

基于化学特征相似性与一致性检验分析中药属性-成分关系

罗奕宇, 尚 晶, 杨晓艳, 丁长松*

湖南中医药大学信息科学与工程学院, 湖南 长沙 410208

摘要:目的 整合化学成分、化学特征探究中药属性与化合物之间的关联性。方法 从公共数据平台收集常用中药的化学成分及结构信息, 计算成分指纹及其相似性, 构建中药属性的化学特征向量, 分析不同属性分类的化学特征。通过 BM25-adpt 算法评价化学特征对中药属性的贡献度, 并计算成分种类得分, 并选取活血药关键成分进行富集分析。采用 Kendall's W 一致性检验评估药性、药味、归经及功效分类成分种类得分一致性。结果 共获得 437 味中药的 44 279 个成分, 以 0.5 为相似度阈值筛选出 5 016 个成分种类。Kendall's W 一致性检验结果显示, 4 个属性的 *c_score* 得分集存在显著差异。余弦相似度分析进一步表明, 不同程度药性分类的化学特征存在差异。药味和归经的复合分类中有相同组成的分类在化学特征上存在一定的联系, 且药味中苦与微苦相似度较高。针对“活血行气止痛”与“活血调经”2 类功效的代表性成分进行基因本体 (gene ontology, GO) 功能和京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG) 通路富集分析, 发现其药理作用可能通过神经活性配体-受体相互作用、类固醇激素生物合成过程、调节氧化还原酶活性等过程实现。结论 中药属性与其化学成分之间存在关联, 每种属性分类存在关键成分群。不同程度的药性分类不可简单视为一类, 药味及归经分类中的复合分类存在一定关联, 相似作用趋势的功效分类在化学特征上体现出一定的相似性。研究结果为中药药效物质基础研究及方剂配伍机制的现代化阐释提供了数据支持。

关键词: 中药; 药性; 功效; 化学特征; Kendall's W 一致性检验

中图分类号: TP18; R285.1 文献标志码: A 文章编号: 0253-2670(2026)05-1814-11

DOI: 10.7501/j.issn.0253-2670.2026.05.019

Analysis of relationship between traditional Chinese medicine properties and components based on chemical characteristic similarity and consistency test

LUO Yiyu, SHANG Jing, YANG Xiaoyan, DING Changsong

School of Information Science and Engineering, Hunan University of Chinese Medicine, Changsha 410208, China

Abstract: Objective To investigate the association between traditional Chinese medicine (TCM) properties and chemical compounds by integrating chemical composition and chemical characteristics. **Methods** Chemical constituents and structural data of commonly used TCMs were collected from public databases. Molecular fingerprints and pairwise similarities were computed to construct chemical feature vectors representing TCM properties, enabling comparative analysis across different property categories. The BM25-adpt algorithm was applied to assess the contribution of chemical features to each TCM property, yielding component-type scores. Key components of blood-activating herbs were selected for functional enrichment analysis. Kendall's W concordance test was used to evaluate the consistency of component-type scores across four property dimensions: medicinal properties, medicinal flavors, meridian tropism, and therapeutic functions. **Results** A total of 44 279 chemical components from 437 TCMs were compiled. At a structural similarity threshold of 0.5, these were clustered into 5 016 distinct component types. Kendall's W test revealed significant discrepancies among the *c_score* distributions across the four property categories. Cosine similarity analysis further demonstrated that different levels of medicinal properties exhibit distinct chemical profiles. Composite classifications within medicinal flavors and meridian tropism that share identical constituent sets showed notable chemical relatedness; in particular, “bitter” and “slightly bitter” flavors exhibited high similarity. Enrichment analyses of gene ontology (GO) functions and Kyoto encyclopedia of genes and genomes (KEGG) pathways were performed on representative components of the two efficacies “blood activating, *qi*-moving and analgesic” and “blood activating to regulate menstruation” implicated key biological processes including neuroactive ligand-receptor interactions, steroid hormone

收稿日期: 2025-11-13

基金项目: 国家自然科学基金面上项目 (82474352); 湖南省中医药管理局重点项目 (A2024011)

作者简介: 罗奕宇, 男, 硕士研究生, 研究方向为中医药大数据分析。

*通信作者: 丁长松, 男, 教授, 博士研究生导师。E-mail: dingcs1975@hnuucm.edu.cn

biosynthesis, and regulation of oxidoreductase activity. **Conclusion** This study establishes a measurable link between TCM properties and their underlying chemical constituents, with each property category characterized by specific clusters of key components. Different gradations of medicinal properties should not be treated as homogeneous, composite categories within medicinal flavors and meridian tropism show meaningful chemical associations; and therapeutic functions with similar clinical actions demonstrate convergent chemical features. These findings provide a data-driven foundation for elucidating the material basis of TCM efficacy and advancing the modern understanding of herbal compatibility mechanisms.

Key words: traditional Chinese medicine; medicinal properties; efficacy; chemical characteristics; Kendall's W consistency test

中药药性及功效等理论是中药学的基础, 这些属性系统概括了中药对人体的作用特点。中药发挥作用依赖于含有的化学成分, 其空间构型、原子组成及化学键性质共同决定了成分的理化特性与生物活性^[1]。近年来, 研究者开始关注中药属性与成分的关系, “分子药性学假说”认为中药的药性特征与其所含化学成分的结构与性质之间存在内在关联^[2]。“性-效-物质”三元论试图构建中药药效、药性与化学成分之间的对应关系^[3-4]。然而, 现有研究多集中于单一属性(如寒、热药性)的物质基础分析^[5]或侧重于预测模型的构建^[6-10], 缺乏对多维度属性的化学成分群及分子结构特征的探讨。

