

基于人工智能算法的中药主归经络判别研究

赵书言, 王书睿, 申镇华, 肖 琿, 翟玉莹, 姜希伟*

沈阳药科大学医疗器械学院, 辽宁 沈阳 110016

摘要: **目的** 基于活性成分属性, 探究药材主归经络判别模型的构建及优化。**方法** 收集药材归经属性及活性成分, 以单经络药材所含活性成分为基础, 构建活性成分结构特征、理化性质分类集合以及活性成分频数分类集合。应用余弦相似度算法, 量化不同经络药材间活性成分的属性差异, 计算药材所含活性成分组成的集合与不同活性成分类别间的相似性, 构建集合相似度矩阵作为建模数据集。应用人工智能算法建立药材主归经络判别模型, 并对模型进行优化及评价。**结果** 收集 209 味药材并建立了药材-归经属性-活性成分数据集, 通过对单经络药材活性成分分类, 获取了 34 个活性成分分类集合。计算每味药所含成分与不同活性成分类别之间的余弦相似度, 得到维度为 209×34 的集合相似度矩阵。使用 K 近邻算法、随机森林、极端梯度提升算法构建药材主归经络判别模型, 并进行超参数调优, 根据最优参数下测试集上的指标, 评价模型的优劣。随机森林算法表现出了最优的性能, 平衡准确率、曲线下面积 (area under curve, AUC) 分别为 0.86、0.90。**结论** 构建的药材主归经络判别模型达到了较好的结果并且运行稳定, 探索了药材归经属性的理论基础, 为揭示药材归经属性与药材活性成分间关系提供了新的研究方向, 为药材归经判别研究提供一种新的思路。

关键词: 中药; 归经判别; 归经理论; 机器学习; 中药药性

中图分类号: TP18; R285 文献标志码: A 文章编号: 0253-2670(2024)16-5573-10

DOI: 10.7501/j.issn.0253-2670.2024.16.019

Discriminant research on main meridians of traditional Chinese medicine based on artificial intelligence algorithms

ZHAO Shuyan, WANG Shurui, SHEN Zhenhua, XIAO Shen, ZHAI Yuxuan, JIANG Xiwei

School of Medical Equipment, Shenyang Pharmaceutical University, Shenyang 110016, China

Abstract: **Objective** To explore the construction and optimization of a discrimination model for the main meridians of medicinal materials based on the properties of active components. **Methods** Collecting the meridian and active component information of medicinal components. Using the active components of single-meridian medicinal materials as a foundation, the structural features of the active ingredients and the classification set of physicochemical properties and the frequency classification set of the active components were constructed. By applying the cosine similarity algorithm, quantifying the differences in properties of active components among various meridians, calculating the similarity between the set of active components contained in medicinal materials and the different active component categories, and constructing a set similarity matrix as the modeling dataset. Artificial intelligence algorithms were then utilized to establish discrimination model for the main meridians of medicinal materials and to optimize and evaluate the model. **Results** Collecting 209 medicinal materials and establishing a dataset comprising medicinal materials, meridians, and active components. By categorizing the corresponding components of single-meridian medicinal materials, we obtained 34 classification sets. Calculating the cosine similarity between the components contained in each herb and different active component categories yielded a similarity matrix with dimensions of 209×34 . K-nearest neighbor algorithm, random forest and extreme gradient lifting algorithm were used to construct the discrimination model for the main meridians of medicinal materials, hyperparameter tuning was carried out and the model's performance based on the test set under the optimal parameters was evaluated. The performance of the random forest algorithm was found to be optimal, with balanced accuracy and area under curve (AUC) of 0.86, 0.90, respectively. **Conclusion** This study successfully constructed a model for classifying the main meridians of medicinal materials, achieving favorable and stable results. It explored the theoretical foundation of meridian entry theory, providing a new direction for understanding

收稿日期: 2024-03-10

基金项目: 辽宁省教育厅基本科研项目 (LJKFZ20220258); 辽宁省教育厅基本科研项目 (JYTQN2023337)

作者简介: 赵书言, 硕士研究生, 研究方向为网络药理学。E-mail: shuyanzhao0106@163.com

*通信作者: 姜希伟, 副教授, 硕士生导师, 从事生物医药信息学研究。E-mail: jiangxiwei@syphu.edu.cn

the relationship between meridian and the active components of medicinal materials, and providing a new approach for the discrimination of meridians for medicinal materials.

Key words: traditional Chinese medicine; meridian discrimination; meridian entry theory; machine learning; properties of traditional Chinese medicines

中医学是中华民族传统的医学理论体系，是中国人千百年来智慧的结晶。在中医理论体系中，中药是用以防病治病的药物总称，在中医临床应用方面发挥着重要的作用^[1]。中药药性是中药的功能以及作用的精炼概括，包括四气五味、升降沉浮、归经、毒性等很多方面。药性从多角度地描述着中药的性质，是人们为指导临床用药总结提炼出来的^[2]。在中医临床研究中，为指导精准用药，人们将药物在人体的选择性作用概括为中药的归经^[3]。临床上，归经理论将中药与人体的官窍、脏腑密切联系起来，说明了中药对于人体特定部位的选择性作用^[4]。归经理论使得中医临床遣方用药更加精准，在中医药临床应用领域发挥重要作用，归经理论为临床用药提供了原则，也是中药方剂学形成的理论基础^[5]。

