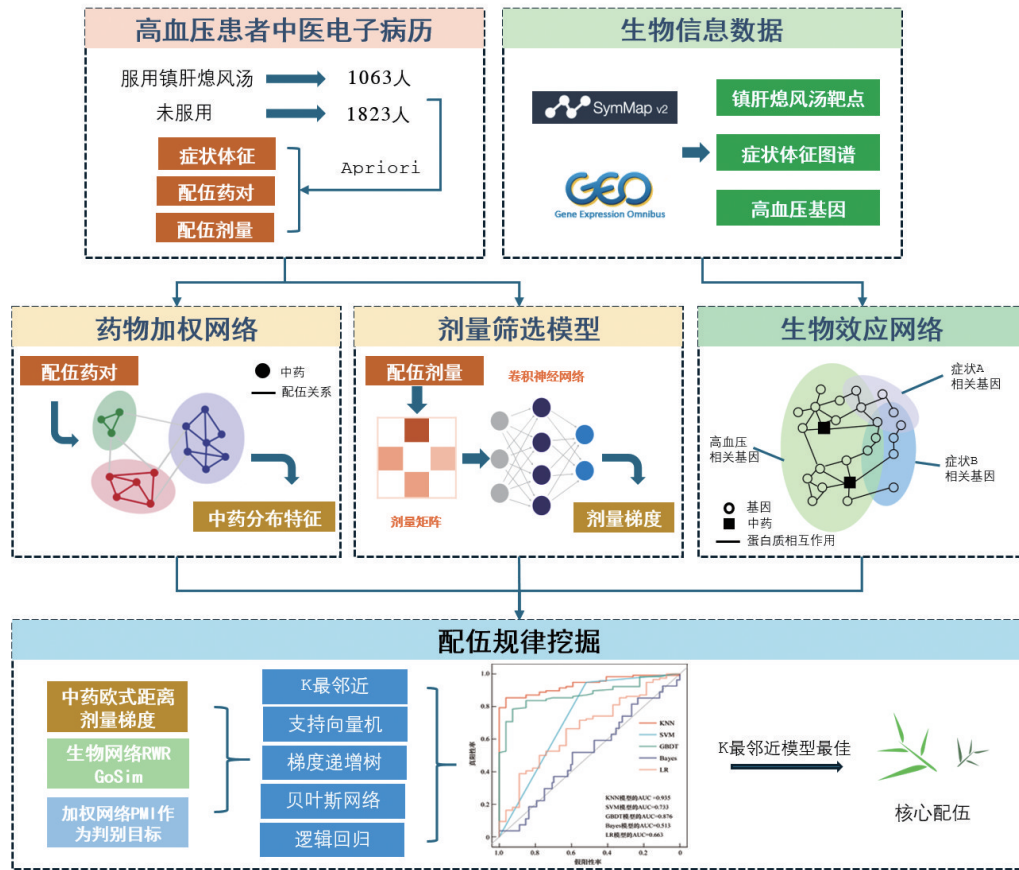


图1 研究流程

### 1.1.3 数据标准化

依据《中国药典》和美国国家医学图书馆的医学主题词(medical subject headings, MeSH)完成了中药处方和诊断的规范化,建立数据库。数据规范过程中,由2名具有15年以上诊疗经验的高级职称医师进行质量控制。

## 1.2 镇肝熄风汤配伍药物

### 1.2.1 发现中药配伍规律

Apriori算法作为经典的关联规则算法常用于发现元组间的相关性,本研究用于量化真实世界处方中镇肝熄风汤组成药物的配伍强度。根据既往研究<sup>[26-27]</sup>,保留支持度大于0.2且置信度大于0.5的配伍药对,在R语言(version 3.6.3)中使用`arules`包(version 1.6-6)实现。

### 1.2.2 构建镇肝熄风汤处方网络

本研究使用快速模块性优化算法(BGLL算法)挖掘临床处方中镇肝熄风汤的应用规律。以中药为

实体节点,以Apriori算法获得的药物间提升度(Lift)为连接权重,建立复杂关系网络,运用BGLL算法,形成镇肝熄风汤组成中药的加权网络社团划分。

## 1.3 配伍剂量模型的建立

本研究使用CNN研究镇肝熄风汤治疗高血压的中药配伍剂量,首先建立中药矩阵,以配伍剂量作为矩阵元素,使用CNN分析矩阵后形成剂量特征模型,用于研究高血压治疗中镇肝熄风汤配伍的剂量变化特征。

### 1.3.1 建立矩阵

以电子病历的中药处方作为数据来源,建立Apriori算法发现的配伍药物与镇肝熄风汤的联合矩阵。为消除不同中药常用剂量的差异,在本研究中使用剂量指数作为矩阵元素,即中药在处方中的剂量除以其在整个数据集中的平均剂量。从数据集中随机抽取80%作为训练集用于CNN模型训练,20%作为验证集。

### 1.3.2 模型训练

本研究基于 TensorFlow (version 2.5.0) 构建 CNN 模型,该模型包括 2 个卷积层,每个卷积层后连接 1 个最大池化层,池化层采用 2×2 滤波器,步幅设置为 2,并在模型末端添加 1 个全连接层。训练时批量大小设置为 32,迭代 50 次,卷积层和全连接层分别以 *relu* 和 *softmax* 为激活函数,并采用 Adam 优化器进行模型权重的更新,损失函数选择交叉熵损失。

模型的训练平台为搭载了 NVIDIA GeForce RTX 3070 GPU (CUDA 11.3) 的计算机。

## 1.4 生物网络的建立

### 1.4.1 数据收集

高血压相关基因收集自 Gene Expression Omnibus (GEO) 数据库内的 GSE24752 数据集,从中检索高血压基因表达数据,使用 GEO2R 工具进行在线分析,并使用公共数据库<sup>[28-30]</sup>进行补充。结合前期研究确定的高血压核心症状体征<sup>[27]</sup>,从 SymMap 中获取高血压症状体征的相关基因和镇肝熄风汤靶点,均使用 Uniport 数据库进行数据规范。

### 1.4.2 建立网络

将高血压相关基因和中药调控靶点输入 STRING<sup>[31]</sup> 平台的 Homo sapiens 数据库,利用其中实验数据、算法预测和文献挖掘等多个数据源的蛋白质相互作用,建立蛋白质-蛋白质相互作用网络。

## 1.5 中药配伍的效力评价

将真实世界数据的中药配伍与高血压的关联程度进行量化,评价不同配伍在临床实际和生物过程中与高血压的密切程度,再使用机器学习模型从各个层次进行汇总分析。

### 1.5.1 配伍的临床相关性评价

为评价真实世界中 2 味中药的相关性,建立高血压处方中镇肝熄风汤配伍药物的应用频数矩阵,计算中药间的欧式距离,将剂量配伍信息和中药效力评价信息输入 CNN 模型,以加权网络节点间的点式互信息 (pointwise mutual information, PMI) 值作为分类标准,据此评价它们在临床应用中配伍的耦合程度。