本研究将中药 4 大基本属性纳入分析框架, 从“属性-成分-结构”的视角系统分析中药属性与化学特征的内在关联。研究选取常用中药为研究对象,

整合公共数据库中的化学成分信息, 基于分子结构相似性对成分进行聚类分析, 并结合 BM25-adpt 算法挖掘不同属性类别下成分群的分布特征与结构偏好, 识别具有代表性的化学成分群, 为中药属性理论的科学内涵阐释提供数据支持与分子层面的证据。研究流程见图 1。

1 数据来源与方法

1.1 中药的筛选及成分获取

本研究选取全国中医药行业高等教育“十三五”规划教材《中药学》中记载的常用中药为研究对象, 在 HERB (本草组鉴, <http://herb.ac.cn/>)^[11]、Herbal Ingredients' Targets Platform (HIT, <http://www.badd-cao.net:2345/>)^[12-13]、Encyclopedia of Traditional Chinese Medicine (ETCM, <http://www.tcmip.cn/ETCM2/front/#/>)^[14]、中药系统药理学平台

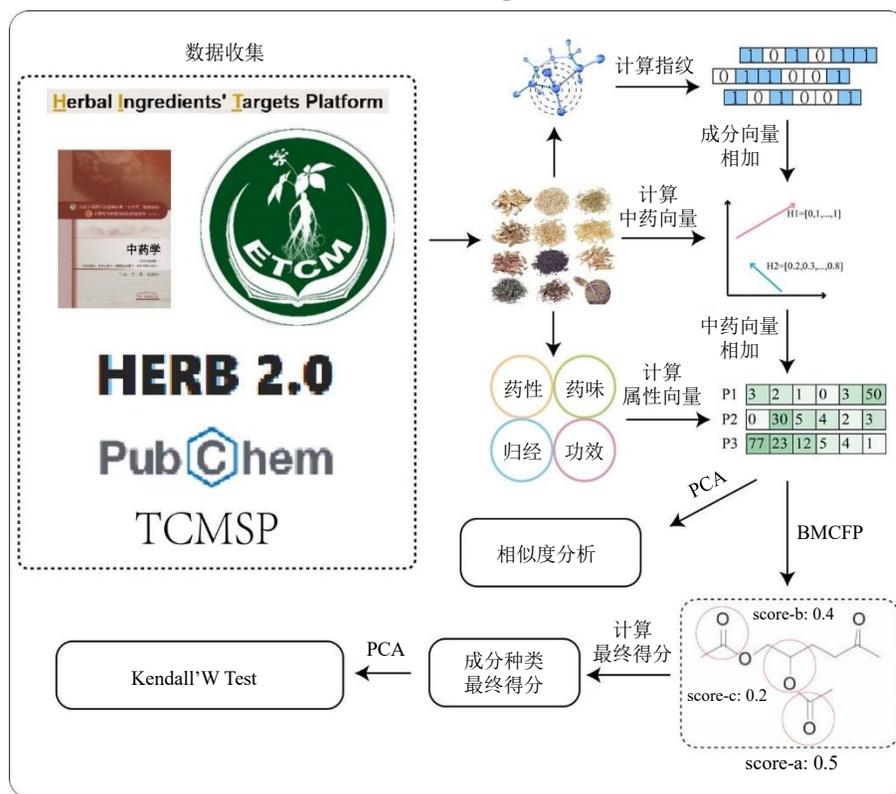


图 1 研究流程图

Fig. 1 Research flow chart

(TCMSP, <https://www.91tcmsp.com/#/home>)^[15]以及 PubChem (<https://pubchem.ncbi.nlm.nih.gov/>)中收集每味中药的所含的成分及成分的简化分子线性输入系统 (simplified molecular input line entry system, SMILES)。

1.2 特征向量构建

1.2.1 成分向量构建 去除重复及无结构信息的中药成分,基于中药成分的 SMILES 计算成分指纹,得到成分的二进制向量。第 i 个中药第 c 个成分向量记作 m_{ic} , 其表示如下。

$$m_{ic} = (v_{ic1}, v_{ic2}, v_{ic3}, \dots, v_{ic2048}) \quad (1)$$

v_{icj} 表示第 i 个中药第 c 种成分在第 j 维指纹特征上的取值

1.2.2 中药化学特征向量构建 为表征中药的化学特征将其所含各成分的向量相加。第 i 个中药的化学特征向量 (H_i) 表示如下。

$$H_i = \sum_{c=1}^{m_i} m_{ic} \quad (2)$$

m_i 表示第 i 个中药所含的成分数

1.2.3 属性分类向量构建 通过将归属于同一属性分类下的所有中药向量相加得到该属性分类的向量。第 k 类属性分类的向量 (P_k) 表示如下。

$$P_k = \sum_{i \in S_k} H_i \quad (3)$$

S_k 表示具有第 k 个属性的中药集合

1.3 关键化学特征评价

BM25-adpt^[16]是基于信息检索经典算法 BM25 的改进,具有更优的性能与鲁棒性。为评估不同成分种类和化学特征对属性的重要性,将其核心思想扩展应用至本领域,即每个属性分类视为一个“文档”,每种化学特征视为一个“词”,通过计算化学特征在特定属性分类中出现的频数得到与属性最匹配的化学特征 (best matching chemical feature for properties, BMCFP) 得分,用以量化其重要性。

1.3.1 归一化化学特征频率 为消除不同属性分类在规模上的差异,需对每个属性分类中的化学特征原始频次进行长度归一化处理。例如,对于属性分类 k 中的化学特征 d ,其归一化频率 $p'(k,d)$ 计算公式如下。

$$p'(k,d) = \frac{p(k,d)}{1-b+b \frac{N_k}{\bar{N}}} \quad (4)$$

$p(k,d) = P_k[d]$ 表示特征位点 d 在属性分类 k 中出现次数;

