

• 数据挖掘与循证医学 •

基于分子结构跨域泛化的中药成分相互作用预测研究

程 宁¹, 曾湘祥², 宋学坤^{1*}

1. 河南中医药大学信息技术学院, 河南 郑州 450046

2. 湖南大学计算机学院, 湖南 长沙 410082

摘要: **目的** 针对中药复方成分复杂、相互作用关系不明导致配伍机制难以解析的问题, 以成分间相互作用预测为切入点, 构建基于深度学习的分子相互作用预测模型, 为识别中药成分间潜在相互作用关系及后续方剂配伍机制研究提供方法学参考。 **方法** 提出基于分子 3D 结构特征与扩展连通性指纹的分子相互作用预测模型 (3D structure and extended connectivity fingerprints molecule-molecule interaction, 3DFPMMI)。模型利用 SchNet 网络提取分子 3D 图结构的深层特征, 同步获取扩展连通性指纹, 以对比学习策略实现多源特征的有效融合; 模型在 DrugBank 药物相互作用数据集上完成训练, 并通过跨域泛化实现中药成分间相互作用的预测。 **结果** 3DFPMMI 模型在 DrugBank 数据集上准确率达到 0.921, 精确率-召回率曲线下面积为 0.972, 均优于其他基准模型。消融实验证实, 分子 3D 结构特征、扩展连通性指纹特征及对比学习特征融合策略均为模型性能提升的核心有效模块。将该模型应用于经典方剂补阳还五汤和半夏泻心汤的成分相互作用预测, 其部分预测结果与已有的实验研究结论高度吻合。 **结论** 构建了一种中药成分相互作用预测的新方法, 并初步验证了其应用潜力, 为从分子层面解读中药方剂配伍机制提供了新的技术手段与方法学借鉴。

关键词: 中药成分; 分子相互作用; 深度学习; 跨域泛化; 分子结构; 配伍机制

中图分类号: TP18; R284 **文献标志码:** A **文章编号:** 0253-2670(2026)17-6867-11

DOI: 10.7501/j.issn.0253-2670.2026.17.018

Prediction of traditional Chinese medicine component interactions via cross-domain generalization of molecular structure

CHENG Ning¹, ZENG Xiangxiang², SONG Xuekun¹

1. School of Information Technology, Henan University of Chinese Medicine, Zhengzhou 450046, China

2. College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

Abstract: Objective Given the complexity of traditional Chinese medicine (TCM) formula constituents and their unclear interactions, a deep learning-based molecular interaction prediction model is constructed taking the prediction of interactions among constituents as the entry point, providing methodological support for identifying potential constituent interactions and subsequent studies on formula compatibility mechanisms. **Methods** We propose a molecule-molecule interaction prediction model based on molecular 3D structural features and Extended Connectivity Fingerprints, designated as 3D structure and extended connectivity fingerprints molecule-molecule interaction (3DFPMMI). The model employs the SchNet network to extract deep features from the 3D graph structure of molecules while simultaneously acquiring Extended Connectivity Fingerprints, and achieves effective fusion of multi-source features through a contrastive learning strategy. The model is trained on the DrugBank drug-drug interaction dataset, and cross-domain generalization is implemented to realize the prediction of interactions among TCM constituents. **Results** The 3DFPMMI model achieves an accuracy of 0.921 and an area under the precision-recall curve of 0.972 on the DrugBank dataset, both outperforming other benchmark models. Ablation studies confirm that molecular 3D structural features, extended connectivity fingerprint features, and the contrastive learning-based feature fusion strategy are all core effective modules for improving model performance. When applied to predict component interactions in the classic TCM formulas Buyang Huanwu Decoction and Banxia Xiexin Decoction, some of the prediction results are highly consistent with the conclusions from existing experimental studies. **Conclusion** This study establishes a novel method for predicting interactions among

收稿日期: 2026-03-25

基金项目: 国家中医药管理局中医药创新团队及人才支持计划项目 (ZYXCXTD-D-202408); 国家自然科学基金项目 (61702164); 河南省卫生健康委国家中医临床研究基地科研专项课题 (2021JDZX2135)

作者简介: 程 宁, 讲师, 主要从事中医药人工智能研究。E-mail: chengning@hactcm.edu.cn

*通信作者: 宋学坤, 教授, 硕士生导师, 主要从事人工智能、计算生物医学信息学研究。E-mail: sxk@hactcm.edu.cn

TCM ingredients and preliminarily validates its application potential. This work offers new technical tools and methodological references for interpreting the compatibility mechanisms of TCM formulas at the molecular level.

Key words: traditional Chinese medicine components; molecular interaction; deep learning; cross-domain generalization; molecular structure; compatibility mechanism

中药复方以多成分、多靶点、多通路的协同作用为核心特征，在多种疾病的临床治疗中展现出不可替代的独特优势^[1]。复方药效的发挥并非依赖单一成分的独立作用，而是源于多种化学成分的协同起效。然而，中药复方配伍体系极具复杂性，其成分间的相互作用关系尚不明确，整体药效的潜在作用机制尚未得到系统全面的阐释。中药成分间的相互作用是解析复方配伍规律的核心切入点，不仅决定复方的临床药效，更与用药安全性密切相关，因此成分-成分相互作用预测是理解中药复方复杂配伍关系的重要基础环节，可为后续构建多组分相互作用网络、筛选关键成分组合及解析整体调控机制提供候选线索，具有重要的理论与实践意义。

传统网络药理学研究多遵循“成分-靶点-通路-疾病”的固定单向分析范式^[2]，核心研究路径为鉴定中药复方中的活性成分，筛选成分对应的潜在作用靶点，富集靶点参与的关键信号通路，最终将药理效应与目标疾病相关联，以此阐释中药治疗疾病的内在作用机制^[3]。然而，该经典研究范式存在显著的局限性：其仅聚焦于单一成分的独立药理效应，逐一解析单个成分与靶点、通路、疾病间的线性关联，却完全忽视了复方中成分与成分间潜在的交互作用^[4]。中药复方是多成分构成的复杂动态体系，成分间既可能通过协同作用实现药效增强，也可能因拮抗作用导致药效衰减，甚至引发毒副反应^[4]。这种成分间的交互作用是决定复方整体药效与作用模式的核心要素，而传统网络药理学对成分交互效应的忽视导致其对中药作用机制的阐释存在片面性与偏差，难以完整揭示中药复方治疗疾病的深层科学内涵与复杂作用规律。

尽管分子间相互作用的机制预测已成为生物医药领域的研究热点^[5-7]，但现有相关研究仍主要聚焦于化学药物^[8-9]，针对中药成分间相互作用的系统性研究仍显不足，尤其缺乏高质量、标准化的中药成分-成分相互作用标准数据集，导致数据驱动的深度机器学习模型难以基于有限的中药数据完成端到端训练与泛化性能优化。针对这一问题，考虑到无论化学药还是中药成分，其分子间是否发生相互作

