

基于人工神经网络和支持向量机算法的药源性急性间质性肾炎计算模型的建立及在中药领域的应用

张文青, 赵 珊, 钱文秀, 阎星旭, 姚雅琦*, 李遇伯*

天津中医药大学中药学院, 天津 301617

摘要: **目的** 通过构建早期预测药源性急性间质性肾炎 (drug-induced acute interstitial nephritis, DI-AIN) 的计算模型, 筛选中药致 DI-AIN 成分。 **方法** 从文献和 SIDER 等数据库中收集得到了 125 个致 DI-AIN 的药物和 122 个未导致 DI-AIN 的药物, 作为预测模型的训练集, 基于人工神经网络 (artificial neural network, ANN) 和支持向量机 (support vector machine, SVM) 2 种算法构建模型。通过文献各选择 45 种药物进行验证, 以评估最优模型的预测性能。并将其应用于筛选 10 种中药的 DI-AIN 成分。 **结果** 共筛选得到 207 种分子描述符参与建模, 其中, ANN 和 SVM 算法搭建的最优模型分别包含 112 和 80 种分子描述符, 2 个模型的特异度、灵敏度、准确度均在 84% 以上。使用 2 种算法搭建的最优模型进行外部验证, 准确度均在 90% 以上。ANN 联合 SVM 模型预测为 DI-AIN 的中药成分有雷公藤甲素、水苏碱、京尼平苷等。 **结论** 首次建立中药成分致 DI-AIN 早期预测的计算模型, 具有良好的预测能力, 对于中药肾毒性预测研究及中药毒理学研究具有一定的应用价值。

关键词: 药源性急性间质性肾炎; 中药; 肾毒性预测; 人工神经网络; 支持向量机; 雷公藤甲素; 水苏碱; 京尼平苷

中图分类号: R285 文献标志码: A 文章编号: 0253-2670(2023)02-0416-09

DOI: 10.7501/j.issn.0253-2670.2023.02.08

Establishment of computational models for drug-induced acute interstitial nephritis based on algorithms of artificial neural network and support vector machine and their applications in traditional Chinese medicine

ZHANG Wen-qing, ZHAO Shan, QIAN Wen-xiu, YAN Xing-xu, YAO Ya-qi, LI Yu-bo

School of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China

Abstract: Objective To construct computational models for early prediction of drug-induced acute interstitial nephritis (DI-AIN) to screen the ingredients of traditional Chinese medicine (TCM)-induced DI-AIN. **Methods** A total of 125 DI-AIN-causing drugs and 122 non-DI-AIN-causing drugs were collected from literatures and SIDER database as training sets for prediction models. The models were constructed through the algorithms of artificial neural network (ANN) and support vector machine (SVM). In order to evaluate the prediction performance of the optimal model, a total of 45 DI-AIN-causing drugs and 15 non-DI-AIN-causing drugs were collected as external validation set through literatures. Two models were applied to screen the DI-AIN components of 10 TCMs. **Results** After filtering, 207 molecular descriptors were obtained to construct the model. The optimal models built by ANN and SVM algorithms contained 112 and 80 molecular descriptors respectively, and the specificity, sensitivity and accuracy of two models were all above 84%. The ANN and SVM models were used for external validation with accuracy above 90%. ANN and SVM models predicted that the TCM ingredients of DI-AIN included triptolide, stachydrine, geniposide. **Conclusion** This study established the computational model with reasonable accuracy for the early prediction of DI-AIN caused by the ingredients of TCM for the first time. It might be a promising method for studying nephrotoxicity and toxicology of TCM.

Key words: drug-induced acute interstitial nephritis; traditional Chinese medicine; renal toxicity prediction; artificial neural network; support vector machine; triptolide; stachydrine; geniposide

收稿日期: 2022-08-27

基金项目: 国家自然科学基金资助项目 (81573825); 国家中医药管理局青年岐黄学者支持项目

作者简介: 张文青, 硕士研究生。E-mail: zwq257595@163.com

*通信作者: 姚雅琦, 博士, 讲师, 从事手性药物研究。E-mail: yaopharmacy@163.com

李遇伯, 博士生导师, 教授, 从事计算毒理学、中药毒理评价研究。E-mail: yaowufenxi001@sina.com

急性间质性肾炎 (acute interstitial nephritis, AIN) 是一种肾脏病变, 具有起病急、进展快、预后差的特点^[1]。相关研究表明, 有 70%~90% 的 AIN 患者均由药物诱导^[2]。药源性急性间质性肾炎 (drug-induced acute interstitial nephritis, DI-AIN) 是一种对药物的特异性全身过敏反应, 发病时仅有少数患者会出现发热、嗜酸性粒细胞增多和皮疹等症状, 大多数 DI-AIN 患者的早期症状并不明显, 在开始用药到发生肾功能丧失之间有很长的潜伏期, 存在 DI-AIN 延迟发作的现象, 很容易造成误诊和漏诊, 从而延误 DI-AIN 的治疗^[3], 而一旦治疗不及时则会使病情加重, 一些患者会继续发展为慢性肾病, 甚至有可能由于持续的肾小管间质炎症和纤维化形成而导致永久性肾损伤。近些年来, 服用中药后发生 DI-AIN 的病例逐渐增多, 中药致 DI-AIN 的患者起病更为隐匿, 通常就诊时已经出现明显的肾功能损害。此外, 在 DI-AIN 前瞻性队列中, 单纯由中药所致者约占 10%, 其中多数病例为服用多组分的汤药, 其余病例服用中成药, 均无法确认单味的致病药物, 造成疾病诊断的困难。鉴于上述原因, 早期识别某一中药成分是否致 DI-AIN 并提出预警性认识至关重要^[4]。