$N_k = \sum_{d=1}^{2048} P_k[d]$ 表示属性分类 k 所有特征出现的总频次; \bar{N} 表

示所有属性分类的平均长度,即 N_k 的平均值; b 为控制长度归一化强度的参数

1.3.2 化学特征增益 (chemical feature gain, CG) CG 用于衡量特定化学特征 (指纹位点) 对区分不同属性分类的贡献度。第 d 个特征位点在所有属性分类中出现 t 次的增益 (CG'_d) 定义如下。

$$CG'_d = -\log_2 \left(\frac{cf_d^t + 0.5}{K+1} \right) + \log_2 \left(\frac{a_d \cdot cf_d^{t+1} + 0.5}{a_d \cdot cf_d^t + 1} \right) \quad (5)$$

K 表示属性分类的总数, cf_d^t 表示特征位点 d 在所有属性分类中至少出现 t 次的统计频数

a_d 为特征 d 的调整因子,用于反映该特征在不同属性分类中频次分布的离散程度,其本质是特征 d 在所有属性分类中频数的方差加 1,具体计算公式如下。

$$a_d = \frac{1}{K} \sum_{k=1}^K [p'(k,d) - \mu_d]^2 + 1 \quad (6)$$

$\mu_d = \frac{1}{K} \sum_{k=1}^K p'(k,d)$ 为特征 d 在所有属性分类中归一化频率的平均值

1.3.3 自适应参数 (k_1) 为在 BM25 框架下实现参数 k_1 随化学特征进行自适应调整,使用最小二乘法估计化学特征 d 的最佳 k_1 , 算法方法如下。

$$k_1(d) = \arg \min_{k_1} \sum_{t=1}^{2048} \left(\frac{CG'_d}{CG_d^1} - \frac{(k_1+1) \cdot t}{k_1+t} \right)^2 \quad (7)$$

1.3.4 BMCFP 得分 为评估化学特征 d 在特定属性分类 k 中的重要性,引入 BMCFP (k,d) 得分,计算公式如下。

$$\text{BMCFP}(k,d) = \frac{p'(k,d)(k_1(d)+1)}{k_1(d)+p'(k,d)} \cdot CG_d^1 \quad (8)$$

$k_1(d)$ 表示化学特征 d 的自适应参数

1.4 成分及成分种类得分计算

1.4.1 计算成分对属性分类重要性得分 为量化化学成分对属性分类的重要性,定义化学成分 c 对属性分类 k 重要性得分 $i_score(k,c)$ 。该得分通过对化学成分指纹向量与对应属性分类的位点权重向量间的点积得到,具体公式如下。

$$i_score(k,c) = \sum_{d=1}^{2048} m(c,d) \cdot \text{BMCFP}(k,d) \quad (9)$$

$m(c,d)$ 表示化学成分 c 的指纹向量在第 d 个特征位点的值
复杂程度不同的化学成分向量中为 1 的位点数量不一样,使得不同化学成分得分难以比较,为解决这一问题,在得到一个属性分类的全部成分得分

后进行 Yeo-Johnson 变换。

1.4.2 构建成分化学相似性矩阵 基于成分间的化学相似性构建成分化学相似性矩阵，化学相似性使用 Tanimoto 相似性系数。任意 2 种化学成分 m_a 与 m_b 之间的相似度定义如下。

$$T(m_a, m_b) = \frac{m_a \cdot m_b}{\|m_a\|^2 + \|m_b\|^2 - m_a \cdot m_b} \quad (10)$$

$m_a \cdot m_b$ 表示成分向量 m_a 与 m_b 的点积， $\|m_a\|^2$ 和 $\|m_b\|^2$ 分别表示向量 m_a 和 m_b 的平方模长

1.4.3 计算成分种类得分 采用 complete-linkage 层次聚类，以 Tanimoto 相似度为度量，并设定阈值 $\tau=0.5$ 对成分进行分类。记成分种类 C_m 对属性 k 的算术平均值为 $c_score(k, C_m)$ ，其计算公式如下。

$$c_score(k, C_m) = \frac{1}{|C_m|} \sum_{c \in C_m} i_score(k, c) \quad (11)$$

$|C_m|$ 表示成分种类 C_m 中包含的化学成分数量

1.5 数据降维

为了降维去噪与保留主要变异结构，为后续相似性计算提供稳定输入，使用主成分分析 (principal component analysis, PCA) 对属性分类向量和成分种类得分矩阵进行降维。在进行 PCA 之前，首先对原数据按特征进行 Z-score 标准化处理，以消除不同属性间量纲和方差差异对主成分方向的影响。然后，对原始数据进行中心化处理，计算协方差矩阵，对矩阵进行特征值分解，求得特征向量和对应的特征值，选择前 k 个最大特征值对应的特征向量作为主成分方向，将原始数据投影到由选取的 k 个主成分所构成的子空间中。

1.6 成分种类得分差异性检验

对降维后的成分种类得分矩阵采用肯德尔 W 检验 (Kendall's W test) 评估不同属性标签下各类别得分排序的一致性。一致性检验数 (W) 具体公式如下。

$$W = \frac{12 \times \sum_{m=1}^n (R_m - \bar{R})^2}{N_p^2 (T_{\max}^3 - T_{\min})} \quad (12)$$

m 为第 m 个成分种类或化学特征的秩次总和， \bar{R} 为平均秩次， N_p 为当前属性中属性标签数量， T_{\max} 为 2 048，表示化学特征数量

1.7 属性分类化学组成关联性分析

为从化学特征的角度评价不同属性标签之间的关系，研究计算属性分类向量降维后的结果中不同属性分类之间的余弦相似度，计算公式如下。

$$\cos(P_e, P_l) = \frac{P_e \cdot P_l}{\|P_e\| \|P_l\|} \quad (13)$$

$P_e \cdot P_l$ 表示向量 P_e 和 P_l 的点积， $\|P_e\|$ 和 $\|P_l\|$ 分别为向量 P_e 、 P_l 的模长

1.8 关键成分群筛选及基因本体 (gene ontology, GO) 和京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG) 富集分析