用、以何种方式作用，本质上均由分子自身的结构特征决定^[10]，基于分子结构信息的跨域迁移在化学原理层面具备理论可行性，因此本研究引入跨域泛化的研究范式^[11]，使用化学药物数据训练分子相互作用预测模型，并将其应用于中药成分相互作用预测中。当前研究的技术难点主要在于现有基于化学药物数据训练的模型难以直接泛化至中医药领域，一些化学药物分子特征提取是基于药物-药物相互作用（drug drug interactions, DDI）网络、药物-靶点相互作用（drug target interactions, DTI）网络等关系网络，与中医研究常用的“方-药-成分-靶点”异构图在语义空间上存在差异，化学药物特征缺乏中药层级节点及其与方剂的配伍关系，导致特征空间分布不一致，因此必须采用仅依赖分子自身固有结构的特征提取方法。分子间相互作用的预测性能取决于分子表征的质量，高精度的分子表征能够完整刻画分子的结构特性与理化属性，因此，如何从分子固有结构中挖掘高维度、强鲁棒性的高阶非线性特征是突破成分间相互作用预测性能瓶颈的核心关键^[12]。分子的 3D 结构是其生物学功能的基础，能够直观反映分子间的相互作用机制。扩展连通性指纹特征（extended connectivity fingerprint, ECFP）作为一种基于分子拓扑结构的特征表示方法，能够量化分子的化学性质和结构信息。通过结合 3D 结构和 ECFP 信息，可以提取更高阶的分子特征。此外，近年来深度学习技术在生物医学领域取得了突破性进展^[13]，其强大的高维特征自动提取与非线性模式识别能力能够从复杂的分子结构数据中挖掘核心关联规律^[14]。

因此，本研究提出一种基于分子结构的成分-成分相互作用深度学习预测模型，该模型通过融合分子的 3D 空间结构特征与 ECFP 特征，实现了分子间相互作用预测；同时基于跨域泛化思路，探索将化学药物数据训练得到的预测模型迁移应用于中药成分间相互作用预测场景的可行性。本研究旨在从成分间相互作用层面为中药方剂配伍关系研究提供候选线索和方法学参考，并为后续构建多组分相互作用网络及解析复方整体调控机制奠定基础。

1 方法

1.1 数据来源

1.1.1 药物相互作用数据 从 DrugBank 数据库^[15]中获取化学药物分子相互作用数据作为模型训练的数据源,通过该数据集完成分子相互作用预测模型的训练后,进一步采用跨域泛化的方式,将训练好的模型应用于中药成分间相互作用的预测研究。

具体而言,从 DrugBank 数据库获取包含 1 707 种化学药物的 191 807 条 DDI 记录的标准数据集,用于后续分子相互作用预测模型的训练与参数优化。该数据集包含 86 种详细的 DDI 类型,如“增加新陈代谢”“提高血药浓度”等药物代谢结果,如表 1 所示。每条 DDI 数据包含药物 1、药物 2 及其具体相互作用描述,如表 2 所示。

表 1 DrugBank 数据库 DDI 类别示例

Table 1 Examples of DDI categories in DrugBank database

DDI 类别	描述
type 10	药物 2 与药物 1 合用可提高血清浓度
type 13	药物 2 与药物 1 合用会降低治疗效果
type 15	药物 1 可能会降低药物 2 的代谢率,从而导致血清水平升高
type 17	药物 1 可能会降低药物 2 的心脏毒性活性
type 26	药物 1 与药物 2 联合使用不良反应的风险或严重程度可能会增加
type 46	药物 1 可能增加药物 2 的抗凝血活性
type 63	药物 1 可能增加药物 2 的低钙活性
type 64	药物 1 可能增加药物 2 的降血糖活性

表 2 DrugBank 数据库药物相互作用示例

Table 2 Examples of drug-drug interactions in DrugBank database

药名 1	药名 2	相互作用描述
环磷酰胺	利血平	两者联合应用可增加新陈代谢
青蒿素	阿巴美吡	两者联合应用可提高血药浓度
麻黄素	醋丁洛尔	两者联合应用可提高降血压治疗效果
氟康唑	阿西美辛	两者合用高钾血症的副作用可能会增加
布洛芬	阿巴卡韦	布洛芬可能会降低阿巴卡韦的代谢率,从而导致血清水平升高

1.1.2 中药数据 选取补阳还五汤、半夏泻心汤 2 首经典名方作为研究对象,方剂均出自中医经典医籍^[16-17],药材明确、配伍规范。2 首方剂的组成见表 3。

表 3 方剂及其包含中药

Table 3 Herbal formulas and their contained traditional Chinese medicines

方剂	包含中药
补阳还五汤	黄芪、当归尾、赤芍、地龙、川芎、红花、桃仁
半夏泻心汤	半夏、黄连、黄芩、干姜、甘草、大枣、人参

由于当前国内外生物医药领域主流专业数据库均未建立针对中药多成分体系的成分-成分相互作用规范化标注数据,因此本研究基于上述方剂的标准组方,从《中国药典》2025 年版^[18]及其相关文献报道^[19-20]中完成方剂对应各单味药核心药效物质,即有效成分的系统性筛选与规范化收录。在此基础上,依托本草组鉴 (HERB) 数据库^[21],获取有效成分的简化分子线性输入规范 (simplified molecular input line entry system, SMILES) 序列,将其泛化应用于利用化学药物分子训练好的相互作用模型,进而实现中药成分分子间相互作用的预测。其中,中药及其对应主要有效成分如表 4 所示,部分成分及其 SMILES 序列如表 5 所示。

表 4 中药及其主要包含成分

Table 4 Traditional Chinese medicines and their main ingredients

中药	成分
黄芪	黄芪甲苷 (astragaloside IV) 毛蕊异黄酮葡萄糖苷 (calycosin-7-O-beta-D-glucopyranoside) 芒柄花素 (formononetin) 毛蕊异黄酮 (calycosin)
红花	羟基红花黄色素 A (safflomin A) 山柰酚 (kaempferol)
当归	阿魏酸 (ferulic acid) 藁本内酯 (ligustilide)
桃仁	苦杏仁苷 (amygdalin)
川芎	川芎嗪 (tetramethylpyrazine)
赤芍	芍药苷 (paeoniflorin)
地龙	次黄嘌呤 (hypoxanthine)
半夏	麻黄碱 (ephedrine) 葫芦巴碱 (gynesine)
黄连	小檗碱 (berberine) 黄连碱 (coptisine)
黄芩	黄芩苷 (baicalin)
干姜	6-姜酚 (6-gingerol)
甘草	甘草苷 (liquiritin) 甘草酸 (glycyrrhizic acid)
大枣	齐墩果酸 (oleanolic acid)
人参	人参皂苷 Re (ginsenoside Re) 人参皂苷 Rg ₁ (ginsenoside Rg ₁) 人参皂苷 Rb ₁ (ginsenoside Rb ₁)