传统的中药肾毒性研究方法一般采用体内外试验模型, 其优点是结果可靠且全面, 但基于细胞系的肾毒性体外评价模型缺乏靶器官毒性的特异性, 不能区分化合物的细胞毒性效应和肾毒性, 且当使用该方法进行高通量筛选时, 通常耗资较大, 时程较长^[5]。随着计算机技术的飞速发展, 基于计算机模拟方法预测药物性损伤及毒性成分的早期筛选已成为重要的研究方向。计算机模拟方法可以根据化合物的结构及毒性效应数据创建数学模型, 以预测其他已知结构化合物对机体的潜在毒性作用^[6-7], 从而减少分子生物学实验和传统药理学的耗时和成本, 并在短时间内实现对大量化合物毒性的预测。其中, 人工神经网络 (artificial neural network, ANN) 和支持向量机 (support vector machine, SVM) 算法常被用来搭建计算机模型。ANN 算法是模拟人脑神经建立起来的一种信息处理系统, 具有自主适应和自主学习的能力。SVM 是一种强大的监督学习方法, 通过使用几个核函数识别最优超平面来执行分类, 在高维空间中区分阳性和阴性^[8], 广泛应用于解决分类问题。这 2 种算法可通过预先提供的一批相互对应的输入、输出数据, 分析掌握两者之间潜

在的规律, 最终根据这些规律, 用新的输入数据来推算输出结果, 具有较好的非线性拟合能力^[9]。已有许多研究者根据药物的结构建立毒性预测模型, Zhao 等^[7]构建定量构效关系模型预测中药肝毒性, 模型的预测准确率达到 80.2%。然而, 这些模型预测的一般为整体毒性^[10], 目前较少基于其临床表现开发的计算机模型来预测特定的药物性损伤终点。此外, 建立药物分子结构与毒性之间的计算机模型并将其应用于中医药领域, 提供简单、精准、可靠度高的毒性物质初步筛选工具, 为复杂的中药体系的毒性研究提供思路^[5]。

以 DI-AIN 与计算模型为主题词进行在线组合检索, 未见预测模型报道。鉴于此, 本研究建立一种利用 ANN 和 SVM 算法搭建计算机模型来快速检测中药成分致 DI-AIN 的方法。将药物的分子描述符作为输入变量, 将对应的肾损伤作为输出变量, 开发了早期预测 DI-AIN 的模型, 并对模型进行内部和外部验证, 以评估预测的性能, 最终找到具有最优性能的模型并应用于中药成分筛选, 为 DI-AIN 的早期预测和及时诊断提供了基础依据, 避免发展为更严重的肾脏疾病, 同时为单味中药、复方中药的肾脏损伤做出快速的评价, 推动中医药现代化的进程。

1 材料和方法

1.1 数据集

从 PubMed、SIDER (<http://sideeffects.embl.de>) 和 AdisInsight 数据库 (<https://adisinsight.springer.com/>) 中进行检索, 选择文献或数据库中可能造成 DI-AIN 的药物, 共得到 170 种化合物作为阳性药物数据集^[11-28]; 在 AdisInsight 数据库的不良事件中, 以 NOT 肾小球肾炎、肾小管间质性肾炎、急性肾损伤、肾纤维化、肾病综合征、肾结石、肾功能衰竭、肾脏病进行检索, 随机选择不会造成肾损伤的药物, 共得到 167 种化合物作为阴性药物数据集。随后, 从阳性药物数据集中选择文献中常见的致 DI-AIN 的 45 种药物, 从阴性药物数据集中选择文献中未报道致 DI-AIN 的 45 种药物, 建立外部验证集, 并将剩余药物作为训练集, 数据集的具体组成见表 1。

表 1 数据集组成

Table 1 Data sets used for analysis

类别	训练集 (含内部验证)	外部验证集
阳性组	125	45
阴性组	122	45

1.2 分子描述符

化学结构需转化为分子描述符以表示结构的各种参数。目前 Dragon、PaDEL-Descriptor、CODESSA 等多种软件可用于计算分子描述符^[29]。本研究使用 Dragon 的远程在线版 E-Dragon (<http://www.vcclab.org/lab/edragon/>) 计算分子描述符。Dragon 提供了 1600 多种分子描述符, 这些描述符被划分为 20 个逻辑块, 包括结构描述符、拓扑描述符、几何描述符、电荷描述符、分子特性等^[30]。在 PubChem 中获取 SMILES 数据导入 E-Dragon 中, 即可获取所有化合物的分子描述符数据集^[31]。而计算得到的一些分子描述符都包含相似的信息, 并且过多的分子描述符也会导致过拟合现象, 因此需要使用筛选后的分子描述符构建模型^[32]。对分子描述符数据进行处理, 主要包括缺失数据处理以及相关性分析等。首先去除 0 或者存在缺失值的描述符^[33], 然后利用皮尔森相关系数的方法对描述符进行筛选, 其通过描述符与化合物性质之间的相关性, 排除对化合物性质贡献较低的分子描述符, 进而保证用于建模的描述符对化合物的某些性质影响较大^[29]。计算并筛选好分子描述符后用于模型的搭建。

1.3 模型构建

本研究分别应用 ANN 和 SVM 2 种算法构建 DI-AIN 的预测模型, 将药物的分子描述符作为输入变量, 将对应的肾损伤情况作为输出变量, 即造成 DI-AIN 的药物分为一类, 不造成 DI-AIN 的药物分为另一类。ANN 算法是人脑及其活动的一个理论化的数学模型, 由大量的神经元连接构成, 具有高度容错、并行分布处理、自适应、自组织以及自学习功能, 可以分析较为复杂的非线性系统^[34]。ANN 算法建立了化学物质的分子描述符与其毒理活性之间的非线性关系。它是由输入层、隐藏层和输出层构成, 输入层接受信息(描述符), 乘以一个系数(权重)并且该乘积被发送到包含神经元的隐藏层, 经过隐藏层激活函数的作用, 得到隐藏层的输出, 将其作为输出层的输入信息传递至输出层, 经过输出层激活函数作用后得到整个网络的输出。此外, 在模型训练时, 验证集采用“early stopping”方法, 当验证集的误差比上一次训练结果差时便停止模型的训练, 并使用上一次迭代结果中的参数应用于模型, 从而防止 ANN 的过度训练。SVM 是基于 Vapnik 统计学习理论的机器学习算法, 对于线性不可分数据集, 其通过核函数(径向基函数)将数据映射到 1