关键成分群的筛选与计算：首先，在每个属性分类 k 中，选取 $c_score(k, C_m) > 0$ 的成分种类 C_m ；其次，统计种类 C_m 在属于该属性分类 k 的中药中出现的频数 $F(k, C_m)$ ，并据此计算其最终得分；最终，选取得分排名前 10 的成分种类，构成该属性分类的关键成分群。其中种类 C_m 对属性分类 k 的最终得分 $Fc_score(k, C_m)$ 按以下公式计算。

$$Fc_score(k, C_m) = F(k, C_m) \cdot c_score(k, C_m)^2 \quad (14)$$

为探究关键成分群发挥功效的作用机制，以功效分类中活血行气止痛药和活血调经药的关键成分群为例，通过数据库查找成分群中成分的作用靶点，导入 Metascape 数据库进行 GO 功能和 KEGG 通路富集分析。以 $P < 0.01$ 为筛选条件，将排名前 10 条的条目利用微生信平台^[16]进行可视化。

2 结果

2.1 中药筛选及成分获取

检索《中药学》中的 47 类 443 味中药，在 3 个数据库检索每味中药的成分，共收集 62 272 个成分，基于获取的成分名在 HERB、TCMSP、HIT 数据库和 PubChem 数据库检索成分的 SMILES 表达式，剔除 8 127 个无 SMILES 式的成分。合并 3 个数据库的结果后使用 Python 软件的 RDKit 库计算上述成分的 ECFP4 指纹，对中药所含成分以 ECFP4 指纹为标准进行去重，共获得 437 味中药的 44 279 个成分，14 177 个独特成分。检索时间截至 2024 年 8 月 25 日。

2.2 数据预处理

存在多个药味和归经的中药拆分为单个分类，如白芍苦酸将其同时归入苦味和酸味。由于无法确认寒与凉、热与温的区别是否仅在于偏性程度，计算时不将药性简单分为寒与热，而是单独计算不同分类。去除成分数量少于 5 的中药和中药数量少于 5 的属性分类，最终得到 437 味中药的 7 个药性、8 个药味、11 个归经、39 个功效。

2.3 特征向量构建及关键化学特征评价

从成分指纹得到 14 177 个成分的 2 048 位二进制向量 m_{ic} 。基于公式 (2) ~ (3)，得到 4 个属性共

65 个属性向量。根据公式 (4) ~ (8) 分别计算所有化学特征在 4 种属性不同属性分类中的 BMCFP 得分, 得到 4 种属性的 2 048 位化学特征得分矩阵。

2.4 结果检验及关键成分种类筛选

根据公式 (10) 得到 14 177×14 177 大小的相似性矩阵, 再使用 complete-linkage 层次聚类并设定相似度阈值 $\tau=0.5$, 将全部成分划分为 5 016 个成分种类, 对成分种类进行编号, 以编号代表不同成分种类。基于公式 (9) 和公式 (11) 计算成分种类最终得分, 得到 5 016 个成分种类对不同属性分类的 c_score 。对每个属性的 c_score 得分矩阵先进行 Z-score 标准化, 随后采用 Python 软件 sklearn 库中的 PCA 方法对标准化后的数据进行降维, 设定累计方差贡献率为 95%, 自动确定所需保留的主成分数目, 最终药性得到 3×7 的降维矩阵, 药味得到 4×8 的降维矩阵, 归经得到 4×8 的降维矩阵, 功效得到 20×39 的降维矩阵。

根据公式 (12) 对 4 个降维矩阵采用 Kendall's W 检验 (表 1), 结果显示, 3 个属性的 P 值均大于 0.05, 一致性未达到统计显著性水平, 无法拒绝不同属性分类的结果为随机的零假设, 说明不同属性分类的得分不完全一致, 具有差异性。而功效虽 P 值小于 0.001, 但其一致性系数极小, 说明功效不同属性分类的得分不完全一致。

表 1 不同属性分类的 Kendall's W 检验结果

属性	一致性系数	P 值
归经	0.114	>0.05
功效	0.086	<0.001
药味	0.206	>0.05
药性	0.245	>0.05

根据公式 (14) 计算成分种类的最终得分并筛选每个属性分类的关键成分种类。药性、药味、归经属性仅展示中药数量超过中位数的分类, 功效属性展示中药数量前 5 的分类以及活血行气止痛药和活血调经药, 结果如表 2~5 所示。

2.5 属性分类化学组成关联性分析

对属性分类向量的 2 048 个特征进行 PCA 降维, 将药性和药味的特征维度降到 4 维, 归经降到 6 个维度, 功效变为 26 个维度。

药性余弦相似度结果 (图 1-A) 显示, 凉与热、微温相似度高。降维结果显示, 三者 PCA1 及 PCA2 得分差距较小, 累积方差贡献达到 82.3%, 究其原因可能与数据集分布不均有关, 3 个药性分类包含中药数量分别为 22、25、11, 远小于均值 56.857。中药数量不足导致属性分类向量缺少足够的特征信息, 而体现出较高的余弦相似度。将药性数据集按照均值切割, 分别计算 2 个部分的余弦相似度。

表 2 药性关键成分种类最终得分 (中药数量超过中位数)

Table 2 Final score of key ingredient types for medicinal properties (number of traditional Chinese medicines exceeds median)

药性分类	关键成分种类最终得分									
	1	2	3	4	5	6	7	8	9	10
寒	3 007	4 510	3 152	3 965	1 641	3 738	3 758	4 511	3 717	3 752
微寒	4 510	3 007	1 641	3 738	4 502	3 152	1 631	1 749	1 596	3 151
平	3 007	4 510	4 505	3 152	3 965	3 712	4 511	1 641	3 758	3 106
温	3 007	4 510	3 738	3 965	3 152	1 641	3 758	3 150	3 161	4 801