表 5 成分 SMILES 序列数据

Table 5 SMILES sequence of ingredients

成分名	SMILES 序列
黄芪甲苷	CC1(C(CCC23C1C(CC4C2(C3)CCC5(C4(CC(C5C6(CCC(O6)C(C)(C)O)C)O)C)OC7C(C(C(C(O7)CO)O)O)OC8C(C(C(CO8)O)O)O)C
山柰酚	C1=CC(=CC=C1C2=C(C(=O)C3=C(C=C(C(=C3O2)O)O)O)O
阿魏酸	COC1=C(C=CC(=C1)C=CC(=O)O)O
麻黄碱	CC(C(C1=CC=CC=C1)O)NC
黄连碱	C1C[N+](C1)=C(C=C3C=CC4=C(C3=C2)OCO4)C5=CC6=C(C=C51)OCO6
黄芩苷	C1=CC=C(C=C1)C2=CC(=O)C3=C(C(=C(C(=C3O2)OC4C(C(C(C(O4)C(=O)O)O)O)O)O)O
齐墩果酸	CC1(CCC2(CCC3(C(=CCC4C3(CCC5(C4(C(C(C5(C)C)O)C)C)C21)C)C(=O)O)C
人参皂苷 Re	CC1C(C(C(C(O1)OC2C(C(C(OC2OC3CC4(C(CC(C5C4(CCC5C(C)(CCC=C(C)C)OC6C(C(C(C(O6)CO)O)O)O)C)O)C7(C3C(C(C(C7)O)C)C)C)CO)O)O)O)O

1.2 3DFPMI 模型构建

模型整体架构如图 1 所示。针对每个目标分子，模型分别利用 SchNet 和 ECFP 提取 3D 结构与拓扑特征，通过特征融合与对比学习实现多模态对齐，随后将双分子特征拼接，通过多层感知器输出最终预测结果。

1.2.1 分子 3D 结构特征提取模块 为获取分子 3D 结构特征表示，将分子 SMILES 序列特征转换为 3D 图结构。分子 3D 图结构由 4 元组 $G=(U, V, E, P)$ 构成，其中 U 代表全局特征向量， V 代表原子节点特征向量， E 代表边（即化学键）集合， P 代表每个原子节点的坐标信息。 $E_i=(e, s, r) \in E$ ， e 代表边的特征， s 和 r 代表发送方和接收方节点的索引，其传递更新如公式（1~3）所示。

$$e'_k = \phi^e(e_k, v_{r_k}, v_{s_k}, E_{s_k}, u, \rho^{p \rightarrow e}(\{r_h\}_{h=1}^n)) \quad (1)$$

$$v'_i = \phi^v(v_i, \rho^{e \rightarrow v}(E_i), \rho^{p \rightarrow v}(\{r_h|h \in \{i\} \cup N_i\})) \quad (2)$$

$$u' = \phi^u(\rho^{e \rightarrow u}(E'), \rho^{v \rightarrow u}(V'), \rho^{p \rightarrow u}(\{r_h\}_{h=1}^n)) \quad (3)$$

其中， ϕ 表示同一类特征的更新函数， ρ 表示从一类特征到另一类特征的传递函数。不同分子图的特征传递类型可能存在差异，如边特征 e'_k 可以通过 $\rho^{p \rightarrow e}$ 接收旧的边特征、边两端的节点特征 v_{r_k} 和 v_{s_k} 、指向 s_k 的边集合的特征 E_{s_k} 、全局特征 u 以及 $r_k \cup s_k \cup N_{s_k}$ 集合特征。节点特征 v'_i 和全局特征 u' 的传递同上。

SchNet 等^[22]构造连续 n 个原子节点特征的过程如公式（4）所示，其中，原子 X^l 由一组特

征 $(x_1^l, \dots, x_n^l)$ 来描述，它们的位置信息为 $R=(r_1, \dots, r_n)$ ， $r_i \in R^D$ 。SchNet 节点特征更新函数如公式（5~7）所示，边特征 e_k 通过转换 3D 特征中的 P 更新，因此对于 3 个位置聚合函数 SchNet 仅需 $\rho^{p \rightarrow e}$ 对边进行更新。

$$X_i^{l+1} = (X^l * W^l)_i = \sum_j x_j^l \circ W^l(r_i - r_j), W^l: R^D \rightarrow R^F \quad (4)$$

$$e'_k = \phi^e(v_{r_k}, \rho^{p \rightarrow e}(\{r_h|h=r_k \cup s_k\})) \quad (5)$$

$$v'_i = \phi^v(v_i, \rho^{e \rightarrow v}(E_i)) \quad (6)$$

$$u' = \phi^u(u, \rho^{v \rightarrow u}(V')) \quad (7)$$

对于得到的分子 3D 特征，使用 Python 中的 PyG 和 DIG 存储库进行重建，通过对分子 D 的三维重建，得到对应的 $pos \in R^{N1,3}$ 和节点类别号 Z 。将这些特征输入到 SchNet 中得到位置张量的径向基函数坐标，最终的特征 F 如公式（8）所示。

$$F(pos, Z) = \text{SchNet}(D) \quad (8)$$

1.2.2 基于对比学习的分子特征融合模块 根据 3D 分子结构特征及 ECFP 特征搭建多模态对比学习模型，以获得分子的多模态特征。具体而言，分别将分子 3D 结构特征与 ECFP 特征表示映射到同一空间上。其中，分子 3D 结构特征通过“1.2.1”项的特征提取模块获取，ECFP 特征通过 python 的 RDKit 工具提取。特征编码器经过卷积神经网络编码后得到特征表示 h_i ，再通过由 2 层多层感知器^[23]（multilayer perceptron, MLP）再次非线性变换，最后通过 g 函数将特征 h 进行非线性变换为 z ，再将特征映射到相应的表示空间中。

在对比学习过程中，正例与负例的定义如下：对于任意分子，其正例对由该分子自身的多模态特征视图构成，即同一分子的 3D 结构特征与其 ECFP 特征互为正例。其他不同分子相关的特征均作为当前分子的负样本，通过最大化不同分子表征间在嵌入空间中的距离，增强模型对分子个体差异的区分能力。基于上述定义，模型采用的损失函数如公式（9）所示，其中 z_i 表示正例在表示空间的特征向量。通过设定损失函数指引模型的训练方向，使得模型正例间距离更近，负例间距离更远，从而将同一分子的 3D 结构信息与拓扑指纹信息在语义空间中拉近对齐，实现多模态特征的有效融合。

$$\text{Loss}_i = -\log \left(\frac{\exp(S(z_i, z_i^+)/\tau)}{\sum_{j=0}^K \exp(S(z_i, z_j)/\tau)} \right) \quad (9)$$