个新的高维表示, 在 2 个不同类别之间找到 1 个超平面作为决策边界, 并尽可能使超平面与每个类别最近的数据点距离最大化, 从而进行分类^[35]。

1.4 模型验证和性能评估

内部验证可以考察模型的稳健性, k 折交叉验证是常用于进行内部验证的方法, 其原理是将训练集随机分成 k 等份, 每次用 $(k-1)$ 份建模训练, 剩下的 1 份用作测试集进行验证, 如此循环直到每份都被当作测试集使用过为止。相同的过程执行 k 次, 取所有 k 次结果的均值给出最终的性能评估。本研究采用 10 折交叉验证方法对模型进行内部验证, 通过计算评价模型性能的指标, 选出最优模型^[36]。

模型性能的评估指标有准确度 (accuracy, Acc)、马修斯相关系数 (Matthew's correlation coefficient, MCC)、灵敏度 (sensitivity, SE)、特异度 (specificity, SP) 和受试者工作特征曲线下面积 (area under curve, AUC), 它们被广泛用于定义任何计算预测的真实性和有效性^[37]。其中, Acc 定义为所有化合物中正确预测的阳性和阴性总数占总数的百分比^[38]; SE 表示正确预测的阳性数量占阳性总数的百分比; SP 代表正确预测的阴性数量占阴性总数的百分比^[39]; AUC 表示模型能够区分类别的程度。其中一些参数的计算公式如下所示, 在这些等式中 TP (true positives)、TN (true negatives)、FP (false positives) 和 FN (false negatives) 分别是真阳性、真阴性、假阳性和假阴性的数量^[40]。

$$SE = TP / (TP + FN)$$

$$SP = TN / (TN + FP)$$

$$Acc = (TP + TN) / (TP + TN + FP + FN)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FN)(TN + FP)}}$$

本研究共进行了 2 次验证。首先, 对于内部验证, 训练集中的化合物用于开发计算模型, 根据内部验证集的预测性能确定最终模型。并分别计算各模型的 Acc、MCC、SE、SP 及 AUC, 根据计算结果选择 2 种算法构建的最优模型。随后, 将最优模型应用于完全外部的验证集, 以查看模型的泛化能力。模型的外部验证是指使用不参与模型构建的化合物对模型预测能力进一步验证。无法根据 DI-AIN 化合物的体内数据确定一个外部验证集, 以评估所建计算模型的预测性能。因此, 从文献中发现了一些与 DI-AIN 密切相关的化学物质作为外部验证集。

共选择 45 种致 DI-AIN 的化合物和 45 种不会致 DI-AIN 的化合物作为外部验证集, 应用于最优模型并计算 Acc、MCC、SE、SP 及 AUC, 进一步验证模型的预测能力^[41]。

2 结果

2.1 分子描述符的计算和筛选

在数据集中具有足够的代表性、并且与其他描述符的相关性低的描述符才能被保留用于建模。因此, 最终共筛选出 207 种分子描述符应用 ANN 和 SVM 算法构建模型。

2.2 模型的构建和内部验证

本研究共收集了 247 种药物构成训练集, 应用 ANN 或 SVM 算法开发 DI-AIN 早期预测模型。通过在训练集中划分内部验证集, 以评估 ANN 和 SVM 算法的性能, 最终筛选出最优模型。ANN 算法的最优模型包含了 112 种分子描述符和 1 个神经元, Acc 为 92.3%, SE 为 90.4%, SP 为 94.3%, MCC 为 0.847, AUC 为 0.923。SVM 算法的最优模型纳入了 80 种分子描述符, 其 Acc 为 89.1%, SE 为

84.8%, SP 为 93.4%, MCC 为 0.785, AUC 为 0.891。2 种算法的最优模型内部验证结果见图 1。ANN 和 SVM 模型的 Acc 和 SE 均大于 84%, 产生了良好的 DI-AIN 预测性能。

2.3 模型的外部验证

将外部验证集应用于 ANN 和 SVM 模型并计算模型的 Acc、SE、SP、MCC 及 AUC, 用于评价所建模型的预测能力。结果显示, ANN 模型的 Acc 为 90%, SE 为 93.3%, SP 为 86.7%, MCC 为 0.802, AUC 为 0.900。SVM 模型的 Acc 为 92.2%, SE 为 93.3%, SP 为 91.1%, MCC 为 0.845, AUC 为 0.922。2 种算法最优模型外部验证结果见图 2。此外, 用于外部验证的 90 种化学成分的全部预测结果汇总于图 3, 所有成分以紫色表示, 其中阳性预测错误的成分以红色字体标注, 阴性预测错误的以绿色字体标注。ANN 模型和 SVM 模型的外部验证集中阳性组预测结果较好, 双氯芬酸^[42]、头孢吡肟^[43]等可以导致 DI-AIN 的药物均被正确分类。因此, 2 种模型可以联合筛选中药致 DI-AIN 的成分。