表 3 药味关键成分种类最终得分 (中药数量超过中位数)

Table 3 Final score of key ingredient types for medicinal flavors (number of traditional Chinese medicines exceeds median)

药味分类	关键成分种类最终得分									
	1	2	3	4	5	6	7	8	9	10
咸	4 502	3 007	4 509	3 735	3 965	3 736	3 738	4 506	4 604	4 852
甘	4 510	3 007	3 152	4 512	3 738	1 081	4 505	3 717	3 758	3 965
苦	3 007	4 510	3 152	3 965	3 738	3 758	1 081	1 641	3 150	1 070
辛	4 505	4 654	4 504	4 510	3 738	4 507	3 007	3 766	2 939	3 735

表 4 归经关键成分种类最终得分 (中药数量超过中位数)

Table 4 Final score of key ingredient types for meridian tropism (number of traditional Chinese medicines exceeds median)

归经分类	关键成分种类最终得分									
	1	2	3	4	5	6	7	8	9	10
心	3 007	4 510	3 738	3 152	3 965	3 758	3 712	3 150	3 717	1 070
肝	3 007	4 510	3 152	3 965	3 738	4 511	3 758	3 150	1 631	1 081
肺	3 007	4 510	3 152	3 965	3 724	1 081	3 727	3 161	3 733	3 150
肾	3 007	4 510	3 738	3 717	3 152	3 965	3 766	3 727	3 752	4 539
胃	3 007	4 510	3 152	3 965	3 738	3 758	3 727	3 733	1 070	3 161
脾	3 007	4 510	3 738	3 152	3 965	3 717	1 081	3 758	1 070	4 512

表 5 功效关键成分种类最终得分 (中药数量超过中位数)

Table 5 Final score of key ingredient types for efficacy (number of traditional Chinese medicines exceeds median)

功效分类	关键成分种类									
	1	2	3	4	5	6	7	8	9	10
发散风寒药	4 510	1 641	3 007	1 628	3 688	4 803	1 936	3 161	4 557	1 637
活血行气止痛药	4 510	3 738	3 007	4 797	3 717	1 330	3 962	384	4 610	3 748
活血调经药	3 247	3 007	2 754	4 656	4 510	3 106	4 805	3 152	4 490	4 803
清热解药	4 511	4 510	4 504	4 513	3 738	3 007	1 637	4 801	1 628	1 075
理气药	4 801	1 631	4 510	3 007	1 637	3 152	1 081	4 805	1 638	3 747
祛风湿湿药	3 766	3 738	2 942	963	383	4 801	4 805	4 796	4 539	3 007
补阳药	4 510	4 509	3 007	4 654	1 631	4 604	4 804	4 801	1 638	4 805

大于均值的部分寒、平、温相似度均小于 0，符合中药药性理论；小于均值的部分仅热与凉相似度达到 0.4，其余均小于 0.1。

药味余弦相似度结果 (图 1-B) 显示，咸、淡、涩、微苦 4 个分类相似度较高，苦、辛、甘 3 个分类之间相似度也较高。由于药味分类是复合药味的药物同时属于多个分类，可能造成分类之间重复中药较多，使属性分类向量受到拆分方法的影响。为探究拆分归类对结果的影响，考虑到药味初始分类多达 38 个，且较多分类包含的中药较少，在排除中药数少于 5 个分类后，将原始数据集中单个药味分类及相关的 2 种药味的分类重新构建属性向量并计算余弦相似度。如图 1-C 所示，苦、辛、甘的相似度均较高，而三者与其相关的复合药味相似度较低。对药味数据集进行切割，在切割后的结果中，甘、苦、辛 3 个分类的相似度均小于 0，甘、微苦与甘、苦的相似度较高。计算相同组成的分类的平均相似度，结果基本大于全局平均相似度。

归经余弦相似度结果 (图 1-D) 显示，肝、肺、肾、胃、脾相似度均较高，为排除拆分归类影响，在排除原始数据集中少于 5 个中药的归经分类和 3 个归经的复合分类后，重新计算归经分类向量并计

算余弦相似度。如图 1-E 所示，肺、肝、肺、胃，肺之间相似度较高，心、胃与心、肺之间相似度也较高。计算相同组成的复合归经分类的平均相似度，发现相同组成复合分类的平均相似度基本大于全局平均相似度。

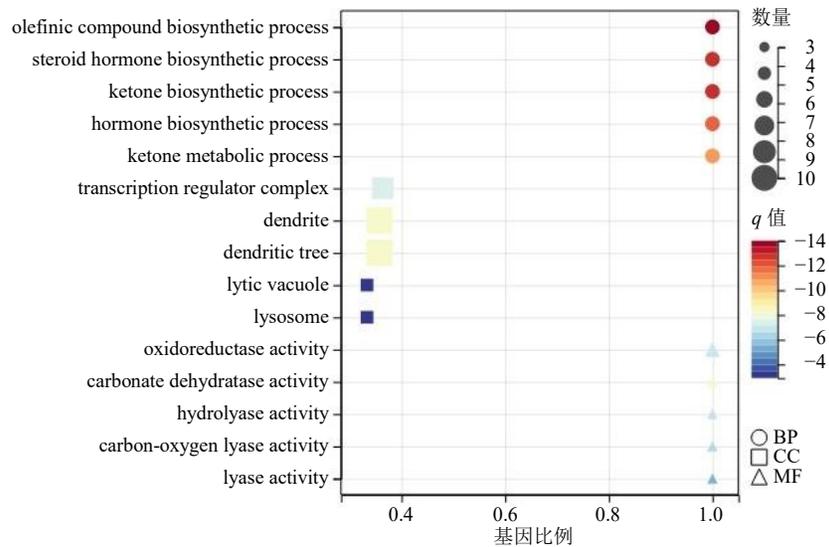
由于功效分类较多，且数据分布不均匀，为了减少影响，在降维前将中药数量小于均值的功效分类剔除，仅分析数量较多的功效分类。如图 1-F 所示，清热泻火药、清热化痰药、祛风湿热药、补阴药同属寒凉相似度较高。然而固精缩尿止带药与利尿通淋药相似度较高，前者为收涩之品与后者作用趋势相反。在降维的结果中，功效分类的特征被降至 12 维，前 4 个维度累积方差解释比例达到 68.67%，而二者前 4 个维度的值相似且为负数，说明 2 个分类均缺少有前 4 个维度所代表化学特征的化学成分。