1.2.3 分子相互作用关系预测模块 通过 DrugBank 数据库的 DDI 数据，基于分子 3D 特

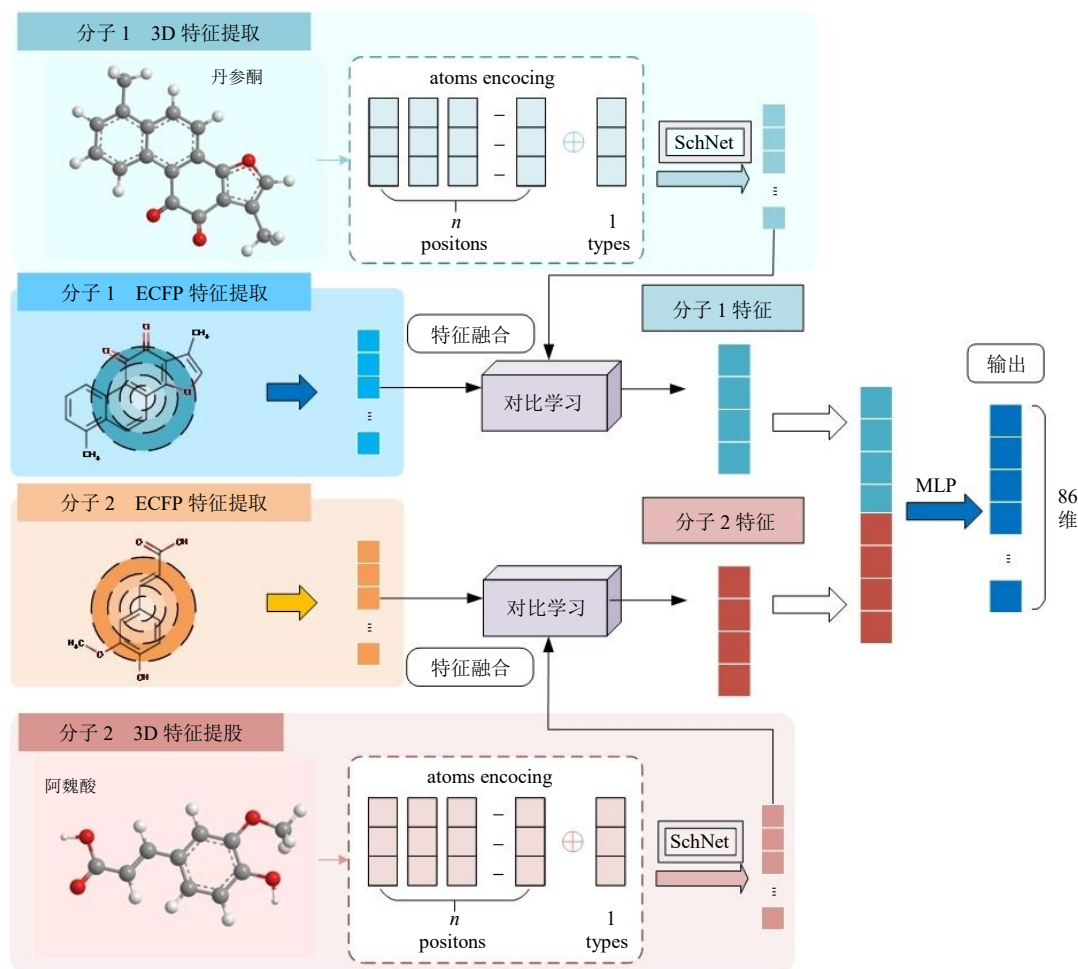


图 1 3DFPMI 分子相互作用预测模型

Fig. 1 Architecture of 3DFPMI molecular interaction prediction model

征与 ECFP 的对比学习获取药物分子特征, 将需要预测的 2 个药物特征拼接得到特征 R , 并对特征 R 进行归一化处理, 输入 MLP 使用 softmax 预测 DDI 关系, 如公式 (10) 所示。

$$y' = \text{softmax}(\text{linear}(\text{layerNorm}(R))) \quad (10)$$

对于中药成分, 根据公式 (11) 定义来描述中药 $Herb$ 中包含的成分 $ingred$, 每个成分 $ingred_i$ 是 $Herb$ 集合中的 1 个元素。对于每味中药, 均构建 1 个包含其所有有效成分的集合。在此基础上, 针对方剂内不同单味中药所属的有效成分集合, 通过集合间的笛卡尔积运算完成全组合两两成分对的遍历生成, 将其输入至上述泛化应用后的模型中, 构建方剂全成分组合的成分-成分对分析数据集, 随后将该数据集中每一组成分对应的标准化分子结构表征信息依次输入分子相互作用预测模型, 进行中药成分间相互作用预测。

$$Herb = \{ingred_1, ingred_2, \dots, ingred_n\} \quad (11)$$

1.2.4 模型性能评估 采用分类任务中常用的一系列评价指标开展模型性能评估工作, 包括精度 (accuracy, Acc)、ROC 特征曲线下面积 (area under curve, AUC)、精确率-召回率曲线下面积 (area under the precision-recall curve, AUPR)、精确率、召回率以及精确率与召回率调和平均数 (F1 分数) [24], 以全面、客观地评价模型预测性能。

目前, 药物相互作用研究在化学药物领域已取得较为丰富的成果, 但由于中药与化学药理论体系存在本质差异, 部分复杂的化学药物相互作用模型无法直接泛化应用于中医药领域, 如药物-药物相互作用多模态深度学习模型 (drug-drug interaction multimodal deep learning, DDIMDL) [25] 需要药物亚结构、靶点、酶等多种模态的数据来进行训练, 而中医药领域目前缺少相关数据。基于 transformer 的自适应共注意力药物-药物相互作用模型 (transformer-based adaptive co-attention

drug-drug interaction, Taco-DDI)^[26]和基于多核图卷积网络的药物相互作用预测模型 (graph convolutional network with multi-kernel, GCNMK)^[27]等基于图神经网络的模型训练依赖于先根据已有 DDI 构建图数据,而中医药领域目前尚无中药成分相互作用相关数据,导致该类模型训练完成后无法通过泛化应用的方式适配中医药领域的研究需求。由于中药成分和化学药物分子在化学结构类似,都可以通过 SMILES 序列来表示,因此,仅基于分子序列结构构建的药物相互作用模型具备向中医药领域泛化应用的可行性。综上,本研究筛选仅基于分子序列开展药物相互作用预测的 DeepDDI^[28]、DeepDrug^[29]和 MHAFFR-DDI^[30]模型作为对比模型。

DeepDDI 模型以药物的分子序列作为输入数据。该模型借助结构相似谱来计算分子向量特征,采用主成分分析技术对所计算的特征进行降维处理,从而得到最终用于预测分子相互作用的特征表示。DeepDrug 模型采用循环神经网络及其变种长短期记忆网络^[31]来处理药物的序列信息。将学习得到的特征输入全连接层,对特征进行进一步的非线性变换和组合,最终实现分子相互作用预测。MHAFFR-DDI 模型采用编码器和图神经网络 (graph neural network, GNN) 架构处理药物 SMILE 序列,将 SMILES 输入编码器提取序列特征,同时将 SMILES 解析为分子图输入 GNN 提取结构特征。两路特征融合后经关系图神经网络和全连接层进行非线性变换,实现药物-药物相互作用预测。