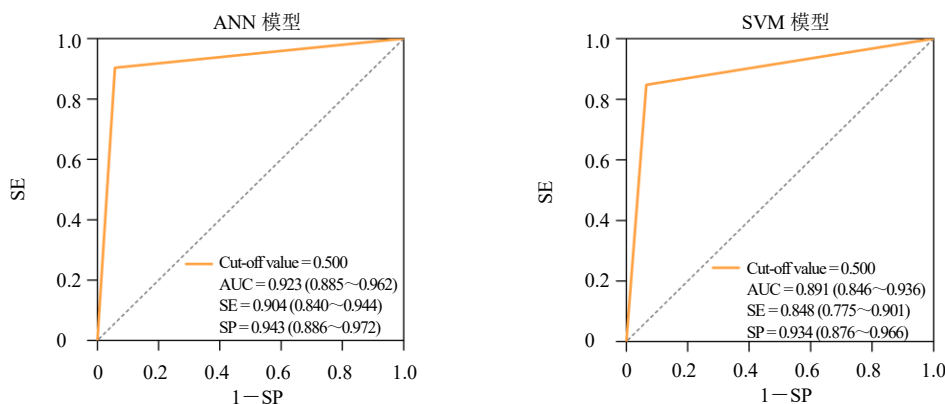


图 1 ANN、SVM 模型的内部验证性能

Fig. 1 Internal validation performance of ANN and SVM models

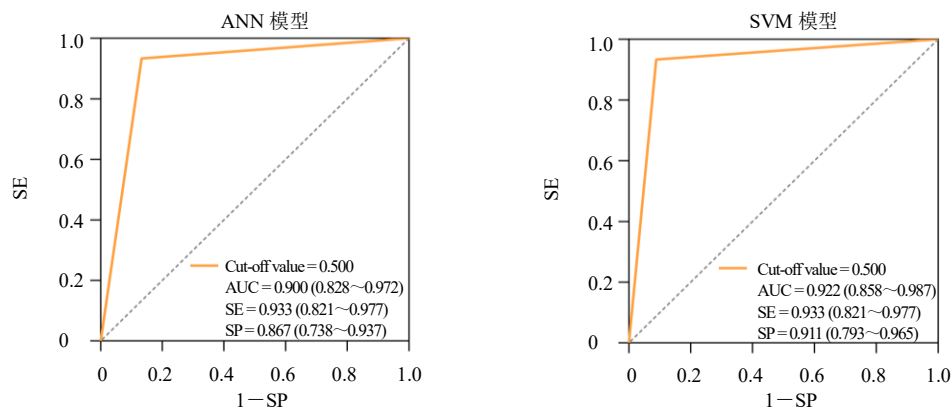


图 2 ANN、SVM 模型的外部验证性能

Fig. 2 External validation performance of ANN and SVM models

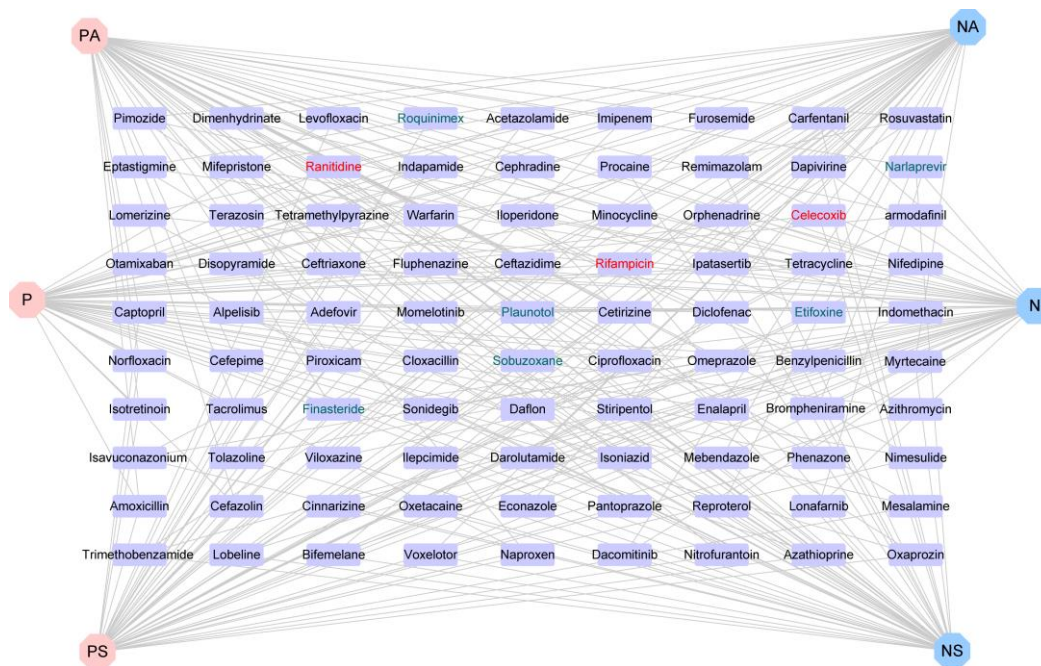


图3 外部验证结果 (P=positive, N=negative, A=ANN, S=SVM)

Fig. 3 External validation results (P = positive, N = negative, A = ANN, S = SVM)

2.4 模型的应用

国外文献及CFDA报道的具有肾毒性风险的中药主要分为马兜铃酸类中药、雷公藤制剂、穿心莲制剂及其他中药^[44]。除有毒中药外，一些传统中草药如泽泻、益母草也发现了不同程度的肾毒性^[45]。因此，使用所搭建的ANN和SVM模型筛选雷公藤、穿心莲、泽泻、益母草、栀子、大黄、苍耳子、商陆、斑蝥、木通10种中药中潜在的致DI-AIN成分。利用中药系统药理学数据库与分析平台(TCMSP)对这10种中药进行检索，以口服生物利用度(oral bioavailability, OB) ≥ 30%和类药性(drug-likeness, DL) ≥ 0.18为条件^[46]，并结合相关文献，共筛选得到雷公藤20种成分、穿心莲25种成分、