### 1.5.2 配伍的生物调控评价

重启随机游走 (random walk with restart, RWR) 算法被用于分析蛋白质相互作用网络 (protein-protein interaction network, PPIN),以评价中药配伍后对与高血压相关生物过程的调控程度。本研究在 Python (version 3.8.10) 中使用 pyrwr 包 (version 1.0.0),以 PPIN 中的高血压相关基因作为种子节点集,设置重启概率为 0.75<sup>[32]</sup>,进行 RWR 计算,得到稳定扩散概率,求和用于量化配伍对生物网络的干预。同时在 R 语言 (version 3.6.3) 中使用 GOSemSim 包 (version 2.12.1),以 BioConductor 为注释数据,进行 GO 语义相似性度量 (the semantic comparison of gene ontology annotations, GoSim)<sup>[33-34]</sup>,用于评价中药与高血压相关生物过程的紧密程度。

### 1.5.3 机器学习模型进行效力评价

为了综合真实世界数据和生物数据,进行镇肝熄风汤配伍组合对高血压调节效力评价。本研究在 R 语言 (version 3.6.3) 中使用 kkn (version 1.3.1), e1071 (version 1.7-8), gbm (version 2.1.8) 和 klaR (version 0.6-15) 包,以临床相关性评价信息和生物调控评价信息作为输入信息,实现 KNN、SVM、GBRT、Bayes 和 LG 模型训练。为防止模型过拟合,根据已有研究<sup>[35]</sup>,设置 10 次交叉验证用于配伍的有效性信息综合评价和筛选。最佳模型同样被用于分析前期研究获得的高血压肾损害数据,以验证模型效力。

## 1.6 通路富集分析

对机器模型筛选的有效配伍信息进行基于超几何分布的 KEGG 富集分析和 GO 富集分析,以 org.Hs.eg.db (version 3.10.0) 为基因注释。为了减少假阳性,将控制 False Discovery Rate (FDR) 获得的  $q$  值和经过 Benjamini-Hochberg (BH) 校正后的  $P$  值 (adjust  $P$ ) 均小于 0.05 作为筛选阈值<sup>[36]</sup>,使用 clusterProfiler 包<sup>[37]</sup> (version 3.14.3) 在 R 语言 (version 3.6.3) 中实现。使用层次聚类 (hierarchical clustering, HCT) 对中药和通路进行分类,用于区分其干预的生物过程。同时使用 Metascape 进行包括蛋白质组学、代谢组学的多基因荟萃分析。

## 2 结果

### 2.1 纳入患者特征

如表1所示,本研究共筛选1063名高血压患者

服用镇肝熄风汤,另有1823名未服用镇肝熄风汤的患者作为对照,2组在年龄、性别、高血压等级、心血管风险等级等方面没有统计学差异。

表1 纳入患者特征

特征	服用镇肝熄风汤	未服用镇肝熄风汤	P值
人数	1063	1823	—
年龄:均值±标准差/岁	57.58±13.65	55.16±15.41	0.332
性别:人数			
男性	461(43.37)	750(41.14)	
女性	602(56.63)	1073(58.86)	0.258
高血压等级:			
1级	178(16.74)	301(16.51)	
2级	312(29.35)	598(32.80)	0.142
3级	573(53.91)	942(51.69)	
低危	15(1.41)	24(1.31)	
心血管风险等级:			
中危	56(5.27)	69(3.78)	
高危	193(18.16)	309(16.95)	0.195
极高危	799(75.16)	1421(77.96)	

注:数据采用均数±标准差或数据(百分数)的形式展示,P值使用t检验或卡方检验计算获得。

### 2.2 镇肝熄风汤的常用配伍

通过Apriori算法筛选镇肝熄风汤的常用配伍中药,包括桂枝、黄柏、合欢皮、郁金、天麻、延胡索、丹参、酸枣仁、川芎、青黛、乳香、没药(图2)。BGLL算法将中药复杂关系网络分为6个社团(图3),模块度为0.423,表明各社团内部中药连接紧密,社团划分理想,社团1中的牛膝、玄参、白芍、牡蛎、茵陈、甘草,位于网络中央,与其余社团广泛配伍,印

证了中医方剂学理论下,镇肝熄风汤以牛膝为“君”,牡蛎、白芍为“臣”,发挥核心功效,补益肝肾、益阴潜阳,“佐”以玄参(社团1)、天冬(社团2)、茵陈(社团3)、龟甲(社团5)、麦芽(社团4)、川楝子(社团0),清泄肝热,梳理肝气,“使”以甘草(社团1)和胃安中,君臣佐使相互配伍,共奏镇肝熄风、滋阴潜阳之功。

### 2.3 镇肝熄风汤配伍的剂量特征

研究结果显示,通过建立中药配伍的剂量指数矩阵并应用CNN模型进行训练,得到了剂量特征模型,其准确度97.72%。进一步的梯度相关分析获得的剂量梯度表明,镇肝熄风汤组成的龙骨、牡蛎、龟甲、赭石、玄参、白芍与配伍的天麻、延胡索、丹参、酸枣仁在高血压的治疗中用量均有所增加,同时社团网络(图3)中,镇肝熄风汤组成的核心中药连接紧密,且数量远多于外围中药。CNN模型筛选出的前30位最佳剂量组合如表2所示,其中丹参-天冬、牛膝-甘草、川芎-玄参、酸枣仁-茵陈配伍剂量在模型中的表现尤为突出,表明临床诊疗更加倾向于对此进行调整,以优化疗效。

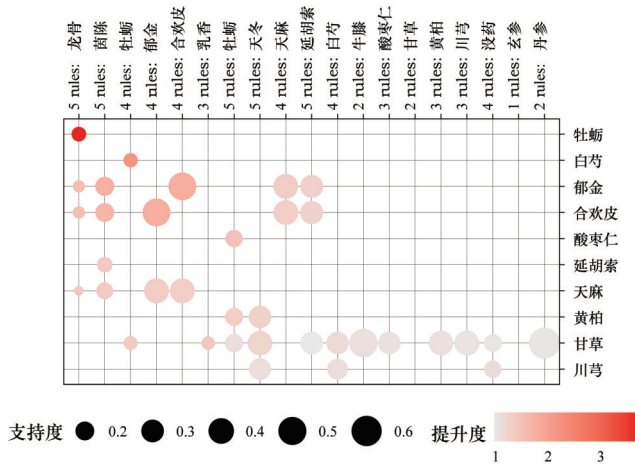


图2 镇肝熄风汤配伍中药间的关联规则

中药

- A 龙骨
- B 茵陈
- C 牡蛎
- D 白芍
- E 酸枣仁
- F 龟甲
- G 牛膝
- H 赭石
- I 天冬
- J 玄参
- K 麦芽
- L 川楝子
- M 桂枝
- N 黄柏
- O 合欢皮
- P 郁金
- Q 天麻
- R 延胡索
- S 丹参
- T 甘草
- U 川芎
- V 青黛
- W 乳香
- X 没药