中药具有作用途径复杂、多成分协同调控的特征，这使得传统方法难以诠释中医药的原理。随着计算机技术的发展，对传统中医药的研究方法也日益丰富，为揭示中药的原理提供了新途径。在中药质量控制研究中，人工智能技术被用于利用多源数据进行综合分析、建立预测模型以识别中药的质量特征以及探索质量控制与功效之间的关联性^[6-8]。在中医药药理以及临床配伍的研究中，人工智能技术被用于挖掘药材-疾病、药材-药性之间的网络关系、分析中药化合物之间的复杂相互作用以及探究药物与药物之间的相互作用^[9-10]。随着中医药在国际的影响日渐加深，中药研究迫切的需要引入现代化方法，以促进中医药的国际传播。

在中药药性理论的现代化研究中，中药归经的量化一直是研究的重点内容。研究人员通过机器学习、图卷积网络等多学科交叉的方法，展开了对中药归经理论的研究。Wang 等^[11]确定了 646 种中药及其活性成分的分子特征，构建中药归经的判别模型。Yeh 等^[12]使用图卷积神经网络探究了中药活性成分与所属经络之间的关系。王小雪等^[13]通过文献定量可视化分析，发现中药材化学成分、药理作用、临床应用具有明确的关联性，且与归经属性一致。Zhang 等^[14]通过抗癌中药的活性成分拓扑结构信

息，采用图卷积网络和递归神经网络组成的深度学习混合模型，对药材的多经络信息进行预测。总体来讲，当前的研究主要集中在药材物质基础与其经络归属的联系方面，常忽略药材所属经络的主次问题，很少探究药材经络归属的原理。

药材的归经属性是指药物对机体某部分特殊的选择性作用，概括说明了药效所在的部位，药材归属于单经络或多经络的不同，主要基于它们的药理作用范围和治疗的特异性差异，同时，已有研究表明药材通过活性成分作用于人体，最终综合体现在经络的选择上。在这一基础上，本研究提出药材的归经属性与药材所含活性成分的属性类别相关，并基于药材-归经属性-活性成分数据，构建了药材主归经络判别模型并进行优化、验证。为探索药材经络归属原理、揭示药材归经属性与药材活性成分间关系、预测新药材归经属性提供新思路。

1 数据集获取与方法

1.1 数据集获取

1.1.1 药材纳入标准 相比于归属单一经络的药材，药材的归经属性越多，其对应的活性成分也会越复杂，对机体的作用也就越发广泛。本研究以归单经络药材所含有的活性成分作为基础，将只归单一经络的药材对应的活性成分简称为该经络的关联成分，并对经络关联成分划分类别，通过集合的方式探寻成分特征在不同经络的差异、探究经络属性与药材之间的复杂关系。同时，为了最大化药材在经络归属方面的差异，本研究对主归肺经、肝经的药材分别进行集合的划分。

收集《中国药典》2020 年版中所记录的所有药材及其归经属性。为了探究归肺经、肝经药材的差异，验证所提出经络归属与药材成分属性相关的猜想，本研究筛选出首位归经信息为肺经、肝经且不同时归肺经、肝经的药材，根据归经数量的不同，将药材分为单经、双经、多经药材。同时，使用中药系统药理学数据库与分析平台（TCMSP）（<https://tcmsp.com/>）收集了药材对应的活性成分信息^[15]，由于矿物药、动物药的活性成分信息难以获取，并且与植物药的活性成分差异较大，因此本研

究主要针对植物药展开研究。为保证数据完整性，未对活性成分进行筛选。

1.1.2 特征变量选取 药物化学研究中认为，相似的分子结构通常具有相似的生物活性^[16]。本研究使用化合物的结构特征以及理化性质间接表示化合物的生物活性。通过化合物的结构特征与理化性质，对经络关联成分进行分类，量化归属不同经络的药材之间的差异。使用扩展连通性指纹（extended connectivity fingerprints, ECFP）表示经络关联成分的结构特征^[17]；使用密度、沸点、饱和蒸气压等表示经络关联成分的理化特征；使用经络关联成分在肺经、肝经单归经药材中出现的频数表示活性成分对于归经属性判别的重要性。

1.2 数据集清洗

关联成分结构特征数据维度高、特征稀疏，为

了实现对关联成分的结构特征进行有效分类，降低成分结构特征数据的维度，对本研究数据集进行了清洗。数据清洗主要处理流程如图 1 所示。数据清洗主要应用化合物相似性计算、特征去重以及特征相关性筛选等方法。化合物相似性应用谷本系数^[18]计算，计算方法如公式（1）所示。

$$Sim_{A,B} = \frac{[\sum_{j=1}^n x_{jA}x_{jB}]}{[\sum_{j=1}^n (x_{jA})^2 + \sum_{j=1}^n (x_{jB})^2 - \sum_{j=1}^n x_{jA}x_{jB}]} \quad (1)$$

$Sim_{A,B}$ 表示分子间相似度， x_{jA} 、 x_{jB} 表示分子 A、B 的第 j 个特征向量， n 表示 ECFP 指纹的特征维度

理化性质特征数据的数据量纲可能相差多个数量级，为了消除不同特征之间的量纲影响，提高模型的准确性和收敛速度，更好地获取关联成分分类集合，对肺经、肝经关联成分理化性质特征数据进行了归一化处理，处理流程如图 2 所示。