泽泻7种成分、益母草10种成分、栀子17种成分、大黄17种成分、苍耳子11种成分、商陆21种成分、斑蝥5种成分、木通8种成分。将最终得到的成分分别应用于ANN和SVM模型。结果显示，ANN和SVM模型均预测为阳性的中药成分有雷公藤9种、穿心莲10种、泽泻2种、益母草9种、栀子11种、大黄12种、苍耳子4种、商陆16种、斑蝥4种、木通5种，具体结果见图4。其中，雷公藤甲素、水苏碱、京尼平苷、苍术苷、羧基苍术苷、马兜铃酸B、常春藤皂苷元、穿心莲内酯等成分被模型预测为阳性，推测其可以导致DI-AIN。ANN联合SVM模型预测为阳性的其他中药成分具体见表2。

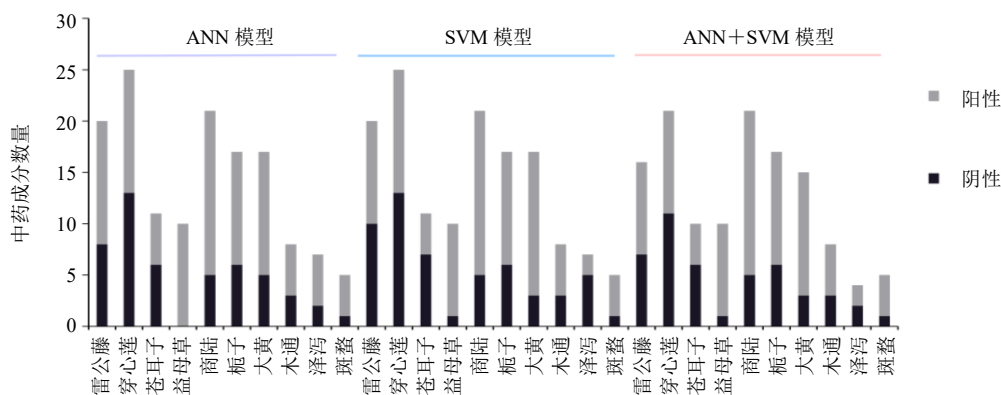


图4 10种中药DI-AIN成分预测结果

Fig. 4 Prediction results of DI-AIN ingredients in 10 TCMs

表 2 ANN 联合 SVM 模型筛选的中药 DI-AIN 成分

Table 2 Screening of DI-AIN ingredients of TCMs based on ANN and SVM models

中药	成分	中药	成分
雷公藤	15-羟基雷公藤内酯醇 清艾菊素 A 常春藤皂苷元 山海棠二萜内酯 A 白桦脂酸 雷公藤氯内酯醇 雷公藤甲素 雷公藤内酯酮 浙贝素	栀子	异山柰素 绿原酸 藏红花酸 藏红花素 京尼平 京尼平苷 常春藤皂苷元 山柰酚 槲皮素 芦丁 熊果酸
穿心莲	(3Z,4S)-3-[2-[(1R,4aS,5R,6R,8aS)-6-羟基-5,8a-二甲基-2-亚甲基-5-羟甲基十氢化萘-1-基]亚乙基]-4-羟基-四氢呋喃-2-酮 14-去氧-11-氧代-穿心莲内酯 14-去氧-12-甲氧基穿心莲内酯 14-脱氧穿心莲内酯 穿心莲内酯 穿心莲内酯19-β-D-葡萄糖苷_qt 去氧地胆草素 倍半萜内酯 B 倍半萜内酯 C 潘尼枯苷元	大黄	(-)-儿茶素 芦荟大黄素 大黄素 大黄素-1-O-β-D-吡喃葡萄糖苷 没食子酸-3-O-(6'-O-没食子酰)葡萄糖苷 掌叶二萜酮 A 大黄素甲醚双葡萄糖苷 原花青素 B-5,3'-O-没食子酸酯 大黄酸 番泻苷 D 番泻苷 E torachryson-8-O-β-D-(6'-oxayl)-glucoside
苍耳子	芦荟大黄素 苍术苷 羧基苍术苷 cynarin(e)	木通	ariskanin A 马兜铃酸 A 马兜铃酸 B 马兜铃苷 常春藤皂苷元
泽泻	泽泻醇 C 泽泻醇 C 醋酸酯	斑蝥	斑蝥素 亚油酸 去甲斑蝥素 油酸
商陆	esculentagenin 6,7-二羟香豆素 商陆皂苷 A esculentoside A_qt 商陆皂苷 B esculentoside C_qt 商陆皂苷 E esculentoside F_qt esculentoside J_qt esculentoside L_qt esculentoside N_qt esculentoside P_qt esculentoside Q_qt 加利果酸 齐墩果酸 phytolaccagenin	益母草	花生四烯酸 甘草素 异胆固醇 异鼠李素 山柰酚 益母草碱 卵磷脂 槲皮素 水苏碱

3 讨论

中药肾损害是目前国内外研究热点课题，中药毒性实验研究多为组分或部位的急性毒性和长期毒性试验，并没有进一步分离和确定其毒性物质，这就需要探讨中药成分肾损伤，进行毒性控制和炮制、配伍等减毒措施，为临床提供药理数据支撑，保证用药安全、合理^[47]。

雷公藤甲素是雷公藤引起肾毒性的成分，而雷

公藤可致急性间质性肾炎^[48]；水苏碱是益母草具有显著生理活性的成分之一，也是控制益母草药材质量的重要指标，体内实验发现大剂量长期使用益母草可致肾小管间质细胞损伤^[49]。还有研究以益母草成分盐酸水苏碱为受试药，观察其对小鼠 ig 给药的急性毒性，提示盐酸水苏碱可能为低毒化学物质^[50]；京尼平苷是栀子的有效成分之一，王晓慧等^[51]考察了京尼平苷的肾脏病理学特征，发现大鼠连续 3 d