社团

- 1 (38.9%)
- 3 (23.8%)
- 4 (12.59%)
- 5 (10.76%)
- 2 (10.07%)
- 0 (3.89%)

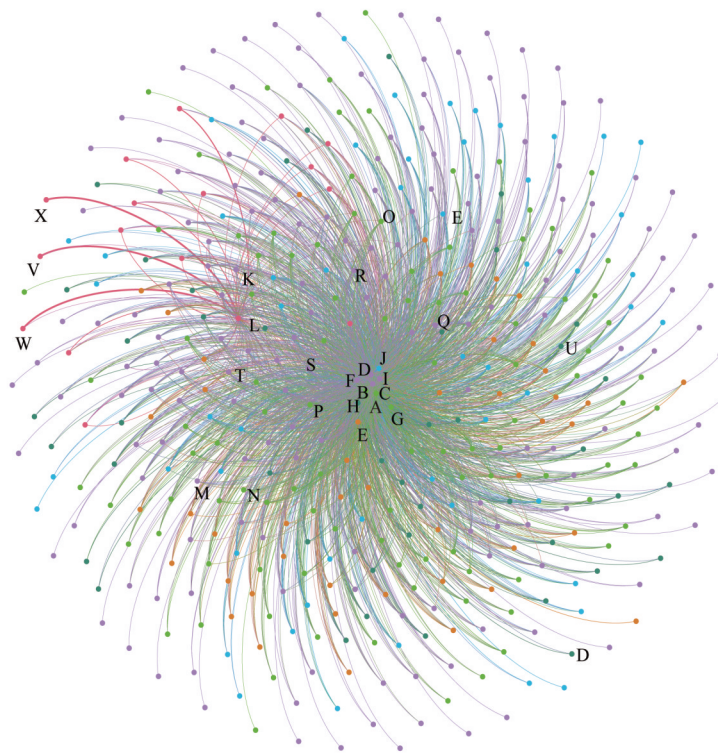


图3 镇肝熄风汤配伍中药的社团划分

表2 CNN模型筛选的前30位最佳剂量组合

中药 A	中药 B	平均剂量/g		剂量指数	中药 A	中药 B	平均剂量/g		剂量指数
		中药 A	中药 B				中药 A	中药 B	
天冬	丹参	24.45	20.45	10	麦芽	天麻	24.28	16.07	6
牛膝	甘草	14.02	7.40	10	牛膝	合欢皮	14.02	20.46	6
玄参	川芎	19.28	14.66	10	白芍	丹参	16.20	20.45	6
茵陈	酸枣仁	23.90	26.59	10	茵陈	黄柏	23.90	12.24	6
川楝子	延胡索	10.15	24.84	10	玄参	郁金	19.28	14.60	6
牛膝	桂枝	14.02	10.79	10	茵陈	川芎	23.90	14.66	6
白芍	天麻	16.20	16.07	10	龙骨	川芎	26.62	14.66	6
玄参	丹参	19.28	20.45	8	天冬	甘草	24.45	7.40	6
麦芽	酸枣仁	24.28	26.59	8	川楝子	酸枣仁	10.15	26.59	6
牛膝	延胡索	14.02	24.84	8	龟甲	酸枣仁	21.10	26.59	6
天冬	郁金	24.45	14.60	8	牡蛎	天麻	23.96	16.07	6
白芍	黄柏	16.20	12.24	8	牛膝	乳香	14.02	7.50	6
川楝子	合欢皮	10.15	20.46	8	赭石	延胡索	15.04	24.84	6
白芍	川芎	16.20	14.66	8	白芍	郁金	16.20	14.60	4
牡蛎	酸枣仁	23.96	26.59	8	茵陈	丹参	23.90	20.45	4

2.4 PPIN

研究结果显示,从GSE24752数据集中鉴定出LINC01420、MIER3、MSRB2、MYSM1、TMEM43、PIP4K2A等92个高血压差异表达基因,并与其他

数据库的信息合并后,共获得206个差异基因。此外收集高血压症状体征相关基因340个,中药靶点共4634个。通过这些靶点导入STRING数据库,构建了包含3188个节点和23333条边的蛋白质相

互作用网络。进一步对节点进行随机游走重启(RWR)计算,得到143组中药配伍的稳定扩散概率,发现牛膝-桂枝,川楝子-桂枝,牛膝-丹参,麦芽-黄柏组合的扩散概率最高,表明其可在高血压生物网络中发挥潜在影响。

## 2.5 配伍中药的有效性评价

### 2.5.1 模型性能

如图4所示KNN模型效果最佳,ROC曲线下面积(Area Under the Receiver Operating Characteristic Curve, AUC)为93.5%,作为对照,GBDT、SVM和LR的AUC值分别为87.6%、73.3%、66.3%,Bayes的效果最差,AUC值仅为51.3%,表明KNN模型能够有效处理异构数据,验证了其在临床数据与微观数据融合中的可靠性。

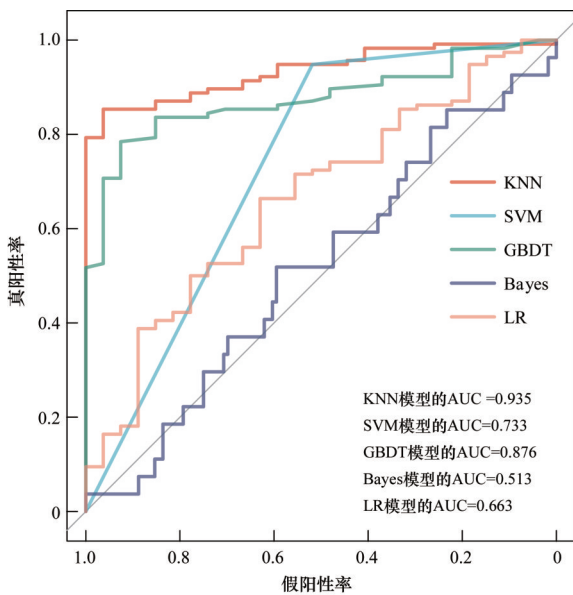
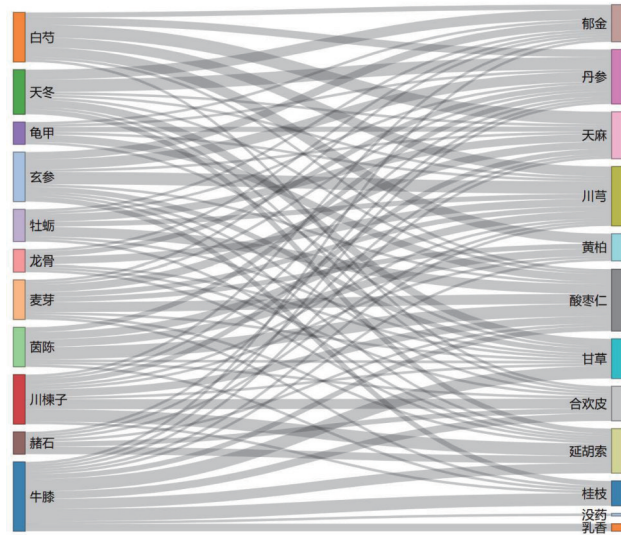