1.3 超参数调整

模型超参数的设置在很大程度上决定了模型的最终性能表现^[32]。分子相互作用预测实验中,为使模型达到最优性能,本研究对模型超参数进行了系统调试与优化,重点针对 ECFP 特征维度、分子 3D 特征维度、模型隐藏层神经元个数及学习率这 4 个核心参数分别分析了不同参数取值对模型性能的影响程度,并结合实验需求设定了各参数的取值范围及调整步长。具体参数设置如下: ECFP 特征维度与分子 3D 特征维度的取值范围均设定为 [128, 512],其中 ECFP 特征维度的调整步长为 128,分子 3D 特征维度的调整步长为 64;模型隐藏层神经元个数的取值范围设定为 [128, 320],调整步长为 64;学习率的取值范围设定为 [0.000 1, 0.012 5],采用 5 倍递增的方式进行梯度调整。

1.4 消融实验

为验证 3DFPMMI 模型中不同特征提取方法及特征融合方法的有效性,本研究在实验中针对特征提取环节设计并开展了消融实验,通过构建模型变体、对比各变体性能差异,明确各核心模块的作用价值。具体消融实验设置如下。

1.4.1 3DFPMMI-T 去除模型中的分子 3D 结构特征提取模块。由于该变体仅保留 ECFP 分子特征,无多类型特征需要融合,因此同步剔除模型中原有的基于对比学习的特征融合方法,确保模型结构与特征输入相匹配。

1.4.2 3DFPMMI-E 去除模型中的分子 ECFP 特征提取模块。与 3DFPMMI-T 变体一致,因仅保留分子 3D 结构特征,无需进行多特征融合,故同步剔除基于对比学习的特征融合方法。

1.4.3 3DFPMMI-C 保留分子 3D 结构特征提取模块与 ECFP 特征提取模块,仅去除基于对比学习的特征融合方法,将 2 种特征直接进行拼接后输入 MLP,完成分子相互作用预测任务。

1.5 成分相互作用预测

在完成 3DFPMMI 分子相互作用预测模型的训练任务后,将该模型跨域泛化应用于中药成分-成分相互作用的预测研究。选取补阳还五汤、半夏泻心汤 2 首经典名方作为具体应用案例,将 2 种方剂中不同中药所含的两两不同有效成分进行组合,构建形成待预测成分组,将其作为 3DFPMMI 模型的输入数据,开展相互作用预测。目前,补阳还五汤与半夏泻心汤中已明确的有效成分见表 4。模型在输出每一条成分间相互作用预测结果的同时,会同步给出对应的置信分数,该分数用于量化评估预测结果的可靠性与准确性,置信分数越高,表明模型对该条预测关系的判断确定性越强。

2 结果

2.1 模型对比实验

为了全面、准确地评估模型性能,本研究使用 10 折交叉验证,将收集的包含 1 707 种化学药物的 191 807 条 DDI 标准数据集随机划分为 10 个子集,每次验证从 10 个子集中选取 9 个子集,同时为其分别匹配等量随机经过相似度筛选后的负样本以规避潜在的未发现相互作用。然后将二者合并构建训练集,用于分类模型的训练。剩余子集以同样方式匹配负样本形成测试集。3DFPMMI 模型训练 epoch 设定为 100,批处理长度为 32,学习率为 $1 \times$

10^{-4} , 随机种子为 42。

不同模型在 DrugBank 的 DDI 标准数据集上的预测结果见表 6。从表中数据能够清晰地看出, 本研究提出的 3DFPMMI 模型在 DDI 标准数据集上展现出了最优的预测性能, Acc 达到 0.921, AUPR 为 0.972, 相较于其他模型效果更佳。值得注意的是, 几

种模型的各项评价分数均在 0.9 以上, 部分分数甚至高达 0.97。这一结果表明, 基于分子化学序列和结构进行相互作用预测, 已经能够达到一个较为理想的性能水平, 同时也验证了本研究模型及所选对比模型预测结果的高可信度, 为中药成分相互作用预测领域的相关研究提供了坚实的理论依据与实践参考。

表 6 DrugBank 数据集下的成分相互作用预测结果

Table 6 Prediction results of ingredient interactions under DrugBank dataset

模型	Acc	AUC	AUPR	召回率	精确率	F1
3DFPMMI	0.921 (0.005)	0.972 (0.017)	0.954 (0.014)	0.976 (0.007)	0.955 (0.004)	0.965 (0.011)
DeepDDI	0.914 (0.021)	0.951 (0.012)	0.938 (0.006)	0.941 (0.019)	0.934 (0.004)	0.937 (0.025)
DeepDrug	0.917 (0.005)	0.961 (0.014)	0.944 (0.018)	0.957 (0.013)	0.944 (0.007)	0.950 (0.019)
MHAFR-DDI	0.919 (0.003)	0.963 (0.022)	0.944 (0.015)	0.967 (0.021)	0.947 (0.005)	0.955 (0.009)

括号内为 10 折交叉验证的标准差, 表 7 同。

Values in parentheses are standard deviations of 10-fold cross-validation, same as table 7.

2.2 超参数调整结果

不同参数设置下模型的性能表现如图 2 所示。从图 2 可明确看出, 当 ECFP 特征维度为 384, 3D 分子特征维度为 192, 隐藏神经元个数为 128 且学习率为 0.000 5 时, 模型性能最优。

2.3 消融实验结果

消融实验的具体结果如表 7 所示, 不同模型变体的性能均出现了一定程度的下降。与原始

3DFPMMI 模型相比, 3 种模型变体的各项性能评价指标均出现不同程度的下降, 充分证明了模型中各特征提取模块及特征融合模块的必要性。其中, 3DFPMMI-T 模型的性能下降最为显著, 这一结果表明, 分子 3D 结构特征提取方法具有更强的特征表征能力, 能够更深入地挖掘分子的高阶特征信息。在分子特征提取过程中, 高阶特征往往蕴含着分子结构、理化性质等更为丰富且复杂的关键信息,

