

基于机器学习筛选抗人参根腐病尖孢镰刀菌的天然药物分子

赵贵萍^{1,2}, 杨若琪³, 李洁¹, 陈莹莹^{1,2}, 于大德^{1*}, 李西文^{1*}

(1. 中国中医科学院中药研究所, 北京 100700; 2. 云南中医药大学中药学院, 云南 昆明 650500;
3. 山东中医药大学药学院, 山东 济南 250355)

摘要: 尖孢镰刀菌 *Fusarium oxysporum* 广泛存在于农田土壤中, 是根腐病的主要致病真菌之一, 严重影响植物的生长发育, 进而影响了经济作物的产量, 往往造成严重的损失。为了更经济、高效地筛选出抑制尖孢镰刀菌活性的天然化合物, 本研究基于机器学习算法, 利用 ChEMBL 数据库中已知的抑菌化合物信息构建了随机森林 (random forest)、支持向量机 (support vector machine)、人工神经网络 (artificial neural network) 等 3 种预测模型, 筛选对尖孢镰刀菌有抑制作用的新型天然药物分子, 并对筛选的药物进行抑菌活性验证。结果显示, 3 种模型预测准确率分别达到了 77.58%、83.03%、81.21%, 经抑菌实验验证, 芒柄花苷 (ononin) 抑制效果最佳 (MIC = 0.312 5 mg·mL⁻¹)。本研究提出的虚拟筛选方法可为天然产物来源的农药研发和创制提供思路, 筛选出的芒柄花苷可作为尖孢镰刀菌新型抑制剂开发的潜在先导化合物。

关键词: 尖孢镰刀菌; 天然药物分子; 随机森林; 支持向量机; 人工神经网络; 人参

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2023)06-1713-09

Screening of natural drug molecules against *Fusarium oxysporum* of ginseng root rot based on machine learning

ZHAO Gui-ping^{1,2}, YANG Ruo-qi³, LI Jie¹, CHEN Ying-ying^{1,2}, YU Da-de^{1*}, LI Xi-wen^{1*}

(1. Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China;
2. College of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Kunming 650500, China;
3. School of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan 250355, China)

Abstract: *Fusarium oxysporum* widely exists in farmland soil and is one of the main pathogenic fungi of root rot, which seriously affects the growth and development of plants and often causes serious losses of cash crops. In order to screen out natural compounds that inhibit the activity of *Fusarium oxysporum* more economically and efficiently, random forest, support vector machine and artificial neural network based on machine learning algorithms were constructed using the information of known inhibitory compounds in ChEMBL database in this study. And the antibacterial activity of the screened drugs was verified thereafter. The results showed that the prediction accuracy of the three models reached 77.58%, 83.03% and 81.21%, respectively. Based on the inhibition experiment, the best inhibition effect (MIC = 0.312 5 mg·mL⁻¹) of ononin was verified. The virtual screening method proposed in this study provides ideas for the development and creation of new pesticides derived from natural products, and the screened ononin is expected to be a potential lead compound for the development of novel inhibitors of *Fusarium oxysporum*.

Key words: *Fusarium oxysporum*; natural drug molecule; random forest; support vector machine; artificial neural network; ginseng

收稿日期: 2022-12-22; 修回日期: 2023-01-24.

基金项目: 国家重点研发计划项目 (2019YFC1710601); 中国中医科学院科技创新工程项目 (CI2021A04505, CI2021A03910); 中央级公益性科研院所基
本科研业务费专项资金 (ZZ15-YQ-031, ZXKT21033).