图4 KNN、SVM、GBDT、Bayes和LR模型的ROC曲线

### 2.5.2 有效配伍筛选

当参数 $k=23$ 时,KNN模型的效力最佳,根据其最佳截止点,得到了镇肝熄风汤组成与其他11味药的87个药对配伍关系,如图5所示,在KNN模型内综合中药的生物特征后,牛膝-酸枣仁、川芎-丹参为核心配伍药物。镇肝熄风汤的主要配伍方向包括滋阴通络、调补肝肾和清热安神。滋阴通络配伍包括玄参-川芎、玄参-丹参、白芍-丹参、白芍-天麻、天冬-丹参;疏理肝肾包括牛膝-乳香,龟甲-酸枣仁,赭石-延胡索,川楝子-延胡索;清热安神



注:连接线的宽度表明机器学习模型计算筛选的配伍紧密度。

图5 KNN模型筛选的中药配伍关系

包括茵陈-酸枣仁、茵陈-黄柏、牡蛎-酸枣仁。

同时,使用关联规则对配伍中药和患者症状的相关性进行量化,体现镇肝熄风汤配伍时的症状特征。如表3所示,牛膝、白芍、川芎、天麻、赭石主要针对眩晕、头痛、易怒、夜尿增多等高血压常见症

表3 中药与症状的关联规则

中药	症状	支持度	提升度
牛膝	眩晕	0.206	1.781
白芍	眩晕	0.206	1.781
川芎	眩晕	0.233	1.471
牛膝	夜尿增多	0.205	1.412
郁金	肢体疼痛	0.203	1.310
酸枣仁	失眠	0.242	1.309
川芎	头痛	0.232	1.289
天麻	头痛	0.236	1.286
延胡索	肢体疼痛	0.205	1.284
龙骨	失眠	0.287	1.252
龟甲	夜尿增多	0.287	1.252
茵陈	易怒	0.374	1.243
黄柏	自汗	0.374	1.243
赭石	易怒	0.342	1.230
川楝子	易怒	0.342	1.230
合欢皮	失眠	0.250	1.228
白芍	头痛	0.226	1.199
桂枝	水肿	0.219	1.199
牡蛎	失眠	0.256	1.196
丹参	心悸	0.203	1.175

状。心悸、水肿、失眠、自汗作为常见合并症状可分别配伍应用丹参、桂枝、酸枣仁、黄柏等药物。

### 2.6 配伍药对的疗效机制

如图6所示,KEGG富集共保留73条富集基因数大于上四分位数的核心通路,建立中药与通路的关系矩阵,HCT将中药分为5组(表4),其中K1、K2、K3组在炎症信号通路有明显的生物功能聚集,包括:NF-κB信号通路、MAPK信号通路、TNF信号通路、HIF-1信号通路、Ras信号通路等,与炎症细

胞因子诱发高血压相关病理变化密切相关;相关镇肝熄风汤组成中药包括K1组中的天冬、玄参、川楝子、麦芽,K2组的白芍、川芎,K3组的牛膝、甘草;核心配伍中药包括K1区的延胡索、乳香、郁金,K3区的丹参、桂枝、黄柏。表明镇肝熄风汤配伍间或直接或间接调控TNF-α、IL-1β、IL-6等炎症因子活性来抑制NF-κB信号通路的过度激活,进一步抑制这些炎症细胞因子造成的炎症反应。此外,针对炎症反应的下游通路,K1组与K3组镇肝熄风汤配伍