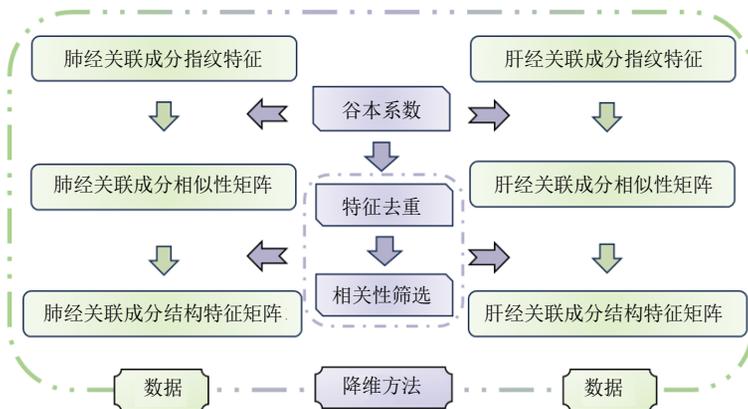


图 1 归经关联成分结构特征矩阵降维流程

Fig. 1 Dimensionality reduction process for matrix of structural characteristics of meridian-related components

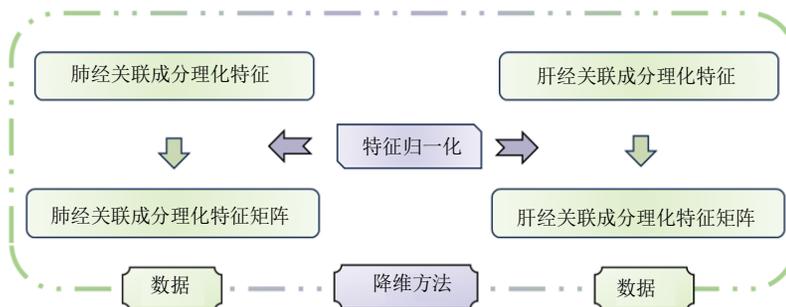


图 2 归经关联成分理化性质特征矩阵降维流程

Fig. 2 Dimensionality reduction process for matrix of physicochemical properties of meridian-related components

1.3 关联成分分类

本研究通过划分集合的方式，探索经络与药材间的复杂关系，探究不同经络药材所含活性成分的差异。以经络关联成分为基础，定义了 34 个活性成

成分分类集合，其中包含 15 个关联成分结构特征分类集合、9 个关联成分理化性质特征分类集合以及 10 个关联成分频数分类集合。应用 K 均值聚类算法^[19]，对两经络关联成分结构特征矩阵和理化性质特征矩

阵进行聚类分析。采用误差平方和 (sum of squared errors, SSE) 指标作为度量畸变程度的损失函数, 并结合肘部法确认最佳聚类个数 (k)。对于两经络关联成分的频数特征数据, 根据优势成分临界域方法, 绘制了关联成分-频数图, 采用肘点法选取优势成分临界域, 构建出频数分类集合。将 3 个维度的分类集合合并, 最终获得了关联成分分类集合。

1.4 集合相似度计算

为了探索活性成分属性在归属不同经络的药材中的差异, 揭示经络与药材之间复杂的关联, 本研究引入集合相似性的概念, 计算中药成分集合与关联成分分类集合之间的相似度, 从而获得关于药材含有的活性成分属性差异的量化指标。

余弦相似度算法在计算原理上能有效地量化 2 个向量在方向上的接近程度, 而忽略它们的绝对大小差异。即便 2 个中药的活性成分数量相差悬殊, 只要它们在活性成分的属性分布上相似, 余弦相似度就能高效地捕捉到这种相似性。本研究应用余弦相似度^[20]来计算药材成分集合与关联成分分类集合之间的相似度, 建立集合相似度矩阵作为建模数据集。余弦相似度使用 Sim_{set} 表示, 其计算公式如下。

$$Sim_{set} = \frac{t}{(uv)^2} \quad (2)$$

u 为药材成分集合中的成分个数, v 为关联成分分类集合中的成分个数, t 为两集合中相同的活性成分个数

1.5 药材归经判别模型构建

1.5.1 模型构建

基于 K 近邻算法 (K-nearest neighbors, KNN)、随机森林算法 (random forest, RF)、极端梯度提升算法 (eXtreme gradient boosting, XGBoost) 3 种人工智能算法构建了药材主归经络判别模型, 预测药材不同的主归经络属性。KNN 基于邻近样本, 通过计算新数据点与训练集中所有数据点的距离, 基于最近的 K 个邻居的类别进行投票得出最终结果^[21]。RF 通过随机的选取特征与样本构建决策树, 综合得到最终结果, 适用于小样本的情况^[22-23]。XGBoost 使用多个弱学习期迭代循环, 不断优化损失函数以构建出强大的模型, 具有很强的泛化能力以及鲁棒性^[24]。这 3 种算法通常用于分类、回归问题。

数据按照肺经、肝经药材的个数分层抽样。外层采用留出法, 80% 的数据被随机选择作为训练集, 剩余的 20% 被用作测试集。内层使用 5 折交叉验证,

采用网格搜索的方法进行参数调优。模型构建基于 R 语言 (R-4.3.1), 流程如图 3 所示。

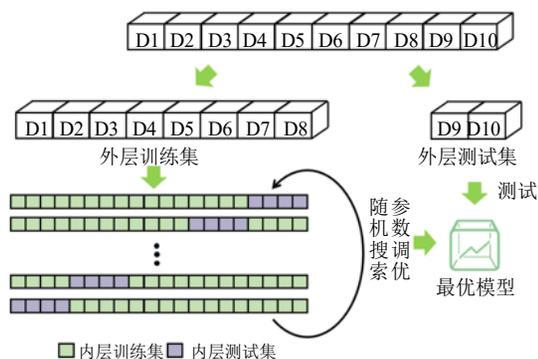


图 3 机器学习流程

Fig. 3 Workflow of machine learning

1.5.2 模型性能评估

通过构建预测结果的混淆矩阵, 计算真阳性 (TP)、真阴性 (TN)、假阳性 (FP) 和假阴性 (FN), 可以得到不同模型的精确率、召回率和 F1 值 (用于评估分类模型性能的指标), 从而对模型性能进行评估。

$$\text{精确率} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{召回率} = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = 2 \times [\text{精确率} \times \text{召回率} / (\text{精确率} + \text{召回率})] \quad (5)$$