2.6 活血行气止痛药与活血调经药关键成分富集分析

在 BMCFP 得分矩阵中，活血行气止痛药与活血调经药得分排名前 10 的成分种类共涉及 356 个化学成分，包括甘草皂苷 A3 (licoricesaponine A3)、环氧叶黄素 (lutein epoxide)、维生素 P (vitamin P)、

(broad substrate specificity ATP-binding cassette transporter, ABCG2)、ABL 原癌基因 1 (tyrosine-protein kinase1, ABL1)、ABL2、Bcl-2 样蛋白 1 (Bcl-2-like protein 1, BCL2L1)、大麻素受体 2 (cannabinoid receptor 2, CNR2)、凝血因子 III (tissue factor, F3)、11 β -羟基类固醇脱氢酶 1 型 (11beta-hydroxysteroid dehydrogenase 1, HSD11B1)、白细胞介素-6 (interleukin-6, IL6)、基质金属蛋白酶 9 (matrix metalloproteinase 9, MMP9)、肿瘤坏死因子 (tumor necrosis factor, TNF) 等。将上述靶点导入 MetaScape 平台, 限定物种为智人, 并进行 GO 功能注释与 KEGG 通路富集分析。以校正后 $P < 0.01$ 为阈值筛选关键通路 (图 2、3)。结果表明, 在生

物过程层面, 显著富集的条目包括烯烃类化合物生物合成过程 (olefinic compound biosynthetic process, GO:0120255) 和类固醇激素生物合成过程 (steroid hormone biosynthetic process, GO:0120178) 等; 在细胞组分层面, 主要在转录调控复合物 (transcription regulator complex, GO:0005667) 与树突 (dendrite, GO:0030425) 等显著富集; 在分子功能层面, 主要富集于氧化还原酶活性 (oxidoreductase activity, GO:0016491) 和碳酸酐酶活性 (carbonate dehydratase activity, GO:0004089) 等。KEGG 通路分析进一步显示, 这些活性成分可能通过调控神经活性配体-受体相互作用 (neuroactive ligand-receptor interaction, hsa04080)



BP-生物过程; CC-细胞成分; MF-分子功能。

BP-biological process; CC-cellular component; MF-molecular function.

图 2 GO 富集分析结果。

Fig. 2 Results of GO enrichment analysis

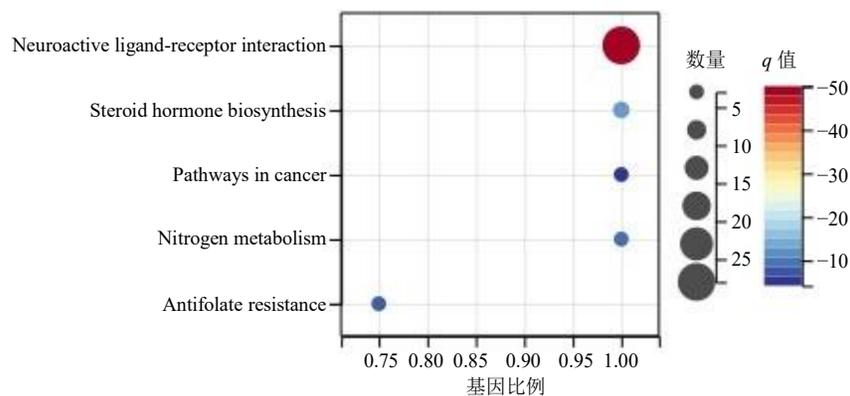


图 3 KEGG 富集分析结果

Fig. 3 Results of KEGG pathway enrichment analysis

和类固醇激素生物合成 (steroid hormone biosynthesis, hsa00140) 等信号通路发挥药理作用。

3 讨论

随着科学技术的发展, 中医药研究已逐步进入微观层面。为了更好地阐述中医药防病治病的原理, 并推动中医药知识的传承与创新, 深入分析中药属性与化学成分间的关联关系具有重要的研究意义。本研究改进了以往仅重视性质而忽略程度的分类方法, 聚焦于常见中药, 通过成分指纹计算属性特征, 获得与属性密切相关的成分种类得分矩阵, 并筛选出对属性分类贡献最大的前 10 个成分种类, 作为属性分类关键成分群。

苦味排名首位的关键成分种类是编号为 3007 的成分种类, 该成分种类包含多种植物甾醇糖苷, 可通过激活人类舌面的味觉受体 2 型成员 1 (taste receptor type 2 member 1, TAS2R) 使人感知到苦味^[17]。寒性属性方面, 编号为 3512 的成分种类多属于黄酮醇多糖苷类化合物, 此类化合物可通过清除自由基, 激活核因子 E2 相关因子 2 (nuclear factor erythroid-2-related factor 2, Nrf2) /抗氧化反应元件 (oxidative response element, ARE) 通路等机制发挥抗氧化、抗炎等作用, 进而实现清热解毒、凉血等寒性功效^[18-22]。关于归经属性, 肺经相关的关键成分种类主要包含双苄基异喹啉类生物碱 (如成分种类编号 1081), 该类中的粉防己碱被证实可用于治疗呼吸系统疾病^[23-25]。对于活血行气止痛药物而言, 其关键成分种类 3738 主要由具有五环三萜骨架在 C-3 位连接复杂的寡糖链的齐墩果烷型和常春藤皂苷元型三萜皂苷组成, 具有抗炎、抗氧化、调节血管内皮功能等活性^[26-27]。特别是续断皂苷 B 在脑卒中^[28]和心脏病^[29-30]的治疗中有一定作用。