*通讯作者 E-mail: ddyu@icmm.ac.cn; xwli@icmm.ac.cn

DOI: 10.16438/j.0513-4870.2022-1404

根腐病是引起经济作物减产的主要植物病害之一。尖孢镰刀菌 (*Fusarium oxysporum*) 是世界五大植物病原真菌之一^[1], 通过土壤传播^[2], 是根腐病的主要致病菌。其菌丝沿着植物根部生长, 破坏植物表皮进而侵染整个根部, 导致根腐烂、坏死^[3], 同时引发作物连作障碍^[4]。多种名贵中药材, 如人参^[5]、丹参^[6]、三七^[7]等多年生、以根茎入药的药用植物的栽培往往受到尖孢镰刀菌导致的土传病害影响而严重减产。目前常用的根腐病防治方法为化学杀菌剂消杀, 频繁使用不仅会产生病原菌的耐药性, 也会对土壤微生态环境造成破坏^[8]。天然产物具有多种化学结构及广泛的药理活性, 是开发新药和药物先导化合物的主要来源^[9,10]。天然产物来源的农药目前已占据农药种类的半数以上^[11], 广泛用于除草剂、杀菌剂和杀虫剂等环境友好型农药的开发。因此, 从巨大的天然产物库中筛选抑制尖孢镰刀菌的活性物质并应用于根腐病的防治具有较大的潜力。已有研究报道, 薄荷挥发油^[12]、水杨酸^[13]、大蓟总黄酮^[14]等多种天然产物对尖孢镰刀菌有显著的抑制作用。但由于目前尖孢镰刀菌的致病机制尚不明确^[15], 相应天然产物源农药的研发基础还非常薄弱^[14], 生产上尚缺乏成熟的抑菌产品。传统的农药创制方法包括偶然发现、随机筛选和模仿已知农用药物化学结构^[16]等, 研发、生产过程不确定性较大、周期长、成本高。基于机器学习的虚拟筛选技术可以有效缩短药物筛选的时间, 利用数据库中已有的化合物数据构建预测模型, 对抑菌活性未知的天然药物进行筛选, 从而节省大量的时间和经济成本^[17]。

随着计算机辅助药物设计技术的发展, 农药研发已经从“偶然发现”模式逐渐转型为“理性设计”模式^[18]。机器学习近年来被广泛应用于先导化合物的发现以及分子理化性质和生物学特性的预测。相比于传统的计算方法, 机器学习通过构建不同模型学习现有数据中复杂的非线性关系对未知数据进行合理的预测^[19,20], 可以缩小目标候选药物范围, 提高筛选效率, 为后续药物开发减少盲目性。目前常用于构建模型的算法中, 随机森林 (random forest, RF) 的基本单元是单一决策树, 主要思想是集成学习, 对于同一个输入的样本, 可以随机从训练集中获取特征, 通过对所有决策树的预测结果进行平等投票, 最终输出为投票最多的类别^[21]。支持向量机 (support vector machine, SVM) 是一种基于核函数的二元分类模型, 其核心思想在于利用间隔最大化寻找最优超平面对训练样本分类, 不依赖于样本空间的维数^[22]。人工神经网络 (artificial neural network, ANN) 的架构由多个分布在不同层的神经元相互连接而成, 首先通过前向传播计算损失, 再通过反

向传播更新权重, 反复进行这个过程直到找到能够使模型损失最小的最佳权重^[23]。赵晓等^[24]基于 RF、神经网络和 SVM 等算法预测了农药对雄蜂和蜜蜂的毒性, 准确率达到 86.7% 和 80.0%。Sakthivel 等^[25]利用同源建模、分子对接和分子动力学模拟筛选了 ZINC 数据库中潜在的对棉花害虫有效的广谱天然药物。针对水稻纹枯病菌, Islam 等^[26]以分子对接成功预测了 D1、D2、D3 潜在化合物是防治水稻纹枯病的有效成分, 为该种病菌的农药精准研制提供了可行性方案。因此, 采用机器学习方法辅助植物源农药设计具有可行性且能极大提高研发效率。

本研究基于天然化合物的分子描述符, 采用机器学习方法构建高精度的预测模型, 筛选潜在的抑制尖孢镰刀菌活性的天然产物, 并对候选化合物进行抑菌活性验证以评估筛选结果的准确性, 从而为探究基于机器学习方法筛选抑制尖孢镰刀菌活性的天然药物提供研究思路。