除了使用准确率 (accuracy, acc) 对模型算法的判别情况进行评价以外, 本研究还计算了平衡准确率 (blanced accuracy, bacc), 用于排除不同经络数据不平衡对准确性的影响。此外, 还绘制了受试者工作特性曲线 (receiver operating characteristic curve, ROC), 并计算了曲线下面积 (area under curve, AUC), 实现全面评估比较不同机器学习模型的性能。

$$bacc = (\frac{TP}{TP+FN} + \frac{TN}{FP+TN}) / 2 \quad (6)$$

$$AUC = \int_{x=0}^1 TPR(FPR^{-1}(x)) dx \quad (7)$$

2 结果

2.1 药材数据收集

本研究共收集了 209 味药材, 包括 90 味归肺经药材和 119 味归肝经药材, 药材数量情况如表 1 所示, 构建了药材-归经属性-活性成分信息表 (表 2)。单归经药材共有 36 味, 其中归肺经药材有 20 味, 对应 1 197 个肺经关联成分; 归肝经药材有 16 味, 对应 490 个肝经关联成分。

表 1 肺、肝经与不同归经位数的药材数量

Table 1 Number of traditional Chinese medicinal materials in lung and liver meridians and different digits

主归经络	不同归经位数的药材数量					总计
	单经	双经	三经	四经	五经	
肺经	20	45	20	4	1	90
肝经	16	66	32	5	0	119
总计	36	111	52	9	1	209

表 2 单经药材归经信息

Table 2 Information of single meridian traditional Chinese medicinal materials

主归经络	药材
肺经	川射干、马勃、白前、百部、芥子、苍耳子、牡荆叶、金荞麦、鱼腥草、前胡、桔梗、射干、浮萍、桑白皮、款冬花、紫花前胡、紫苏子、紫菀、鹅不食草、锦灯笼
肝经	山楂叶、赤芍、青葙子、青黛、罗布麻叶、闹羊花、贯叶金丝桃、茜草、莲房、夏天无、海风藤、桑枝、断血流、密蒙花、蒺藜、蕤仁

2.2 特征变量选取

对于结构特征方面，基于活性成分的简化分子线性输入规范 (simplified molecular input line entry system, SMILES) 式，通过 Python 的 Rdkit 包获取 1 564 个单归经药材对应活性成分半径为 2 的 4 096 位 ECFP 指纹，构建肺经、肝经关联成分结构特征矩阵。

理化性质方面，利用 ChemSpider 数据库 (<http://www.chemspider.com/>) 收集单归经药材对应活性成分的理化性质，构建肺经、肝经成分理化性质特征矩阵。去除掉理化性质不完整以及搜索不到的活性成分，共获取 1 244 个活性成分对应的 22 个理化性质，如表 3 所示。

2.3 特征数据清洗

首先使用谷本系数分别对两经络关联成分的相似性进行计算，对特征进行去重后，使用斯皮尔曼相关系数计算特征之间的相关性以进行相关性筛选，阈值的选取使用肘点法，分别绘制了肺经、肝

经关联成分结构特征矩阵剩余特征数-阈值折线图，如图 4 所示。得到肺经、肝经关联成分结构特征相关性阈值分别为 0.85 和 0.90，最终保留肺经关联成分结构特征数量为 264，肝经关联成分结构特征数量为 185。对于关联成分的理化性质特征数据，对特征进行归一化处理，得到了大小为 983×22 的肺经关联成分理化性质特征矩阵，以及大小为 374×22 的肝经关联成分理化性质特征矩阵。

2.4 集合相似度矩阵构建

2.4.1 关联成分分类集合获取 应用 K 均值聚类算法，分别对两经络关联成分结构理化性质特征矩阵进行聚类，采用肘部法确认最佳聚类个数 *k*。如图 5、6 所示，从 SSE 折线图中观察到在 *k* 为 7、8、4、5 处出现拐点，即为最佳聚类数量。使用最佳聚类数量 *k* 分别对两矩阵进行聚类，获得了 15 个关联成分结构特征分类集合、9 个关联成分理化性质特征分类集合。

表 3 活性成分理化性质

Table 3 Physicochemical properties of active components

理化性质英文名称	理化性质中文名称	理化性质英文名称	理化性质中文名称
density	密度	LogP	油水分配系数
boiling point	沸点	LogD (pH 5.5)	分布系数 (pH 5.5 条件下)
vapour pressure	饱和蒸汽压	BCF (pH 5.5)	生物浓缩系数 (pH 5.5 条件下)
enthalpy of vaporization	气化焓	KOC (pH 5.5)	沉积物-水分配系数 (pH 5.5 条件下)
flash point	闪点	LogD (pH 7.4)	分布系数 (pH 7.4 条件下)
index of refraction	折射率	BCF (pH 7.4)	生物浓缩系数 (pH 7.4 条件下)
molar refractivity	摩尔折射率	KOC (pH 7.4)	沉积物-水分配系数 (pH 7.4 条件下)
H bond acceptors	氢键受体	polar surface area	极性表面积
H bond donors	氢键供体	polarizability	极化率
freely rotating bonds	自由旋转键	surface tension	表面张力
rule of 5 violations	违反五原则的个数	molar volume	摩尔体积