属性向量相似性进一步分析显示, 在中药药性分类中, 即使是微程度差异的药性 (如寒、微寒、凉) 也反映出不同的化学特征, 因此不能简单地将它们视为同一属性分类。此外, 研究发现甘、微苦与甘、苦之间存在较高的相似度, 这暗示着苦与微苦在化学层面上可能有相似之处。同时, 尽管相同药味和归经的复合属性分类并未表现出明显的相似性, 但拥有相同组成的分类在化学特征上的平均相似度高于全局平均水平, 表明它们之间存在一定的关联。值得注意的是, 作用趋势相近的功效分类化学相似性更高, 该发现不仅契合中药功效理论的整体认知, 也与“七情”配伍理论中相须、相使的

协同增效机制相呼应^[31-33]。

活血药常用于脑卒中的治疗, 通过 GO 和 KEGG 富集分析, 研究发现活血药关键成分参与了神经活性配体-受体相互作用、类固醇激素生物合成 (steroid hormone biosynthesis)、癌症通路和溶酶体相关条目等多种生物学过程或通路。关键成分作用的靶点中, 细胞色素 P450 11B2 (cytochrome P450 11B2, CYP11B2)、CYP17A1、CYP19A1、17 β -羟基类固醇脱氢酶 3 型 (17 β -hydroxysteroid dehydrogenase type 3, HSD17B3)、5 α -还原酶 2 型 (3-oxo-5 α -steroid 4-dehydrogenase 2, SRD5A2) 在类固醇激素生物合成通路中发挥重要作用, 活血药可通过影响 CYP11、CYP2C9 等靶点降低类固醇激素的合成而减少血管收缩和炎症反应以改善血液循环影响脑卒中恢复^[34-35]。半胱氨酸天冬氨酸蛋白酶-3 (cysteine aspartate protease 3, CASP3) 作为凋亡家族重要的信号分子, 在治疗脑卒中方面也发挥了一定的作用^[36]。另外, 溶酶体相关条目 (lysosome 和 lytic vacuole) 等条目可参与调节溶酶体膜稳定性, 促进凋亡细胞清除。这些结果不仅揭示了活血化瘀药物通过多靶点、多通路调控血液微循环和抗凝血功能, 还提示了其潜在的抗肿瘤和改善代谢紊乱的效果。

然而, 研究仍存在一定的局限性, 包括数据选择偏倚和成分数据误差等问题。为更全面地揭示中药成分与属性的关系, 未来的工作需要构建更加完善的化学特征评价体系和成分数据库, 并通过实验验证以提高结果的可靠性。