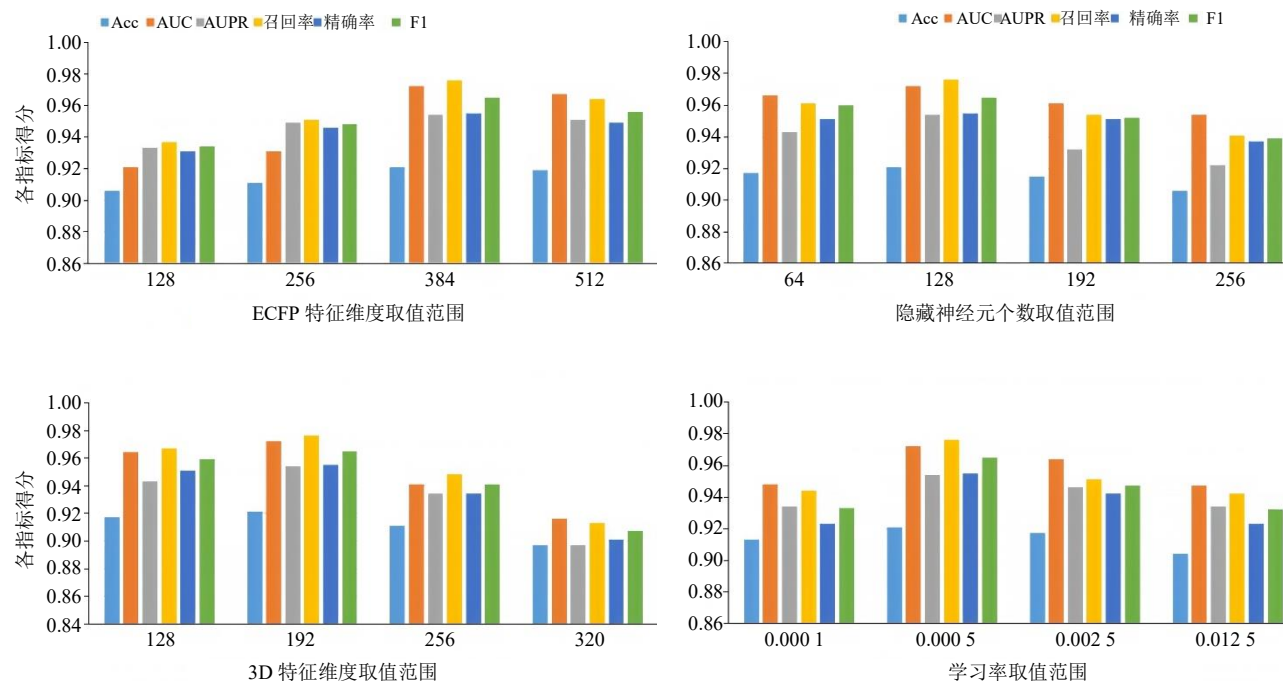


图 2 超参数调整实验结果

Fig. 2 Results of hyperparameter tuning experiments

表 7 消融实验结果
Table 7 Results of ablation experiments

模型	Acc	AUC	AUPR	召回率	精确率	F1
3DFPMMI	0.921 (0.005)	0.972 (0.017)	0.954 (0.014)	0.976 (0.007)	0.955 (0.004)	0.965 (0.011)
3DFPMMI-T	0.894 (0.003)	0.949 (0.006)	0.943 (0.001)	0.955 (0.015)	0.923 (0.016)	0.939 (0.023)
3DFPMMI-E	0.901 (0.019)	0.955 (0.007)	0.951 (0.017)	0.966 (0.003)	0.941 (0.004)	0.953 (0.009)
3DFPMMI-C	0.899 (0.006)	0.952 (0.009)	0.942 (0.006)	0.969 (0.009)	0.934 (0.021)	0.951 (0.016)

对准确解析分子特性、提升相互作用预测精度至关重要。此外，3DFPMMI-E 模型的性能也出现明显下降，这一结果验证了 ECFP 特征在分子特征提取中的有效性。ECFP 特征通过对分子结构进行系统编码，能够有效捕捉分子的局部结构特征与全局特征信息，为模型预测提供了重要的特征支撑，是分子相互作用预测的核心输入特征之一。同时，3DFPMMI-C 模型的各项评价指标也均下降，凸显了基于对比学习的特征融合方法的重要性。在分子相互作用预测任务中，该特征融合方法相较于传统的特征拼接方法表现更具优势：传统特征拼接方法仅简单将不同来源的特征进行组合，未能充分挖掘各类特征之间的内在关联与互补性，易导致特征冗余或信息丢失；而基于对比学习的特征融合方法，可通过学习不同类型特征之间的相似性与差异性，实现各类特征信息的高效整合，有效提升特征的表征能力，进而增强模型在中药成分相互作用预测中的性能表现。

2.4 案例分析

为进一步验证 3DFPMMI 模型预测的中药成分相互作用结果的实际价值与可靠性，便于后续缩小案例研究范围，本研究按置信分数降序选取排名前 20 位（末位并列，共 21 条）组合作为候选，见表 8。本研究选取表 8 中部分高置信度预测结果，结合临床疗效及相关文献研究，对其可靠性进行系统考究。从验证结果来看，表 8 中部分预测的中药成分相互作用关系与当前中医药领域的研究进展有一定契合，具体案例分析如下。

模型预测显示，毛蕊异黄酮葡萄糖苷和阿魏酸的相互作用关系，后者会提升前者的抗凝血活性。Zheng 等^[33]通过实验证明，在黄芪的提取过程中，阿魏酸能够增加毛蕊异黄酮的含量，从而促进黄芪的活血能力，这与模型预测结果相吻合。此外，模型预测山柰酚能够增强黄芪甲苷的活性，Ren 等^[34]通过分子对接实验证明，在抗炎机制中黄芪甲苷和

表 8 补阳还五汤和半夏泻心汤中药成分相互作用部分预测结果

Table 8 Partial prediction results of ingredient interactions in Buyang Huanwu Decoction and Banxia Xiexin Decoction

成分 1	成分 2	类别	置信分数
毛蕊异黄酮葡萄糖苷	阿魏酸	type 46	0.999 9
山柰酚	阿魏酸	type 46	0.999 9
山柰酚	芒柄花素	type 46	0.999 9
山柰酚	川芎嗪	type 46	0.999 9
芒柄花素	川芎嗪	type 46	0.999 8
小檗碱	6-姜酚	type 6	0.999 8
山柰酚	毛蕊异黄酮	type 46	0.999 8
毛蕊异黄酮葡萄糖苷	山柰酚	type 15	0.999 7
6-姜酚	甘草酸	type 17	0.999 6
黄芩苷	小檗碱	type 46	0.999 5
黄芩苷	麻黄碱	type 46	0.999 5
黄芪甲苷	山柰酚	type 10	0.999 5
黄芪甲苷	芍药苷	type 13	0.999 5
芒柄花素	山柰酚	type 46	0.999 5
6-姜酚	小檗碱	type 26	0.999 5
毛蕊异黄酮	阿魏酸	type 46	0.999 4
小檗碱	齐墩果酸	type 6	0.999 5
黄芪甲苷	川芎嗪	type 10	0.999 4
小檗碱	甘草酸	type 17	0.999 4
阿魏酸	芒柄花素	type 26	0.999 3
小檗碱	人参皂苷 Rb ₁	type 6	0.999 3

山柰酚能够与肿瘤坏死因子、Toll 样受体 4 和白细胞介素-10 靶点高效结合，呈现多靶点协同的作用，这也与模型预测结果有相似之处。模型还预测黄芪甲苷和川芎嗪具有协同作用，Rao 等^[35]通过实验证明黄芪甲苷和川芎嗪通过作用于 OIP5-反义 RNA1（OIP5 antisense RNA 1, OIP5-AS1）/微小 RNA-34a（microRNA-34a, miR-34a）/沉默信息调节因子 1（sirtuin 1, Sirt1）通路，在减轻炎症方面具有协同作用，该研究结论与模型预测结果完全一致，进一步验证了模型预测的准确性。此外，有研究通过实验证明 6-姜酚能够利用自身酸性将黄连中更多生

物碱包括小檗碱游离析出^[36]。有学者发现干姜和甘草配伍中,随着甘草用量增大,6-姜酚的含量有所下降^[37],这与预测结果中二者合用可降低6-姜酚的刺激性结果一致。有研究通过小鼠实验说明人参皂苷 Rb₁ 激活氧化物酶体增殖物激活受体 γ (peroxisome proliferator activated receptor γ , PPAR γ) 促进脂肪细胞的分化^[38],而小檗碱抑制 PPAR γ 抑制脂肪的形成,因此,二者可能存在相互干扰制约的现象。