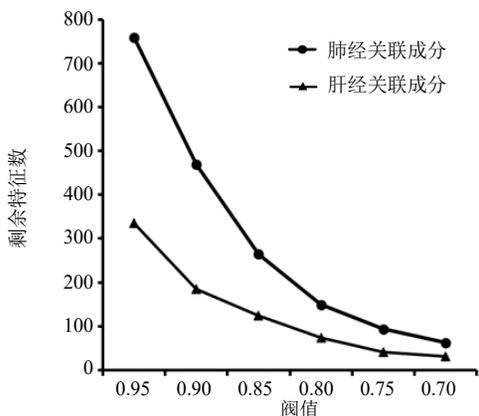


图4 肺经、肝经关联成分结构特征矩阵剩余特征数-阈值折线图

Fig. 4 Residual characteristic number-threshold values line chart for matrix of structural characteristics of lung and liver meridian-related components

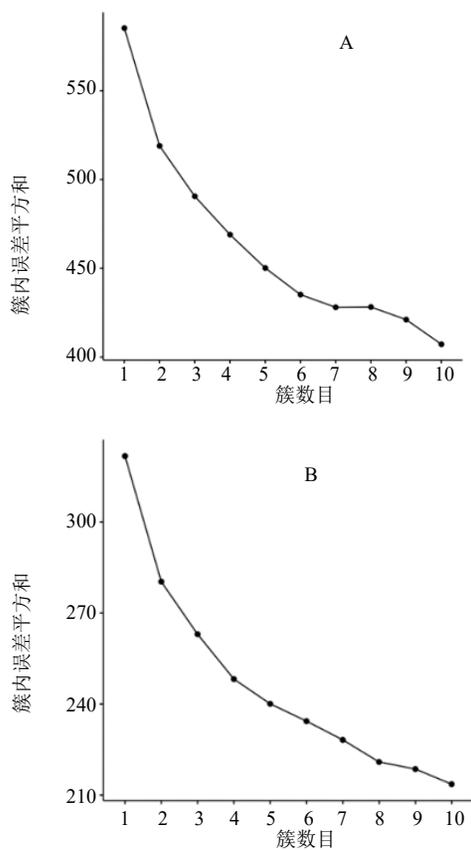


图5 肺经 (A) 和肝经 (B) 关联成分结构特征误差平方和-簇数折线图

Fig. 5 SSE-cluster number line plot of structural characteristics of lung (A) and liver (B) meridian-related components

对于成分频数特征方面, 本研究根据关联成分出现频数定义了10个集合(S1~S10)。根据关联成

分个数与关联成分在肺经、肝经单归经药材中的出现频数之间关系的折线图, 如图7、8所示, 得到肺

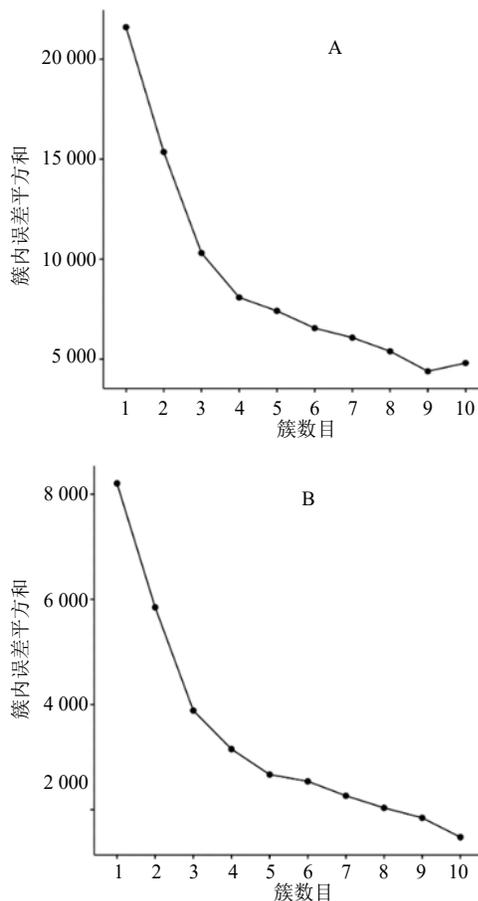


图6 肺经 (A) 和肝经 (B) 关联成分理化性质特征误差平方和-簇数折线图

Fig. 6 SSE-cluster number line plot of physicochemical properties of lung (A) and liver (B) meridian-related components

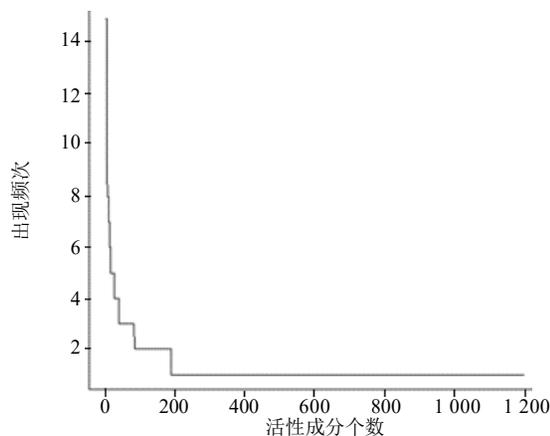


图7 肺经关联成分优势成分临界域获取

Fig. 7 Determining dominant component critical threshold of lung meridian-related components