给予高剂量京尼平苷，存在间质炎症细胞浸润；苍耳子中毒引起的肾脏病理损伤以肾小管和肾间质病变为主，其可能与苍耳子中的苍术苷或羧基苍术苷等成分有关^[52]；木通^[53]、莲必治注射液^[54]可引起急性间质性肾炎，其主要成分有马兜铃酸 B、常春藤皂苷元、穿心莲内酯等。ANN 和 SVM 模型筛查出的部分中药成分有文献佐证其对肾间质的损害，也进一步说明 ANN 和 SVM 模型联合应用于中药筛选致 DI-AIN 成分具有一定的准确性。

近年来，有研究^[55]建立了间质性肾炎（interstitial nephritis, IN）的预测模型，其利用 SVM 建立了含有指纹描述符的肾毒素和非肾毒素的二元分类模型，外部验证的预测准确性优于 83%。AIN 作为 IN 的一种类型，其发病较快，且为急性肾损伤的常见原因，而 DI-AIN 的发病率在 AIN 中占多数。事实上，最近在一项经活检证实的 DI-AIN 患者的回顾性研究中，30% 的患者不能识别出造成 DI-AIN 的药物。早期识别和停用引起 DI-AIN 的药物是初始治疗的基础^[56]。鉴于目前没有关于 DI-AIN 基于计算模型进行早期预测的研究，本研究首次使用 ANN 和 SVM 算法搭建了预测模型并应用于中药，得到较高的准确度。

数据的质量和多样性对于预测模型来说是一个重要问题。本研究的阳性组选择大多被临床报道可以致 DI-AIN 的化合物。鉴于根据特定的化学类别开发的模型适用范围有限^[55]，在此开发了多种药理学化合物肾毒性终点的数据集，阳性化合物组包括质子泵抑制剂、抗生素、非甾体抗炎药等不同类别药物组成。此外，当模型之间的唯一差异是建模算法时，比较了 2 种算法所构建模型的预测性能。在内部验证中，基于 ANN 模型在准确度、SE、SP、AUC 方面略胜于 SVM 模型。这可能是由于 ANN 模型具有更好的学习能力和更高的数据集特征提取能力，从而产生了比 SVM 模型更好的可预测性和泛化能力^[57]。随着中药资源庞大的成分信息不断纳入训练集，模型可进一步完善和优化，其结果为日后中药复杂体系肾脏毒性预测提供参考^[58]，并采用大规模筛查与毒理实验再验证相结合的方法，加快中药临床使用安全性的研究进展。

在模型的适用范围方面，本研究所使用的相似性度量是基于将化合物的相关信息转化为计算机可识别的数学信息。然后通过分子描述符对化合物的结构信息进行编码。该方法与其他方法的主要区别

在于，数学参数可用于表征分子描述符，并计算描述符值与毒性之间的相关性^[59]。虽然 ANN 和 SVM 模型在所提到的数据集中表现良好，但在其他领域中的应用效果还未证实。因此，构建高效且泛化能力强的预测模型是未来研究的重点之一^[60]。计算模型若要在毒性预测领域得到进一步的发展和创新，还应当完善以下 3 方面内容：①加强对毒性数据的整合工作，扩大数据量，并意识到建立大规模公共毒性数据库的重要性。本课题组已建立中药系统毒理学数据库（TCMSTD，<http://www.ehbio.com/Toxicity/front/#/>），这是首个关注中药安全性问题的信息共享平台，可为临床研究者、监管者和使用人员免费提供更为全面、系统的关于有毒中药、不良反应中药制剂、化学成分、毒性成分及毒性靶标等信息。②随着深度学习技术的引入，分子指纹，分子图像等也成为有效的输入形式。未来的研究中，应当研究新型的数据输入形式，丰富数据表达形式，挖掘更全面的数据内容。③加强与先进计算技术的联系，深入探索建模方法，开发高精度的计算与建模技术^[61]。

随着中药及其制剂在国内外的广泛使用，其安全问题日益突显^[62]，特别是中药肾毒性越来越受到关注。本研究建立一种通过计算模型来快速预测中药成分致 DI-AIN 的方法。采用 112 和 80 种分子描述符分别构建 ANN 和 SVM 模型以预测 DI-AIN。结果显示，ANN 和 SVM 模型在外部验证的准确度均达在 90% 以上。将其应用于中药并预测出雷公藤甲素、水苏碱、京尼平苷等成分可能会导致 DI-AIN。综上所述，首次搭建的计算模型可以用于中药致 DI-AIN 成分的筛选，对于正确全面认识中药肾毒性，规避肾脏疾病恶化风险具有重要意义。随着中药资源庞大的成分信息不断纳入训练集，模型可进一步优化，后续研究也可进行毒理实验再验证，加快中药临床使用安全性的研究进展。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Caravaca-Fontán F, Shabaka A, Sánchez-Álamo B, *et al.* Recurrent acute interstitial nephritis: What lies beneath [J]. *Clin Kidney J*, 2020, 14(1): 197-204.
- [2] Eddy A A. Drug-induced tubulointerstitial nephritis: Hypersensitivity and necroinflammatory pathways [J]. *Pediatr Nephrol*, 2020, 35(4): 547-554.
- [3] Moledina D G, Eadon M T, Calderon F, *et al.* Development and external validation of a diagnostic model for biopsy-