综上,结合模型高置信度预测结果与现有文献研究可知,本研究 3DFPMI 模型预测出的部分中药成分相互作用关系与已有的实验研究及临床相关研究存在一定程度的吻合。这一结果充分表明,该模型具有一定实用价值,能够为中医药领域中药成分相互作用的后续研究提供相关理论指导与实践参考。

3 结论

中药复方凭借多成分、多靶点、多通路的整体调控特点,在一些复杂疾病的治疗中展现出独特的临床优势,然而复方中成分间相互作用关系不明,成为制约中药配伍机制深度解析与现代化发展的关键瓶颈。针对上述问题,本研究利用跨域泛化策略构建了一种基于分子 3D 结构特征与扩展连通性指纹的分子相互作用预测模型 3DFPMI,为中药成分间相互作用的预测提供了全新的技术方法,也为从分子层面解析中药复方配伍机制开辟了新的研究思路。

本研究提出的基于分子结构的分子相互作用深度学习预测方法 3DFPMI,以分子自身结构信息为唯一输入,模型通过 SchNet 深度学习框架从分子 3D 图结构中提取高阶空间结构特征,同步采用 ECFP 编码分子的拓扑结构与化学性质信息,再通过对比学习方法实现 2 类异质特征的深度融合,最终基于多层感知机构建分类预测模块,实现分子间相互作用的初步预测。研究采用 DrugBank 数据库中涵盖 1 707 种化学药物、191 807 条 DDI 记录的标准数据集完成模型训练与 10 折交叉验证,结果显示,3DFPMI 模型的预测 Acc 达 0.921, AUPR 达 0.972,各项评价指标均全面优于同样以分子结构为输入的基准模型,证实了模型具备优异的分子相互作用预测性能与泛化能力。消融实验结果进一步验证了模型各核心模块的有效性:去除分子 3D 结构特征提取模块后模型性能下降最为显著,证实

3D 空间结构特征对分子功能相关的高阶信息提取具有核心作用;ECFP 特征的引入可有效补充分子拓扑结构信息,与 3D 结构特征形成良好互补。

为验证模型的实际应用价值,本研究将训练完成的 3DFPMI 模型跨域泛化应用至中药成分相互作用预测场景,选取补阳还五汤、半夏泻心汤 2 首经典名方为研究对象,完成方剂核心有效成分间相互作用的系统预测,并按置信分数降序选取排名前 20 位的结果进行案例分析。结果显示模型预测的多组核心成分相互作用结果,包括毛蕊异黄酮与阿魏酸的抗凝血活性协同作用、黄芪甲苷与山柰酚的抗炎靶点协同效应、黄芪甲苷与川芎嗪的炎症通路协同调控作用、6-姜酚与小檗碱的成分溶出相互影响等,均与已发表的体内外实验研究结果高度吻合,初步证实该模型预测结果具有一定可靠性与应用潜力,可为中药复方配伍机制研究提供候选线索和方法学参考。同时以补阳还五汤和半夏泻心汤为例,利用该模型对方剂中的成分相互作用进行预测,并对预测结果进行了详细的阐释,结果表明提出的成分相互作用预测方法在研究方剂配伍机制内在原理方面具有一定潜力,能够为后续相关研究提供一定启发。

综上,本研究构建了一种仅依赖分子结构信息的分子相互作用预测模型,初步验证了其向中药成分相互作用预测场景迁移应用的可行性。本研究为解决中药成分相互作用数据集稀缺、现有 DDI 模型难以适配中药研究场景的问题,提供了新的技术路径与实证支持,为从分子层面阐释中药复方多成分的配伍规律奠定了方法学基础,也为中药复方的现代化研究与创新药物研发提供了新的视角与方法。

4 展望

本研究构建的 3DFPMI 模型在中药成分相互作用预测中初步展示了可行性与应用潜力,为中药复方配伍机制研究提供了新的技术路径,但研究仍存在一定的局限性,未来可从以下多个方面开展进一步的深入探索与优化完善。(1) 本研究目前主要通过已发表文献佐证预测结果的合理性,尚缺乏系统的体内外湿实验验证和标准化中药成分相互作用测试集。此外,DrugBank 训练数据与中药成分之间可能存在分布差异,影响模型跨域迁移的稳定性与适用范围。后续将针对高置信度预测结果开展实验验证,并补充二者在理化性质、结构复杂度、ECFP 指纹相似性及化学空间分布等方面的比较分