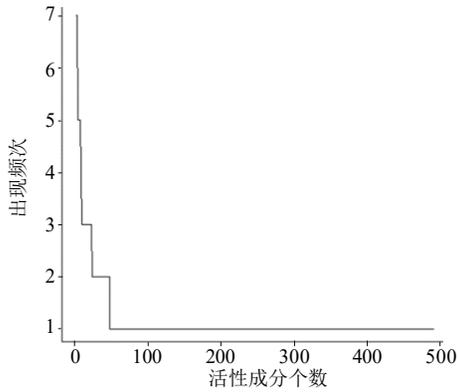


图8 肝经关联成优势成分临界域获取

Fig. 8 Determining dominant component critical threshold of liver meridian-related components

经优势成分临界域为4，肝经优势成分临界域为3，将优势成分的阈值作为分类标准，对关联成分进行分类，10个关联成分频数分类集合之间关系如图9所示，其中S1、S2为肺经、肝经关联成分全集，S3、S4为肺经、肝经优势成分全集，S5、S6为交集

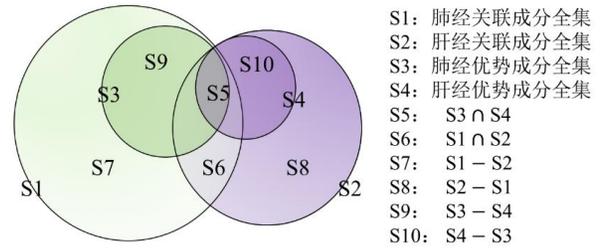


图9 关联成分频数分类集合关系

Fig. 9 Relationship of frequency classification sets of lung and liver meridian-related components

集合，S7、S8为补集集合。

2.4.2 集合相似度计算 应用余弦相似度对药材活性成分集合与关联成分分类集合进行集合相似度计算，最终得到集合相似度矩阵作为建模数据集。如图10、11所示，药材活性成分与关联成分分类集合的相似度数据具有一定差异，可以作为构建药材主归经络判别模型的基础数据。

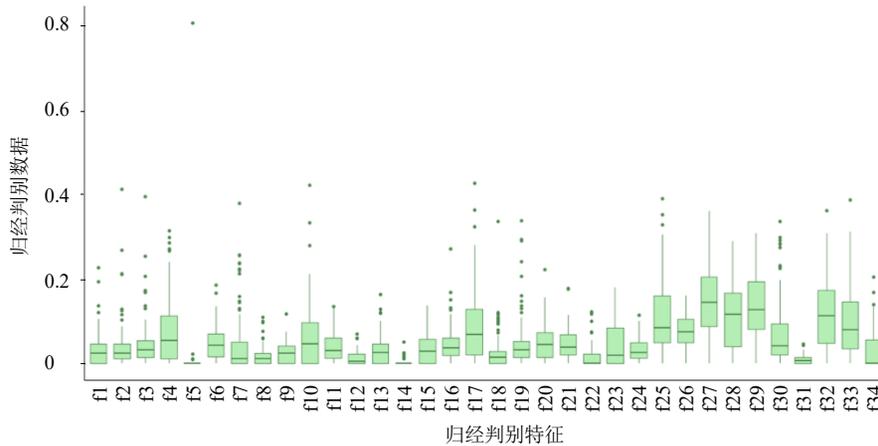


图10 归肺经药材集合相似度特征箱形图

Fig. 10 Boxplot of set similarity features for lung meridian targeted traditional Chinese medicinal materials

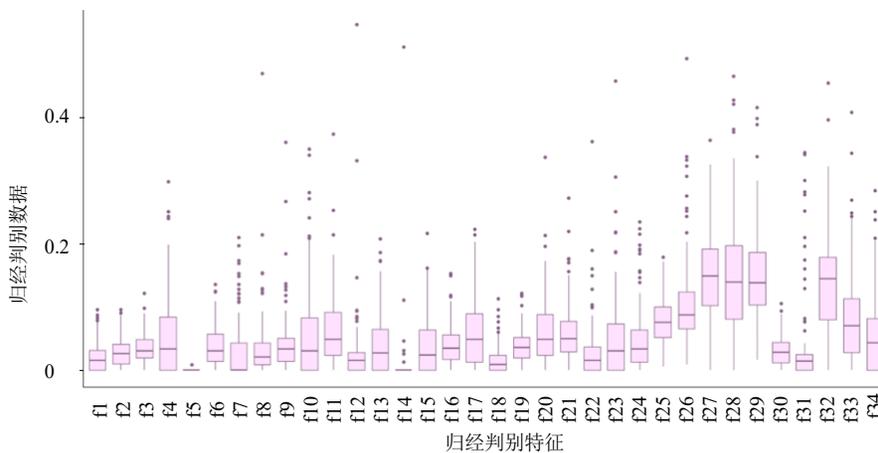


图11 归肝经药材集合相似度特征箱形图

Fig. 11 Boxplot of set similarity features for liver meridian targeted traditional Chinese medicinal materials

2.5 判别模型的构建及性能评估

本研究使用集合相似性矩阵作为建模数据集，其中归肺经药材数据集包含 34 个特征，共 90 个药材样本，归肝经药材数据集包含 34 个特征，共 119 个药材样本，合并共 209 个药材样本，34 个特征。每个药材样本的标签为“肺经”或“肝经”，表示该药材主归经路的类别。并基于 KNN、RF、XGBoost 3 种机器学习算法建立模型。针对 KNN 算法，使用欧氏距离作为距离度量算法，对参数 k 进行调参，其中参数 k 的范围为 1~132，步长为 1；针对 RF 算法，针对 trees、mtry 2 个参数进行调参，trees 的调参范围为取 1~50 的 10 个等间隔的值，mtry 的调参范围为取 1~30 的 10 个等间隔的值；针对 XGBoost 算法，对 eta、nrounds、max_depth 3 个参数进行调参，调参设置为 eta、nrounds、max_depth 分别在 0.01~0.5、1~30、1~10，各取 10 个等间