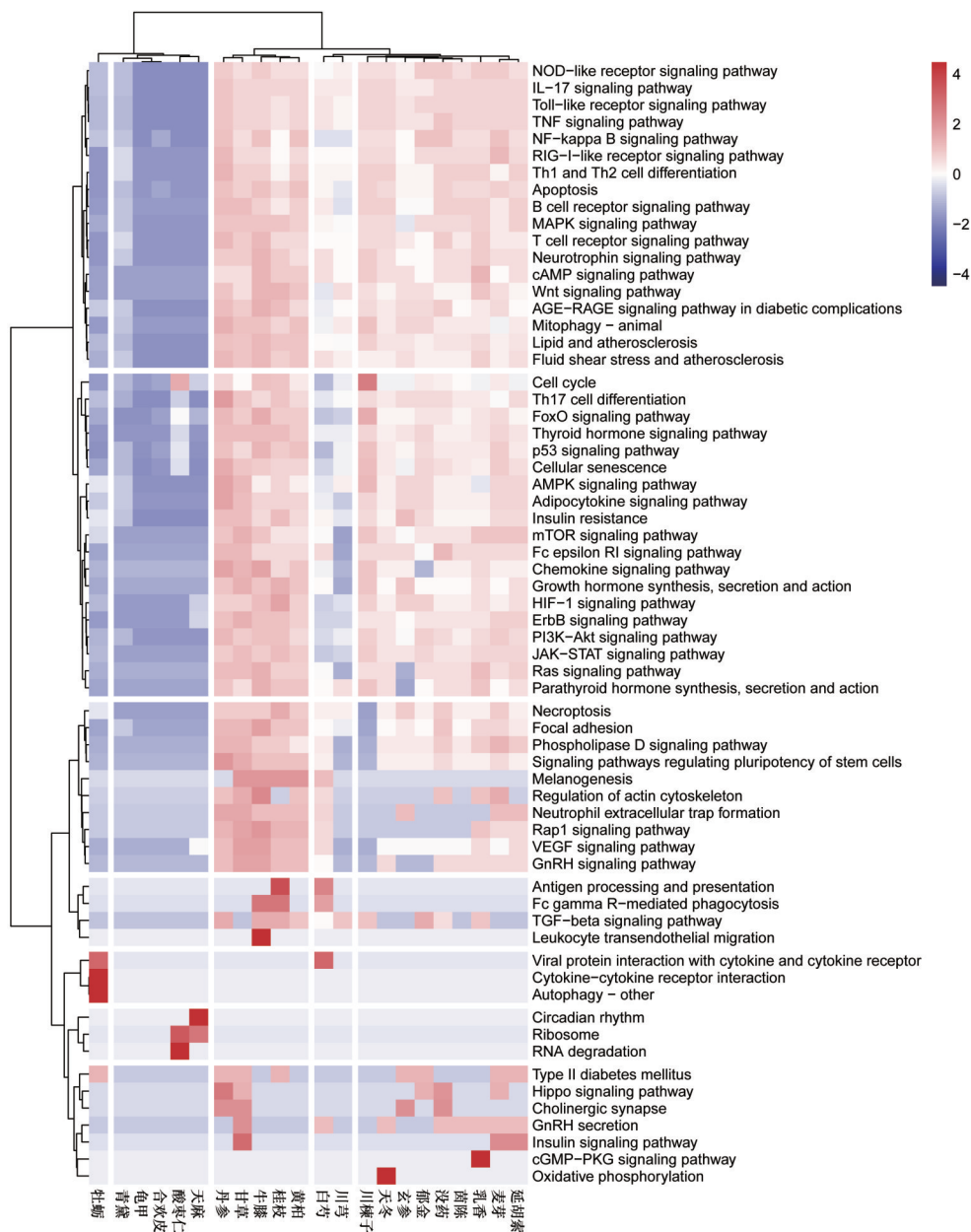


图6 中药的KEGG分析

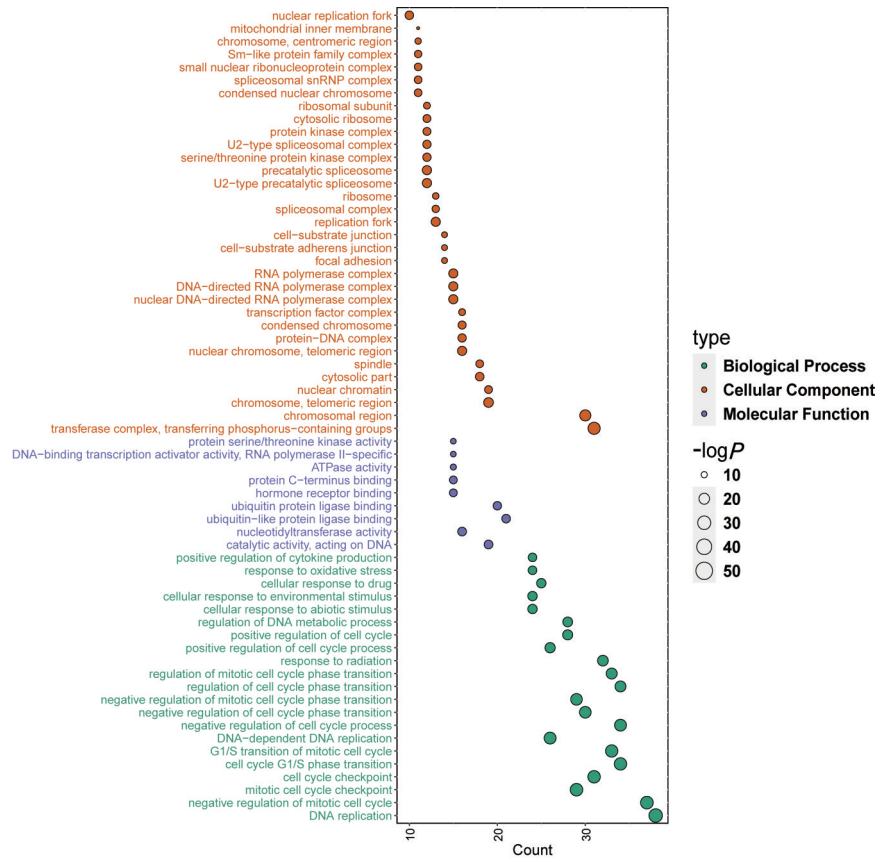
表4 中药在BGLL社团与KEGG通路中的分组

中药	BGLL社团	KEGG通路	中药	BGLL社团	KEGG通路
桂枝	1	K3	甘草	1	K3
黄柏	1	K3	川芎	1	K2
茵陈	3	K1	龟甲	5	K4
牡蛎	1	K5	牛膝	1	K3
白芍	1	K2	天冬	2	K1
合欢皮	3	K4	玄参	1	K1
郁金	3	K1	麦芽	4	K1
天麻	3	K4	川楝子	0	K1
延胡索	3	K1	乳香	0	K1
丹参	1	K3	没药	0	K1
酸枣仁	2	K4	青黛	0	K4

相比有更多中药在胰岛素、胆汁酸和脂质代谢的相关通路有更多富集。K4组天麻、酸枣仁在睡眠节律相关通路中具有富集。此外,在KNN模型筛选获得的87个配伍药对中,在HCT分组与BGLL社团相比没有差异( $P=0.73$ ),进一步表明KNN模型筛

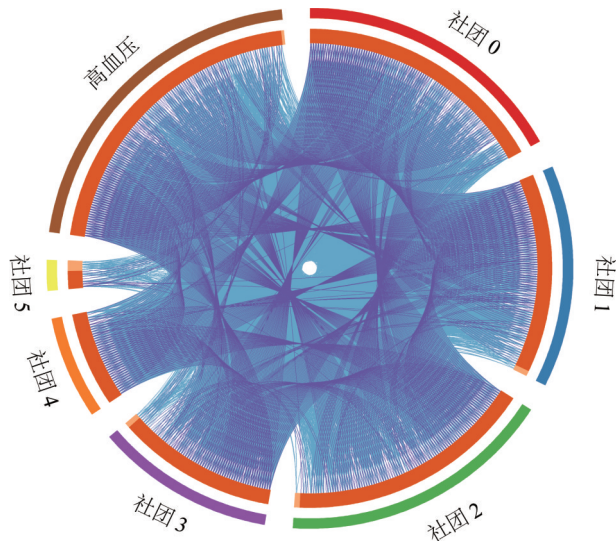
选兼顾了配伍的临床特征和生物特征。

如图7(a)所示,GO富集发现,镇肝熄风汤配伍可参与细胞因子、细胞代谢、生长因子的调节。同时,如图7(b)和图7(c)所示,配伍中药在不同中药社团内与高血压相关基因在GO网络的主要节点