析,以明确模型适用边界;同时逐步构建中药成分相互作用金标准数据集,形成“预测-验证-优化”的迭代研究体系。(2)当前 DDI 相关负样本构造采用随机采样策略,虽已辅以分子相似度筛选,但仍存在将未发现的正例误标为负例的风险。后续将探索基于正-未标记学习等更严谨的负样本生成方法,并考虑其他数据集进行辅助验证。以进一步提升模型训练的可靠性。(3)当前模型仅以成分分子结构信息为输入开展预测,特征维度较为单一,缺乏多模态信息的融合支撑。后续可引入中药药性、药量配比、靶点活性等多维度中医药特色数据,构建多模态特征融合的预测框架,进一步提升模型的预测精度与场景适配性。(4)本模型目前仅可实现两两组分间的相互作用预测,而中药复方的临床疗效源于多成分、多靶点的整体协同调控,多成分间的交互作用普遍存在且动态关联。后续研究将重点突破两两成分相互作用的研究局限,探索多成分组合协同/拮抗效应的整体预测方法,使模型更贴合中药复方的整体作用特点,有望为经典名方的现代化深度开发与中药创新药研发提供新的计算工具与方法学支持。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Gong X, Sucher N J. Stroke therapy in traditional Chinese medicine (TCM): Prospects for drug discovery and development [J]. *Phytomedicine*, 2002, 9(5): 478-484.
- [2] Zhao L, Zhang H, Li N, *et al.* Network pharmacology, a promising approach to reveal the pharmacology mechanism of Chinese medicine formula [J]. *J Ethnopharmacol*, 2023, 309: 116306.
- [3] Wu J S, Zhang F Q, Li Z Z, *et al.* Integration strategy of network pharmacology in traditional Chinese medicine: A narrative review [J]. *J Tradit Chin Med*, 2022, 42(3): 479-486.
- [4] Wang H M, Xu M, Guo Z T, *et al.* DTI-BGCGCN: A novel bipartite graph-based cluster graph convolutional network for drug-target interaction prediction in modern and traditional Chinese medicine [J]. *Comput Biol Chem*, 2025, 117: 108410.
- [5] Lu Y Q, Chen J, Fan N N, *et al.* Machine learning models for drug-drug interaction prediction from computational discovery to clinical application [J]. *NPJ Digit Med*, 2026, 9: 198.
- [6] Hashir Q, Asfand E Yar M, Ullah A, *et al.* Computational approaches for drug-drug interaction prediction: A systematic review of data sources, modeling strategies, and evaluation frameworks [J]. *Front Pharmacol*, 2026, 17: 1816394.
- [7] Ahmad M M, Ali Shah Z B, Nies H W. A systematic review of molecular structures, knowledge graphs, and cold-start scenario in drug-drug interaction prediction [J]. *Comput Biol Med*, 2025, 190: 110122.
- [8] Zhang X X, Lei X J. Predicting miRNA-drug interactions via dual-channel network based on TCN and BiLSTM [J]. *Front Comput Sci*, 2024, 19(5): 195905.
- [9] Wang H P, Meng X J, Zhang Y. Biomolecular interaction prediction: The era of AI [J]. *Adv Sci*, 2025, 12(38): e09501.
- [10] le Maire A, Bourguet W. What structural biology tells us about the mode of action and detection of toxicants [J]. *Annu Rev Pharmacol Toxicol*, 2025, 65: 529-546.
- [11] Li X C, Chen K, Yang J X, *et al.* TLDA: A transfer learning based dual-augmentation strategy for traditional Chinese medicine syndrome differentiation in rare disease [J]. *Comput Biol Med*, 2024, 169: 107808.
- [12] Ren Z H, Zeng X X, Lao Y Z, *et al.* Predicting rare drug-drug interaction events with dual-granular structure-adaptive and pair variational representation [J]. *Nat Commun*, 2025, 16: 3997.
- [13] Budhkar A, Song Q Q, Su J, *et al.* Demystifying the black box: A survey on explainable artificial intelligence (XAI) in bioinformatics [J]. *Comput Struct Biotechnol J*, 2025, 27: 346-359.
- [14] Pan D P, Guo Y L, Fan Y F, *et al.* Development and application of traditional Chinese medicine using AI machine learning and deep learning strategies [J]. *Am J Chin Med*, 2024, 52(3): 605-623.
- [15] Knox C, Wilson M, Klinger C M, *et al.* DrugBank 6.0: The DrugBank knowledgebase for 2024 [J]. *Nucleic Acids Res*, 2024, 52(D1): D1265-D1275.
- [16] 张仲景, 钱超尘, 郝万山. 伤寒论 [M]. 北京: 人民卫生出版社, 2023: 28-188.
- [17] 王清任. 医林改错 [M]. 北京: 中国中医药出版社, 2023: 5-78.
- [18] 中国药典 [S]. 一部. 2025.
- [19] 孙景存, 张丁丁, 王嘉欣, 等. UPLC-MS/MS 法同时测定补阳还五汤中 17 种成分的含量 [J]. *中成药*, 2025, 47(12): 3903-3908.
- [20] 王若玲, 霍静娴, 张辉兰, 等. 半夏泻心汤化学成分、药理作用的研究进展及质量标志物的预测 [J]. *世界中医药*, 2024, 19(5): 719-726.
- [21] Gao K, Liu L, Lei S, *et al.* HERB 2.0: An updated database integrating clinical and experimental evidence for traditional Chinese medicine [J]. *Nucleic Acids Res*, 2025,

- 53(D1): D1404-D1414.
- [22] Schütt K T, Kindermans P J, Felix H E S, *et al.* SchNet: A continuous-filter convolutional neural network for modeling quantum interactions [A] // *Advances in Neural Information Processing Systems* 30 [C]. Long Beach: Curran Associates, Inc., 2017: 992-1002.
- [23] Popescu M-C, Balas V E, Perescu-Popescu L, *et al.* Multilayer perceptron and neural networks [J]. *WSEAS Transactions Circuits Systems*, 2009, 8(7): 579-588.
- [24] Naidu G, Zuva T, Sibanda E M. A review of evaluation metrics in machine learning algorithms [A] // *Artificial Intelligence Application in Networks and Systems* [M]. Cham: Springer International Publishing, 2023: 15-25.
- [25] Deng Y F, Xu X R, Qiu Y, *et al.* A multimodal deep learning framework for predicting drug-drug interaction events [J]. *Bioinformatics*, 2020, 36(15): 4316-4322.
- [26] Qiao J B, Guo X, Jin J R, *et al.* Taco-DDI: Accurate prediction of drug-drug interaction events using graph transformer-based architecture and dynamic co-attention matrices [J]. *Neural Netw*, 2025, 189: 107655.
- [27] Abbas K, Hao C, Yong X, *et al.* Graph neural network-based drug-drug interaction prediction [J]. *Sci Rep*, 2025, 15: 30340.
- [28] Ryu J Y, Kim H U, Lee S Y. Deep learning improves prediction of drug-drug and drug-food interactions [J]. *Proc Natl Acad Sci USA*, 2018, 115(18): E4304-E4311.
- [29] Yin Q J, Fan R, Cao X S, *et al.* DeepDrug: A general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction [J]. *Quant Biol*, 2023, 11(3): 260-274.
- [30] Li M L, Cao C, *et al.* MHAFFR-DDI: A multimodal hierarchical attention fusion and relation-aware architecture for drug-drug interaction event prediction [J]. *Brief Bioinform*, 2026, 27: bbag077.
- [31] Oruh J, Viriri S, Adegun A. Long short-term memory recurrent neural network for automatic speech recognition [J]. *IEEE Access*, 2022, 10: 30069-30079.
- [32] Franceschi L, Donini M, Perrone V, *et al.* Hyperparameter optimization in machine learning [J]. *FNT Machine Learning*, 2025, 18(6): 1054-1201.
- [33] Zheng K, Zhang Z, Du C, *et al.* Ferulic acid enhances the chemical and biological properties of *Astragali Radix*: A stimulator for Danggui Buxue Tang, an ancient Chinese herbal decoction [J]. *Planta Med*, 2014, 80(2/03): 159-164.
- [34] Ren J W, Ding Y T, Li S Z, *et al.* Predicting the anti-inflammatory mechanism of *Radix Astragali* using network pharmacology and molecular docking [J]. *Medicine*, 2023, 102(35): e34945.
- [35] Rao Y J, Li J J, Qiao R F, *et al.* Tetramethylpyrazine and Astragaloside IV have synergistic effects against spinal cord injury-induced neuropathic pain via the OIP5-AS1/miR-34a/Sirt1/NF- κ B axis [J]. *Int Immunopharmacol*, 2023, 115: 109546.
- [36] 黄文娟, 刘晶, 陈良, 等. 黄连-姜的不同比例配伍使用对有效成分含量的影响 [J]. *新疆医科大学学报*, 2018, 41(10): 1298-1301.
- [37] 李洁, 任爱农, 赵妍. 高效液相色谱法同时测定干姜甘草配伍后 6-姜辣素、甘草酸和甘草苷的含量变化 [J]. *中南药学*, 2016, 14(9): 994-997.
- [38] 尚文斌, 郭超, 于希忠, 等. 人参皂苷 Rb₁ 与小檗碱配伍对糖尿病小鼠糖脂代谢的影响 [J]. *时珍国医国药*, 2015, 26(3): 518-521.

[责任编辑 潘明佳]