隔的数值。3 种机器学习算法最优参数设置情况如表 4 所示。不同模型的性能评估结果如表 5 所示，模型的 ROC 曲线如图 12 所示。

由表 5 可以看出，3 种机器学习模型在区分药材的主归经路方面都表现出了不错的性能，bacc 以及 AUC 均达到了 0.8 以上，其中 KNN 表现较差，XGBoost 表现相对较好，RF 算法的表现最佳，RF

表 4 机器学习模型最优超参数设置情况

Table 4 Optimal hyperparameter settings for machine learning models

机器学习算法	超参数设置
K 近邻算法	最近邻居数设置为 44
随机森林算法	树的数量设置为 17，特征选择数为 30
极端梯度提升算法	学习率设置为 0.45，最大深度设置为 1，迭代次数设置为 14

表 5 不同机器学习模型性能评估结果

Table 5 Performance evaluation results of different machine learning models

主归经路	模型	平衡准确率 (bacc)	精确度	召回率	F1 分数
肺经	KNN	0.80	0.81	0.72	0.76
肝经	KNN	0.80	0.81	0.88	0.84
肺经	RF	0.86	0.80	0.89	0.84
肝经	RF	0.86	0.91	0.83	0.87
肺经	XGBoost	0.84	0.93	0.72	0.81
肝经	XGBoost	0.84	0.82	0.96	0.88

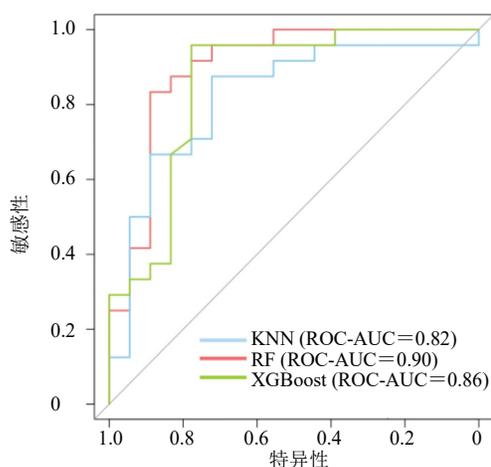


图 12 不同算法 ROC 曲线

Fig. 12 ROC curves of different algorithms

模型的平衡准确率、AUC 分别达到了 0.86、0.90，说明 RF 模型在对于药材的主归经路判别方向具有

较好的判别能力。

3 讨论

随着“一带一路”倡议的深入实施，我国与欧美、日本、东南亚、非洲等多个天然药物市场之间的交流与合作日益频繁。外来中药已然成为中药学领域中不可或缺的一部分，为中医药的丰富和发展注入了新的活力。为了能够更好地运用外来中药，探索外来中药的主治病症，提高用药的准确性，实现外来中药的良性运用和研究，关键在于为其赋予中药药性，尤其是归经的属性。

本研究以肺经、肝经药材对应活性成分为基础，探索了归属不同经路的药材所含有的活性成分的属性分布规律和归经理论的原理，使用人工智能算法对属性分布规律与主归经路间的关系进行学习，构建了药材主归经路判别模型。

在方法的构建方面，本研究首次提出了关联成

分分类集合概念,将集合的概念引入经络分类的研究中,并建立了量化不同经络药材间的活性成分属性差异的方法,利用人工智能算法,多维度地探究了药材活性成分与药材经络归属之间的联系,符合中医药理论的整体性、系统性思想^[25]。本研究使用药材完整的活性成分数据,探索药材主归经络与活性成分间关系,从宏观的角度探索了药材归经属性的理论基础。本研究构建的3个模型平衡准确率均高于0.8,验证了提出方法的准确性,同时,最佳模型的平衡准确率、AUC分别达到了0.86、0.90,证明了模型的准确性和稳定性,说明人工智能算法可以实现肺经与肝经药材主归经络的准确判别,具有较好的应用价值与泛化能力。本研究构建的RF模型可以较为准确地判断药材主归经络,并且可以说明药材所含活性成分与药材的经络归属间存在联系,药材所含活性成分的特征差异是药材归经属性的基础。同时,本研究所提出的方法为后续推广至其余经络的预测提供了研究基础,为未知经络归属或归经信息混乱的药材提供主归经络参考,为解决中药归经属性记载一致性差的问题提供了新的思路。

本研究希望使用人工智能算法,通过构建主归经络判别模型,探索药材归经理论的基础原理,经研究发现多经药材与单经药材相比,在含有的活性成分属性上更加复杂,会出现2种甚至多种的经络类别偏向,不易判别主归经络属性,导致模型无法学习到足够特征来区分它们,因此准确率相对较低。又由于在收集数据过程中发现不同古籍中对于药材的归经信息常有区别,中医强调辨证施治,不同的医家或学派基于各自的经验和理解,对相同的中药也会有不同的归经属性分类判断。为了保证标准化的药材归经描述,本研究以《中国药典》2020年版作为药材的归经信息数据来源,应用TCMSP数据库收集药材活性成分,获取的数据具有一定的局限性,因此,模型难以完整学习到不同经络药材之间的特征差异。针对本研究存在的问题,课题组会在后续研究中加入更全面的活性成分特征数据,如指纹图谱、实验数据、药材的吸收、分布、代谢、排泄和毒性等药理性质,使关联成分集合分类更加完善;通过丰富数据来源,增加双经、多经药材的数据多样性和量级,提升模型的学习能力;应用深度学习、自适应增强等更先进的人工智能算法来捕捉双经或多经药材之间的特征差异。逐步提高模型的