- proven acute interstitial nephritis using electronic health record data [J]. *Nephrol Dial Transplant*, 2022, 37(11): 2214-2222.
- [4] 王亚芳, 苏涛, 李晓玫, 等. 中药致急性肾小管间质性肾炎临床病理特点分析 [J]. 临床肾脏病杂志, 2015, 15(5): 272-276.
- [5] 刘睿, 李新宇, 李亚卓, 等. 网络毒理学及其在中药毒性成分预测中的应用研究 [J]. 药物评价研究, 2018, 41(5): 709-715.
- [6] 章建军, 胡纯琦, 王建平, 等. 新型二芳基三嗪类抗锥体虫病化合物的三维定量构效关系研究 [J]. 中国现代应用药学, 2017, 34(2): 156-160.
- [7] Zhao P, Liu B, Wang C Y, *et al.* Hepatotoxicity evaluation of traditional Chinese medicines using a computational molecular model [J]. *Clin Toxicol*, 2017, 55(9): 996-1000.
- [8] Jaganathan K, Tayara H, Chong K T. Prediction of drug-induced liver toxicity using SVM and optimal descriptor sets [J]. *Int J Mol Sci*, 2021, 22(15): 8073.
- [9] 张琦, 李青, 冷光, 等. 人工神经网络模型预测药物性肝损伤的临床转归及影响因素 MIV 值评价 [J]. 药物流行病学杂志, 2018, 27(7): 444-449.
- [10] Shin H K, Kang M G, Park D, *et al.* Development of prediction models for drug-induced cholestasis, cirrhosis, hepatitis, and steatosis based on drug and drug metabolite structures [J]. *Front Pharmacol*, 2020, 11: 67.
- [11] Nast C C. Medication-induced interstitial nephritis in the 21st century [J]. *Adv Chronic Kidney Dis*, 2017, 24(2): 72-79.
- [12] Hosohata K. Role of oxidative stress in drug-induced kidney injury [J]. *Int J Mol Sci*, 2016, 17(11): 1826.
- [13] Muriithi A K, Leung N, Valeri A M, *et al.* Biopsy-proven acute interstitial nephritis, 1993-2011: A case series [J]. *Am J Kidney Dis*, 2014, 64(4): 558-566.
- [14] Raghavan R, Shawar S. Mechanisms of drug-induced interstitial nephritis [J]. *Adv Chronic Kidney Dis*, 2017, 24(2): 64-71.
- [15] Raghavan R, Eknoyan G. Acute interstitial nephritis-a reappraisal and update [J]. *Clin Nephrol*, 2014, 82(3): 149-162.
- [16] 丁小强, 傅辰生. 药物所致急性间质性肾炎 [J]. 中华肾脏病杂志, 2005(3): 123-124.
- [17] Martinez de la Cruz P, Shabaka A, Mielgo-Rubio X, *et al.* Concomitant acute tubular necrosis and acute interstitial nephritis induced by tipifarnib in a patient with squamous cell carcinoma of the lung [J]. *Am J Med Sci*, 2021, 362(1): 99-102.
- [18] Asnake M, Henock A, Abayneh M, *et al.* Acute interstitial nephritis with prothionamide [J]. *SAGE Open Med Case Rep*, 2022, 10: 2050313X221094076.
- [19] Yilmaz H, Bilgiç M A, Bavbek N, *et al.* Cefpodoxime proxetil-related hemolysis and acute interstitial nephritis [J]. *Curr Drug Saf*, 2013, 8(2): 145-147.
- [20] Singh M, Shailesh F, Tiwari U, *et al.* Phenazopyridine associated acute interstitial nephritis and review of literature [J]. *Ren Fail*, 2014, 36(5): 804-807.
- [21] Rathod N, Pai P. Ertapenem-induced acute interstitial nephritis (AIN) in a case of protein S deficiency and factor V leiden mutation with deep vein thrombosis [J]. *J Assoc Physicians India*, 2016, 64(3): 85-86.
- [22] Chen T, Xu P C, Hu S Y, *et al.* Severe acute interstitial nephritis induced by valsartan: A case report [J]. *Medicine*, 2019, 98(6): e14428.
- [23] DiMaria C, Hanna W, Murone J, *et al.* Direct oral anticoagulant and AKI: Apixaban-induced acute interstitial nephritis [J]. *BMJ Case Rep*, 2019, 12(6): e230371.
- [24] Golbin L, Dolley-Hitze T, Lorcy N, *et al.* Drug-induced acute interstitial nephritis with nifedipine [J]. *Case Rep Nephrol*, 2016, 2016: 1971465.
- [25] Markóth C, File I, Szász R, *et al.* Ibrutinib-induced acute kidney injury via interstitial nephritis [J]. *Ren Fail*, 2021, 43(1): 335-339.
- [26] Zafar F, Iqbal A M, Mubarik A, *et al.* Rivaroxaban-induced acute interstitial nephritis: A case report [J]. *Am J Case Rep*, 2019, 20: 1719-1722.
- [27] Shao E X, Wilson G J, Ranganathan D. Phentermine induced acute interstitial nephritis [J]. *BMJ Case Rep*, 2017, 2017: bcr2017219452.
- [28] Chowdry A M, Azad H, Mir I, *et al.* Drug-induced acute interstitial nephritis: Prospective randomized trial comparing oral steroids and high-dose intravenous pulse steroid therapy in guiding the treatment of this condition [J]. *Saudi J Kidney Dis Transpl*, 2018, 29(3): 598-607.
- [29] 赵朝阳. 有机过氧化物自加速分解温度的预测模型研究 [D]. 大连: 大连理工大学, 2021.
- [30] 万金玉, 刘怡飞. 基于 QSAR 模型的有机磷化合物毒性预测 [J]. 化学通报, 2019, 82(10): 926-936.
- [31] 靳樱子. 基于气味物化特征空间的嗅感评价模型研究 [D]. 广州: 广东工业大学, 2020.
- [32] 陈雅菲. 基于人工神经网络和支持向量机方法的亚硝酸胺定量构效关系研究 [D]. 北京: 北京工业大学, 2017.
- [33] 胡莹. 基于分子动力学和机器学习的水相关性预测研究 [D]. 成都: 四川师范大学, 2021.
- [34] Mao W B, Lyu J Y, Vaishnani D K, *et al.* Application of artificial neural networks in detection and diagnosis of gastrointestinal and liver tumors [J]. *World J Clin Cases*, 2020, 8(18): 3971-3977.