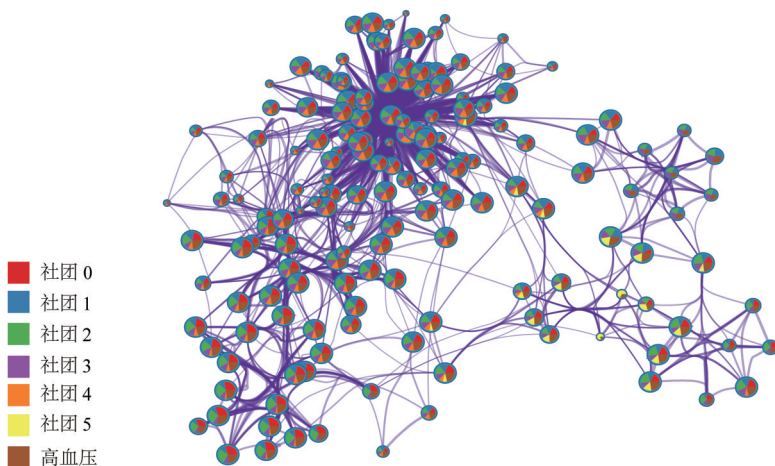


(a) 配伍中药的GO富集结果

图7 配伍中药的GO富集分析



(b) 各个中药社团与高血压之间的GO联系



(c) 各个中药社团与高血压的GO富集网

注:(b) 内圈代表基因信息,浅蓝色曲线表示连接的集合具有相同基因,深蓝色曲线表示连接的集合具有同一GO条目富集,内圈的红色部分表示该集合与其他集合具有相同基因的比例;(c) 镇肝熄风汤与高血压模块富集的GO网络,节点以饼图表示,其组成颜色及比例代表其对应模块和调控基因量。

图7 配伍中药的GO富集分析(续)

上均有公共富集,同样表明本研究总结的镇肝熄风汤配伍规律具有生物学基础。

### 2.7 外部数据测试

为了验证本研究方法构建模型的效力,使用山东中医药大学附属医院标准化电子病历中的高血压肾病数据作为外部数据,开展KNN模型验证,最终AUC=91.3%,并筛选获得丹参-当归、党参-白术、当归-半夏等有效配伍药对的宏微观信息,符

合高血压肾损害的临床治疗经验和相关文献<sup>[38-41]</sup>,进一步验证了此研究的可靠性和外推性。

## 3 讨论

综合各代医家对高血压病的认识,结合现代流行病学研究发现,阴虚阳亢、瘀血阻络、肝肾亏损等证占比逐渐增加<sup>[42-43]</sup>。人体脏腑经络依赖气血津

液濡养,高血压作为慢性疾病,病程日久,属于“终生控制性疾病”,往往由实而虚,由肝及肾;疾病日久,年老体衰,情志不遂,则阴液亏损、气血不畅、虚阳上浮,引动气血上充脑窍,出现眩晕、头痛。镇肝熄风汤是治疗高血压的经典中药处方,已在《高血压中医诊疗专家共识》<sup>[3]</sup>、《2级高血压中西医结合诊疗指南》<sup>[4]</sup>等共识指南中得到推荐。然而,其配伍总结多局限在中医理论层面,缺乏多模态、多维度的临床数据挖掘研究<sup>[45-46]</sup>。

中医处方配伍规律的挖掘是中医现代化研究的重要方向。随着大数据和机器学习技术的发展,越来越多的研究将疾病特征与配伍规律进行匹配<sup>[47-49]</sup>。本研究利用真实世界的电子病历数据,从临床特征和微观机制特征2个方面,分析了镇肝熄风汤在高血压治疗中的针对性及其广泛的临床应用。创新性地采用KNN模型,将处方数据内中药的应用频率、剂量和患者临床症状等信息,与中药的生物特征进行综合分析,优于传统研究中仅考虑处方药物而忽视剂量的研究模式,能够更全面地捕捉配伍的临床影响因素,从而为理解中药治疗高血压提供了更深入的视角。最终结果显示镇肝熄风汤的核心药物为牛膝、玄参、白芍、牡蛎、茵陈、甘草,以镇肝熄风、滋阴潜阳为主要功效;在临床应用时,通过增加龙骨、牡蛎、龟甲、赭石、玄参和白芍的剂量,进一步强化平肝潜阳,清火熄风的效果,从而缓解头痛、眩晕、急躁等症状。另外,针对高血压合并的心悸可配伍丹参活血化瘀,水肿可配伍桂枝温阳利水,失眠可合用酸枣仁、合欢皮养心安神。体现了中医辨证论治的原则,符合因人施治的临床配伍规律。

在微观机制方面,中医药治疗疾病的机制原理仍然不明,且缺乏科学基础,使得传统的化合物分析难以用于中药处方筛选<sup>[47]</sup>。因此,有必要开发模型框架,将中医临床数据与现代生物医学进行关联,本研究关注生物效应网络的拓扑特征,并将其作为调节参数,以挖掘适应症优化的处方配伍规律。此方法弥补了补充传统中药网络药理学研究局限于单一处方且缺乏临床数据关联的不足<sup>[50-53]</sup>。在分析配伍药物生物机制的同时,为挖掘有效的中

药配伍提供了额外数据模块,有助于捕捉潜在的症状-中药关联,确保框架的外推性和外部真实性。

然而,本研究尚未深入探讨中药化合物与疾病的调节关系,未来将优先考虑中药的有效化合物筛选。此外,高血压等慢性疾病病程较长,容易导致心脏、脑、肾脏等靶器官损害,不同进阶段的中药筛选仍然具有挑战性,需要依靠临床大样本的多组学数据,进一步优化模型,为高血压病的全流程治疗提供新的诊疗方案。

## 4 结论

本研究利用机器学习算法进行建模,获得87个包含最佳剂量的镇肝熄风汤药对组合,验证了镇肝熄风汤配伍可以针对高血压发挥以控制炎症反映和调节能量代谢为核心多层次生物调控作用,表明KNN模型可有效融合CNN数据和网络拓扑数据,对传统的关联规则模型进行补充,且具有良好的普遍适用性,为多模态的中药数据集挖掘提供了一种跨学科研究方法,同时为临床决策提供多层次参考。