准确率和泛化能力,为中药材归经研究提供更加全面深入的研究思路,为中医药归经理论的量化研究、推广中药归经理论的临床应用以及中医药的国际化推广提供更有利的支撑。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 韩森,吕爱平,李健,等.网络药理学在中药药性理论研究中的应用概述[J].中国中医基础医学杂志,2019,25(1):127-130.
- [2] Chen Z, Cao Y F, He S B, *et al.* Development of models for classification of action between heat-clearing herbs and blood-activating stasis-resolving herbs based on theory of traditional Chinese medicine [J]. *Chin Med*, 2018, 13: 12.
- [3] 肖斌,陶欧,罗计,等.中药药性与功能靶点的关系[J].中西医结合学报,2011,9(7):789-793.
- [4] 张华敏,杨健,刘思鸿,等.中药归经理论的研究方法及分子描述在中药归经中的研究探讨[J].中国医院用药评价与分析,2019,19(12):1518-1520.
- [5] 金锐,张冰.复杂性科学视野下的中药药性理论——药性形成的多源性[J].中西医结合学报,2012,10(11):1198-1205.
- [6] Shen Q, Wu Y, Li Z, *et al.* Quality assessment of traditional Chinese medicine using quality biomarkers: Compound Danshen dripping pills as an example [J]. *Phytochem Anal*, 2023, 34(5): 580-593.
- [7] Liu S, Liang Y Z, Liu H T. Chemometrics applied to quality control and metabolomics for traditional Chinese medicines [J]. *J Chromatogr B Analyt Technol Biomed Life Sci*, 2016, 1015/1016: 82-91.
- [8] Zhuo L M, Peng J J, Zhao Y L, *et al.* Screening bioactive quality control markers of Qishen Yiqi dripping pills based on the relationship between the ultra-high performance liquid chromatography fingerprint and vascular protective activity [J]. *J Sep Sci*, 2017, 40(20): 4076-4084.
- [9] Jin M L, Ren W W, Zhang W G, *et al.* Exploring the underlying mechanism of shenyankangfu tablet in the treatment of glomerulonephritis through network pharmacology, machine learning, molecular docking, and experimental validation [J]. *Drug Des Devel Ther*, 2021, 15: 4585-4601.
- [10] Yang S L, Shen Y J, Lu W D, *et al.* Evaluation and identification of the neuroprotective compounds of xiaoxuming decoction by machine learning: A novel mode to explore the combination rules in traditional Chinese medicine prescription [J]. *Biomed Res Int*, 2019, 2019: 6847685.
- [11] Wang Y Y, Jafari M, Tang Y, *et al.* Predicting Meridian in

- Chinese traditional medicine using machine learning approaches [J]. *PLoS Comput Biol*, 2019, 15(11): e1007249.
- [12] Yeh H Y, Chao C T, Lai Y P, *et al.* Predicting the associations between meridians and Chinese traditional medicine using a cost-sensitive graph convolutional neural network [J]. *Int J Environ Res Public Health*, 2020, 17(3): 740.
- [13] 王小雪, 卢杉, 郑思悦, 等. 归经中药化学成分、药理作用及临床应用的实证分析 [J]. *中华中医药杂志*, 2018, 33(11): 5193-5197.
- [14] Zhang S, Zhang X W, Du J Y, *et al.* Multi-target meridians classification based on the topological structure of anti-cancer phytochemicals using deep learning [J]. *J Ethnopharmacol*, 2024, 319(Pt 2): 117244.
- [15] Ru J L, Li P, Wang J N, *et al.* TCMSP: A database of systems pharmacology for drug discovery from herbal medicines [J]. *J Cheminform*, 2014, 6: 13.
- [16] Muegge I, Mukherjee P. An overview of molecular fingerprint similarity search in virtual screening [J]. *Expert Opin Drug Discov*, 2016, 11(2): 137-148.
- [17] Rogers D, Hahn M. Extended-connectivity fingerprints [J]. *J Chem Inf Model*, 2010, 50(5): 742-754.
- [18] Willett P. Similarity-based virtual screening using 2D fingerprints [J]. *Drug Discov Today*, 2006, 11(23/24): 1046-1053.
- [19] Ikotun A M, Ezugwu A E. Boosting k-means clustering with symbiotic organisms search for automatic clustering problems [J]. *PLoS One*, 2022, 17(8): e0272861.
- [20] Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? [J]. *J Cheminform*, 2015, 7: 20.
- [21] Deng Z Y, Zhu X S, Cheng D B, *et al.* Efficient kNN classification algorithm for big data [J]. *Neurocomputing*, 2016, 195: 143-148.
- [22] Fralick M, Colak E, Mamdani M. Machine learning in medicine [J]. *N Engl J Med*, 2019, 380(26): 2588-2589.
- [23] Breiman L. Random forests [J]. *Mach Learn*, 2001, 45: 5-32.
- [24] Yuan K C, Tsai L W, Lee K H, *et al.* The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit [J]. *Int J Med Inform*, 2020, 141: 104176.
- [25] Jiang M, Lu C, Zhang C, *et al.* Syndrome differentiation in modern research of traditional Chinese medicine [J]. *J Ethnopharmacol*, 2012, 140(3): 634-42.

[责任编辑 潘明佳]