- [35] 吴思甜. 基于机器学习算法对化学品致突变性的预测研究 [D]. 大连: 大连理工大学, 2021.
- [36] Chaudhary P K, Pachori R B. FBSED based automatic diagnosis of COVID-19 using X-ray and CT images [J]. *Comput Biol Med*, 2021, 134: 104454.
- [37] Khan Z U, Pi D C. DeepSSPred: A deep learning based sulfenylation site predictor via a novel nSegmented optimize federated feature encoder [J]. *Protein Pept Lett*, 2021, 28(6): 708-721.
- [38] Chicco D, Tötsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation [J]. *BioData Min*, 2021, 14: 13.
- [39] Monaghan T F, Rahman S N, Agudelo C W, et al. Foundational statistical principles in medical research: Sensitivity, specificity, positive predictive value, and negative predictive value [J]. *Medicina*, 2021, 57(5): 503.
- [40] Wang H S, Tang Z R, Li W L, et al. Prediction of the risk of C5 palsy after posterior laminectomy and fusion with cervical myelopathy using a support vector machine: An analysis of 184 consecutive patients [J]. *J Orthop Surg Res*, 2021, 16(1): 332.
- [41] Feng H W, Zhang L, Li S M, et al. Predicting the reproductive toxicity of chemicals using ensemble learning methods and molecular fingerprints [J]. *Toxicol Lett*, 2021, 340: 4-14.
- [42] Ozalp F R, Karadeniz T, Alp A. An old enemy: Still nephrotoxic [J]. *J Coll Physicians Surg Pak*, 2021, 31(2): 228-231.
- [43] Mac K, Chavada R, Paull S, et al. Cefepime induced acute interstitial nephritis: A case report [J]. *BMC Nephrol*, 2015, 16: 15.
- [44] 冯雪. 雷公藤制剂安全性 (肾毒性) 的循证评价研究 [D]. 北京: 北京中医药大学, 2017.
- [45] 朴静, 迟继铭. 常用治疗慢性肾脏病的方药中对肾毒性的药物解析 [J]. *黑龙江中医药*, 2016, 45(6): 65-66.
- [46] 林晓彤, 黄重铭, 彭慧婷, 等. 基于网络药理学和分子对接探究斑蝥治疗肝癌的作用机制 [J]. *中国医药导报*, 2021, 18(23): 8-12.
- [47] 孙蓉, 冯群, 赵庆华, 等. 益母草毒性研究进展 [J]. *中国药物警戒*, 2014, 11(2): 70-73.
- [48] 李成, 朱玉华, 孙晓旻, 等. 基于网络药理学和分子对接技术探讨雷公藤致急性肾损伤的多重作用机制 [J]. *中国药学*, 2021, 30(7): 14.
- [49] 刘红燕, 王瑞, 石明, 等. 小剂量盐酸水苏碱对过氧化氢所致肾小管上皮细胞凋亡的保护作用 [J]. *中国中西医结合肾病杂志*, 2008, 9(9): 760-763.
- [50] 石金金, 张美玲, 于扬, 等. 益母草成分盐酸水苏碱对小鼠的急性毒性研究 [J]. *时珍国医国药*, 2015, 26(4): 865-866.
- [51] 王晓慧, 张帆, 席莉莉, 等. 栀子的药效与毒性及其作用机制研究进展 [J]. *中国药物警戒*, 2021, 18(1): 94-99.
- [52] 栾弘, 王传孝, 徐岩, 等. 苍耳子中毒致急性肾功能衰竭一例 [J]. *临床肾脏病杂志*, 2019, 19(4): 303-304.
- [53] 王丽. 急性间质性肾炎的临床表现及治疗方法 [J]. *世界最新医学信息文摘*, 2018, 18(69): 3-4.
- [54] 封华, 马勤, 于兰, 等. 莲必治注射液的临床应用与不良反应 [J]. *中国医药导报*, 2008, 5(11): 97-98.
- [55] Lee S H, Kang Y M, Park H, et al. Human nephrotoxicity prediction models for three types of kidney injury based on data sets of pharmacological compounds and their metabolites [J]. *Chem Res Toxicol*, 2013, 26(11): 1652-1659.
- [56] Fernandez-Juarez G, Perez J V, Caravaca-Fontán F, et al. Duration of treatment with corticosteroids and recovery of kidney function in acute interstitial nephritis [J]. *Clin J Am Soc Nephrol*, 2018, 13(12): 1851-1858.
- [57] Wong Y J, Shimizu Y, Kamiya A, et al. Application of artificial intelligence methods for monsoonal river classification in Selangor river basin, Malaysia [J]. *Environ Monit Assess*, 2021, 193(7): 438.
- [58] 郭晓, 王萌, 朱彦, 等. 中药肾毒性机制研究现状及评价方法研究进展 [J]. *中草药*, 2015, 46(23): 3581-3591.
- [59] Wang Y Y, Xiao Q X, Chen P, et al. In silico prediction of drug-induced liver injury based on ensemble classifier method [J]. *Int J Mol Sci*, 2019, 20(17): 4106.
- [60] 陈颖皓. 几种数据挖掘方法在药毒物 QSAR 预测中的应用研究 [D]. 景德镇: 景德镇陶瓷大学, 2016.
- [61] 查金苗, 郑子廷, 闫赛红. 定量构效关系模型在化合物毒性预测领域的应用、发展与展望 [J]. *环境科学学报*, 2022, 42(10): 1-11.
- [62] 段敬怡, 许妍妍, 王玉明, 李遇伯. 基于有害结局路径的中药毒性进程动态机制研究思路 [J]. *中国药理学与毒理学杂志*, 2020, 34(1): 58-63.

[责任编辑 李亚楠]