## 参考文献(References)

- [1] Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension[J]. *European Heart Journal*, 2018, 39(33): 3021-3104.
- [2] Mach F, Baigent C, Catapano A L, et al. 2019 ESC/EAS guidelines for the management of dyslipidaemias: Lipid modification to reduce cardiovascular risk[J]. *European Heart Journal*, 2020, 41(1): 111-188.
- [3] 中华中医药学会心血管病分会. 高血压中医诊疗专家共识[J]. *中国实验方剂学杂志*, 2019, 25(15): 217-221.
- [4] 中华医学会老年医学分会, 中国医疗保健国际交流促进会高血压病分会. 老年高血压特点及临床诊治流程专家共识(2024)[J]. *中华老年医学杂志*, 2024, 43(3): 257-268.
- [5] 国家卫生健康委员会疾病预防控制局, 国家心血管病中心, 中国医学科学院阜外医院, 等. 中国高血压健康管理规范(2019)[J]. *中华心血管病杂志*, 2020, 48(1): 10-46.
- [6] 中华中医药学会. 松龄血脉康胶囊治疗原发性高血压临

- 床应用专家共识[J]. 中华高血压杂志, 2024, 32(1): 8-15.
- [7] 谢依璇, 闫世艳, 陈波, 等. 新时代针灸的发展[J]. 科技导报, 2023, 41(14): 42-50.
- [8] Xiong X J, Yang X C, Feng B, et al. Zhen Gan Xi Feng Decoction, a traditional Chinese herbal formula, for the treatment of essential hypertension: A systematic review of randomized controlled trials[J]. *Evidence-Based Complementary and Alternative Medicine*, 2013, 2013: 982380.
- [9] Huang Y L, Chen Y H, Cai H R, et al. Herbal medicine (Zhengan Xifeng Decoction) for essential hypertension protocol for a systematic review and meta-analysis[J]. *Medicine*, 2019, 98(6): e14292.
- [10] Yu X Y, Zhang X D, Jin H, et al. Zhengganxifeng decoction affects gut microbiota and reduces blood pressure via renin-angiotensin system[J]. *Biological & Pharmaceutical Bulletin*, 2019, 42(9): 1482-1490.
- [11] Whelton P K, Carey R M, Aronow W S, et al. Correction to: 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the American college of cardiology/American heart association task force on clinical practice guidelines[J]. *Hypertension*, 2018, 71(6): e140-e144.
- [12] Hart P D, Bakris G L. Hypertensive nephropathy: Prevention and treatment recommendations[J]. *Expert Opinion on Pharmacotherapy*, 2010, 11(16): 2675-2686.
- [13] Weir M R. Hypertension and kidney disease[J]. *Journal of the American Society of Hypertension*, 2014, 8(11): 855-857.
- [14] Ren W Q, Liao J Q, Chen J L, et al. The effect of Chinese herbal medicine combined with western medicine on vascular endothelial function for patients with hypertension: Protocol for a systematic review and meta-analysis[J]. *Medicine*, 2019, 98(49): e18134.
- [15] 刘保延, 文天才. 从信息科学与物质科学及其范式认识中西医学体系的科学原理[J]. 科技导报, 2023, 41(14): 5-13.
- [16] 张迪, 雒琳, 文天才, 等. 中医辨证论治疗效评价研究进展[J]. 科技导报, 2023, 41(14): 32-41.
- [17] Shu Z X, Zhou Y N, Chang K, et al. Clinical features and the traditional Chinese medicine therapeutic characteristics of 293 COVID-19 inpatient cases[J]. *Frontiers of Medicine*, 2020, 14(6): 760-775.
- [18] 周雪忠, 王世华, 张迪, 等. 构建中医药特色真实世界临床研究新模式的实践与思考[J]. 科技导报, 2023, 41(14): 22-31.
- [19] 郭然, 曾典, 胡裕涵, 等. 价值医疗视角下的中医药优势评价思路与方法[J]. 科技导报, 2023, 41(14): 14-21.
- [20] Li S. Network pharmacology evaluation method guidance-draft[J]. *World Journal of Traditional Chinese Medicine*, 2021, 7(1): 146-154.
- [21] Li S, Zhang B. Traditional Chinese medicine network pharmacology: Theory, methodology and application[J]. *Chinese Journal of Natural Medicines*, 2013, 11(2): 110-120.
- [22] Dong X, Zheng Y, Shu Z X, et al. TCMPR: TCM prescription recommendation based on subnetwork term mapping and deep learning[J]. *BioMed Research International*, 2022, 2022: 4845726.
- [23] Shu Z X, Wang J J, Sun H L, et al. Diversity and molecular network patterns of symptom phenotypes[J]. *NPJ Systems Biology and Applications*, 2021, 7(1): 41.
- [24] 郭义, 王江, 刘保延, 等. 计算针灸学前沿进展[J]. 科技导报, 2023, 41(14): 51-55.
- [25] Wang N, Li P, Hu X C, et al. Herb target prediction based on representation learning of symptom related heterogeneous network[J]. *Computational and Structural Biotechnology Journal*, 2019, 17: 282-290.
- [26] Huan J M, Su W G, Li W, et al. Summarizing the effective herbs for the treatment of hypertensive nephropathy by complex network and machine learning[J]. *Evidence-Based Complementary and Alternative Medicine*, 2021, 2021: 5590743.
- [27] Huan J M, Li Y L, Zhang X, et al. Predicting coupled herbs for the treatment of hypertension complicated with coronary heart disease in real-world data based on a complex network and machine learning[J]. *Evidence-Based Complementary and Alternative Medicine*, 2022, 2022: 8285111.
- [28] Ru J L, Li P, Wang J N, et al. TCMSP: A database of systems pharmacology for drug discovery from herbal medicines[J]. *Journal of Cheminformatics*, 2014, 6: 13.
- [29] Wu Y, Zhang F L, Yang K, et al. SymMap: An integrative database of traditional Chinese medicine enhanced by symptom mapping[J]. *Nucleic Acids Research*, 2019, 47(D1): D1110-D1117.
- [30] Xu H Y, Zhang Y Q, Liu Z M, et al. ETCM: An encyclopaedia of traditional Chinese medicine[J]. *Nucleic Acids Research*, 2019, 47(D1): D976-D982.
- [31] Szklarczyk D, Gable A L, Lyon D, et al. STRING v11:

- Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets[J]. *Nucleic Acids Research*, 2019, 47(D1): D607-D613.
- [32] Yang J, Tian S S, Zhao J, et al. Exploring the mechanism of TCM formulae in the treatment of different types of coronary heart disease by network pharmacology and machining learning[J]. *Pharmacological Research*, 2020, 159: 105034.
- [33] Yu G C. Gene ontology semantic similarity analysis using GOSemSim[M]//*Methods in Molecular Biology*. New York: Springer, 2020: 207-215.
- [34] Wang J Z, Du Z D, Payattakool R, et al. A new method to measure the semantic similarity of GO terms[J]. *Bioinformatics*, 2007, 23(10): 1274-1281.
- [35] Yu G C, Li F, Qin Y D, et al. GOSemSim: An R package for measuring semantic similarity among GO terms and gene products[J]. *Bioinformatics*, 2010, 26(7): 976-978.
- [36] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing[J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 1995, 57(1): 289-300.
- [37] Yu G C, Wang L G, Han Y Y, et al. clusterProfiler: An R package for comparing biological themes among gene clusters[J]. *OMICS: A Journal of Integrative Biology*, 2012, 16(5): 284-287.
- [38] Yan D, Yue B W, Qian M Y, et al. JYYS granule mitigates renal injury in clinic and in spontaneously hypertensive rats by inhibiting NF- $\kappa$ B signaling-mediated microinflammation[J]. *Evidence-Based Complementary and Alternative Medicine*, 2018, 2018: 8472963.
- [39] Owoicho Orgah J, Wang M, Yang X H, et al. Danhong injection protects against hypertension-induced renal injury via down-regulation of myoglobin expression in spontaneously hypertensive rats[J]. *Kidney & Blood Pressure Research*, 2018, 43(1): 12-24.
- [40] Wu L H, Liu M, Fang Z Y. Combined therapy of hypertensive nephropathy with breviscapine injection and anti-hypertensive drugs: A systematic review and a meta-analysis[J]. *Evidence-Based Complementary and Alternative Medicine*, 2018, 2018: 2958717.
- [41] Li Y Z, Yan S H, Qian L C, et al. Danhong injection for the treatment of hypertensive nephropathy: A systematic review and meta-analysis[J]. *Frontiers in Pharmacology*, 2020, 11: 909.
- [42] 许晴, 宋敏, 姜楠, 等. 原发性高血压中医证型分布规律及其危险因素研究[J]. *安徽中医药大学学报*, 2023, 42(4): 15-19.
- [43] 尚倩倩, 王蕾, 王忆勤, 等. 原发性高血压中医病因病机及证候的研究进展[J]. *河北中医*, 2017, 39(1): 138-142.
- [44] 中华中医药学会. 2级高血压中西医结合诊疗指南[EB/OL]. (2023-01-06)[2024-06-19]. <https://www.cacm.org.cn/2023/01/16/21450>.
- [45] 张红梅, 肖勇, 吕武清. 试论茵陈、生麦芽、川楝子在镇肝熄风汤中的配伍意义[J]. *中国中医药现代远程教育*, 2014, 12(19): 104-105.
- [46] 辛松根. 镇肝熄风汤与建瓴汤的组方意义和运用[J]. *中医研究*, 2014, 27(1): 41-43.
- [47] Gan X, Shu Z X, Wang X Y, et al. Network medicine framework reveals generic herb-symptom effectiveness of traditional Chinese medicine[J]. *Science Advances*, 2023, 9(43): eadh0215.
- [48] Lu K Z, Yang K, Sun H L, et al. SympGAN: A systematic knowledge integration system for symptom-gene associations network[J]. *Knowledge-Based Systems*, 2023, 276: 110752.
- [49] Yang K, Yu Z C, Su X, et al. A optimization framework for herbal prescription planning based on deep reinforcement learning[EB/OL]. [2023-12-04]. <https://arxiv.org/abs/2304.12828v1>.
- [50] Shang L R, Wang Y C, Li J X, et al. Mechanism of Si-junzi Decoction in the treatment of colorectal cancer based on network pharmacology and experimental validation[J]. *Journal of Ethnopharmacology*, 2023, 302: 115876.
- [51] He Q M, Liu C, Wang X H, et al. Exploring the mechanism of curcumin in the treatment of colon cancer based on network pharmacology and molecular docking[J]. *Frontiers in Pharmacology*, 2023, 14: 1102581.
- [52] Duan Z L, Wang Y J, Lu Z H, et al. Wumei Wan attenuates angiogenesis and inflammation by modulating RAGE signaling pathway in IBD: Network pharmacology analysis and experimental evidence[J]. *Phytomedicine*, 2023, 111: 154658.
- [53] Ma Y, Deng Y Y, Li N, et al. Network pharmacology analysis combined with experimental validation to explore the therapeutic mechanism of Schisandra Chinensis Mixture on diabetic nephropathy[J]. *Journal of Ethnopharmacology*, 2023, 302: 115768.

## The study on the formula characteristics of Zhengan Xifeng decoction in the treatment of hypertension based on machine learning models

HUAN Jiaming<sup>1</sup>, CHEN Xiaoqing<sup>2</sup>, YANG Wenqing<sup>3</sup>, LI Jie<sup>3</sup>, HUA Zhen<sup>4</sup>, WANG Yifei<sup>4</sup>, LI Yunlun<sup>1,4,5\*</sup>

1. First School of Clinical Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250355, China
2. Department of Cardiovascular, Hospital of East Gaoxin District of Jinan, Jinan 250101, China
3. Innovation Research Institute of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250355, China
4. Department of Cardiovascular, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan 250014, China
5. Precision Diagnosis and Treatment of Cardiovascular Diseases with Traditional Chinese Medicine Shandong Engineering Research Center, Jinan 250355, China

**Abstract** The high prevalence of hypertension and its wide range of complications have made it a significant risk factor for cardiovascular disease mortality. Zhengan xifeng decoction (ZXD) is a classic traditional Chinese medicine prescription used in the clinical treatment of hypertension. However, there is a lack of systematic analysis of its composition rules and biological effects. In order to further explore the clinical application characteristics of ZXD in the treatment of hypertension, this study established a dataset based on the electronic medical records of hypertension patients from the Affiliated Hospital of Shandong University of Traditional Chinese Medicine from 2011 to 2019. Apriori algorithm was used to quantify the compatibility strength of the Chinese medicines in ZXD, while convolutional neural networks were employed to quantify the dosage characteristics. Topological feature analysis of protein interactions was used to examine biological characteristics. The outcome was then input into k-nearest neighbors, support vector machines, gradient boosting decision trees, Bayesian networks, and logistic regression models to evaluate the efficacy of each model, reflecting the composition rules and mechanisms of action of ZXD from multiple dimensions. The results showed that the k-nearest neighbors algorithm performed the best among the five models (AUC=93.5%), identifying 87 effective combinations, validating the mechanisms by which ZXD regulates cytokines, reduces inflammation, and corrects metabolic disorders. External testing indicated that this research could be extrapolated to other diseases. This study demonstrates that the model effectively integrates convolutional neural network data and network topology data, incorporating dosage information and biological characteristics into the exploration of Chinese medicine combinations. It complements traditional association rule models and has good general applicability and extrapolation capability, providing an interdisciplinary research approach for mining multimodal Chinese medicine datasets.

**Keywords** machine learning; real world data; data mining; zhengan xifeng decoction; hypertension ●



(责任编辑 徐丽娇)