4 结论

本研究基于分子指纹构建中药属性分类向量, 并引入 BMCFP 评分体系, 系统探讨了中药属性与其化学成分之间的关联。此后, 通过计算化学特征得分矩阵, 结合成分间的化学相似性进行层次聚类, 识别出各属性类别中的关键成分群。基于属性分类向量的余弦相似度的进一步分析发现, 不同程度的药性分类在化学特征上存在差异; 同时, 药味、归经的复合分类与表现出类似的作用趋势的功效分类具有一定的相似性, 提示其潜在的作用相似性。此外, 对活血行气止痛药和活血调经药的关键成分进行 GO 和 KEGG 富集分析, 揭示了这些成分发挥作用的相关通路, 为深入理解中药属性的分子机制奠定了可能的基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Li X, Li N, Sui Z Y, *et al.* An investigation on the quantitative structure-activity relationships of the anti-inflammatory activity of diterpenoid alkaloids [J]. *Molecules*, 2017, 22(3): 363.
- [2] 李石生, 邓京振, 赵守训, 等. 中药现代化研究的关键在于建立科学的现代中药理论体系——分子药性假说的提出 [J]. 中国中西医结合杂志, 2000, 20(2): 4-5.
- [3] 付先军, 王振国, 李学博, 等. 基于化学结构相似性的植物类中药“性-构”关系分析 [J]. 中华中医药杂志, 2019, 34(6): 2657-2661.
- [4] 欧阳兵, 王振国, 李峰, 等. 中药四性“性-效-物质三元论”假说及其论证 [J]. 山东中医药大学学报, 2008(3): 182-183.
- [5] 谢芬芬. 基于数据挖掘对不同药性中药的化合物分子特征研究 [D]. 西宁: 青海大学, 2021.
- [6] 林林. 基于 SMILES 结构的中药寒热药性随机森林算法建模及验证 [D]. 北京: 中国中医科学院, 2022.
- [7] 赵书言, 王书睿, 申镇华, 等. 基于人工智能算法的中药主归经判别研究 [J]. 中草药, 2024, 55(16): 5573-5582.
- [8] 贾荣浩, 魏国辉, 赵文华, 等. 基于 K-近邻算法的中药化合物寒热平药性预测研究 [J]. 中华中医药杂志, 2023, 38(4): 1522-1525.
- [9] 张冰冰, 朱琦, 张晶新, 等. 基于深度学习的中药寒热属性预测研究 [J]. 世界中医药, 2023, 18(21): 3047-3052.
- [10] Yang M L, Liu W. Classification of cold and hot medicinal properties of Chinese herbal medicines based on graph convolutional network [J]. *Digital Chin Medicine*, 2024, 7(4): 356-364.
- [11] Fang S S, Dong L, Liu L, *et al.* HERB: A high-throughput experiment- and reference-guided database of traditional Chinese medicine [J]. *Nucleic Acids Res*, 2021, 49(D1): D1197-D1206.
- [12] Ye H, Ye L, Kang H, *et al.* HIT: Linking herbal active ingredients to targets [J]. *Nucleic Acids Res*, 2011, 39(Database issue): D1055-D1059.
- [13] Yan D Y, Zheng G H, Wang C C, *et al.* HIT 2.0: An enhanced platform for herbal ingredients' targets [J]. *Nucleic Acids Res*, 2022, 50(D1): D1238-D1243.
- [14] Xu H Y, Zhang Y Q, Liu Z M, *et al.* ETCM: An encyclopaedia of traditional Chinese medicine [J]. *Nucleic Acids Res*, 2019, 47(D1): D976-D982.
- [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] Shen W T, Song Z G, Zhong X, *et al.* Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform [J]. *Imeta*, 2022, 1(3): e36.
- [17] Meyerhof W, Batram C, Kuhn C, *et al.* The molecular receptive ranges of human TAS2R bitter taste receptors [J]. *Chem Senses*, 2010, 35(2): 157-170.
- [18] 谢孝平, 沈小艳, 李博文, 等. 芦丁通过激活 NRF2/HO-1 通路抑制 Ang II 诱导的血管平滑肌细胞表型转化和炎症反应 [J]. 现代生物医学进展, 2023, 23(22): 4201-4206.
- [19] Guardia T, Rotelli A E, Juarez A O, *et al.* Anti-inflammatory properties of plant flavonoids. Effects of rutin, quercetin and hesperidin on adjuvant arthritis in rat [J]. *Farmaco*, 2001, 56(9): 683-687.
- [20] Procházková D, Boušová I, Wilhelmová N. Antioxidant and prooxidant properties of flavonoids [J]. *Fitoterapia*, 2011, 82(4): 513-523.
- [21] 蒋雨心, 邓岚, 范方宇. 鱼腥草根总黄酮超声辅助酶法提取工艺优化及其抗氧化活性研究 [J]. 食品工业科技, 2023, 44(6): 226-234.
- [22] 王兰, 曾禧, 姚后宗, 等. 雪层杜鹃中的黄酮类化合物及其抗氧化和抗炎活性研究 [J]. 天然产物研究与开发, 2024, 36(8): 1350-1356.
- [23] 王翼遥, 宋玲, 高云航, 等. 雾化吸入粉防己碱对博来霉素致肺纤维化大鼠的药效作用 [J]. 药物评价研究, 2023, 46(12): 2592-2600.
- [24] 谢林艳, 陈子, 闫兆, 等. 粉防己碱抑制支气管哮喘小鼠肺组织核因子- κ B 和诱导型一氧化氮合酶的研究 [J]. 国际呼吸杂志, 2015, 35(9): 641-646.
- [25] Song M Y, Wang J X, Sun Y L, *et al.* Tetrandrine alleviates silicosis by inhibiting canonical and non-canonical NLRP3 inflammasome activation in lung macrophages [J]. *Acta Pharmacol Sin*, 2022, 43(5): 1274-1284.
- [26] 白荣钰, 易欢, 邱静文, 等. 毛冬青三萜皂苷对动脉粥样硬化大鼠粪便和尿液代谢组学的影响 [J]. 南京中医药大学学报, 2022, 38(5): 424-431.
- [27] 贺海波, 秦文, 曾建红, 等. 预知子三萜皂苷体外抗脂质过氧化和红细胞溶血作用的研究 [J]. 湖北农业科学, 2016, 55(7): 1758-1761.
- [28] 任凯迪, 彭紫梅, 罗秀菊, 等. 川续断皂苷 B 抑制线粒体 E3 泛素连接酶 1 减轻缺血性中风大鼠脑损伤 [J]. 中国药理学与毒理学杂志, 2021, 35(10): 795.
- [29] Quan W J, Sun T L, Hu B, *et al.* Dipsacoside B attenuates atherosclerosis by promoting autophagy to inhibit macrophage lipid accumulation [J]. *Biomolecules*, 2024, 14(10): 1226.

- [30] 周淼. 山银花源性川续断皂苷乙缓解力竭运动诱导的心肌损伤机理研究 [J]. 分子植物育种, 2025, 23(14): 4833-4841.
- [31] Cui X, Qian D W, Jiang S, *et al.* Scutellariae Radix and coptidis rhizoma improve glucose and lipid metabolism in T2DM rats *via* regulation of the metabolic profiling and MAPK/PI3K/Akt signaling pathway [J]. *Int J Mol Sci*, 2018, 19(11): 3634.
- [32] 白志尧, 杨雁芸, 周宁, 等. 白术配伍桑白皮对肾病综合征大鼠尿液代谢组学的影响 [J]. 中药材, 2018, 41(7): 1707-1712.
- [33] 李杰, 胡超, 许龙龙, 等. 参麦方对于大鼠心肌损伤保护作用代谢组学研究 [J]. 中国药理学通报, 2016, 32(11): 1559-1565.
- [34] 王玥, 周玥, 赵丽. 缺血性脑卒中关键基因的筛选及治疗药物的预测 [J]. 科技导报, 2023, 41(9): 89-97.
- [35] Duckles S P, Krause D N. Mechanisms of cerebrovascular protection: Oestrogen, inflammation and mitochondria [J]. *Acta Physiol*, 2011, 203(1): 149-154.
- [36] Dehqanizadeh B, Mohammadi Z F, Kalani A H T, *et al.* Effect of early exercise on inflammatory parameters and apoptosis in CA1 area of the hippocampus following cerebral ischemia-reperfusion in rats [J]. *Brain Res Bull*, 2022, 182: 102-110.

[责任编辑 潘明佳]