材料与方法

材料 尖孢镰刀菌 (菌株编号 3.11126), 由中国普通微生物菌种保藏管理中心 (CGMCC) 提供, 分离自吉林省栽参土壤; 筛选后的天然药物分子购自成都普瑞法科技开发有限公司, 药物纯度均 > 96.0%; 多菌灵农药登记号 (PD20070286), 为苯并咪唑类杀菌剂, 购自安徽广信农化股份有限公司; 人参植株采挖自吉林省白山市靖宇县人参栽培基地, 经中国中医科学院中药研究所李西文研究员鉴定为 2 年生人参。

马铃薯葡萄糖琼脂 (potato dextrose agar, PDA) 培养基: 200 g 新鲜去皮马铃薯、15 g 琼脂、20 g 葡萄糖、蒸馏水 1 000 mL。PDA 中不添加琼脂即为马铃薯葡萄糖 (potato dextrose, PD) 液体培养基。

实验中所使用试剂均为国产分析纯。高压蒸汽灭菌锅 (日本雅玛拓公司); 恒温震荡培养箱 (常州普天仪器制造有限公司); 血球计数板 (上海求精生化试剂仪器有限公司); 光学显微镜 (深圳市奥斯微光学仪器有限公司); 人工气候箱 (宁波杨辉仪器有限公司)。

数据预处理、分子表征、特征工程 本研究首先从欧洲生物信息研究所开发的 ChEMBL 靶点与生物活性数据库 (<https://www.ebi.ac.uk/chembl/>)^[27] 中收集关于尖孢镰刀菌的信息, 删除了重复、没有最小抑菌浓度 (MIC) 或者有多个 MIC 的化合物。将 MIC < 50 $\mu\text{g}\cdot\text{mL}^{-1}$ 的化合物标记为 0 (代表抑制剂), MIC \geq 50 $\mu\text{g}\cdot\text{mL}^{-1}$ 标记为 1 (代表非抑制剂)。按照 3:1 的比例将数据集划分为训练集和测试集, 并使用 Python 3.7.0 工具包 Scikit-learn 中 Standard scalar 功能进行数据标准化。使用分子

描述符作为数据集中每个化合物的表征形式。分子描述符由拓扑描述符、组成描述符及几何描述符等不同的类型组成^[28], 其使用数字的特征来描述化合物分子的化学和生物特征。在本研究中, 利用开源的 Python 工具包 RDKit 计算 208 个表示分子理化性质和结构特征的分子描述符。为避免冗余、无关的特征降低模型的泛化能力, 导致过拟合现象, 本研究使用 RFE-递归特征消除来寻找最佳特征, 最终保留了 60 个分子描述符用于机器学习模型的构建。

机器学习模型的生成与评估 通过使用开源的 Python 3.7.0 工具包 Scikit-learn 生成机器学习模型^[29]。本研究使用了 RF、SVM 及 ANN 3 种机器学习算法进行模型的构建。使用网格搜索和学习曲线来调整模型的参数以使模型达到最佳预测结果^[30], 最优的超参数组合为 RF: $n_estimators = 187$, $max_features = 0.271$, $max_depth = 14$; SVM: $gamma = 0.014$, $C = 4.05$; ANN: $hidden_layer_sizes = 76$, $alpha = 0.071$, $max_iter = 1\ 000$ 。

使用训练集的五折交叉验证与测试集来评估模型的预测性能与泛化能力。通过计算不同指标, 包括准确率 (accuracy)、精确度 (precision)、召回率 (recall)、F1 得分 (F1-score)、受试者工作特征曲线下的面积 (area under curve, AUC), 比较各模型的性能。其中, 精确度表示真实为抑制剂的化合物占模型预测为抑制剂的化合物的比例; 召回率代表模型正确预测的所有抑制剂化合物的比例; F1 得分是精密度和召回率之间的平衡系数; AUC 值越接近 1, 模型的分类能力越强。最后通过绘制受试者工作特征曲线 (ROC) 和混淆矩阵图来进一步综合评估模型的性能。具体计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

其中, TP (true positive) 表示正确预测尖孢镰刀菌活性抑制剂的数量, TN (true negative) 表示正确预测尖孢镰刀菌活性非抑制剂的数量, FP (false positive) 则表示被错误分类为尖孢镰刀菌活性抑制剂的数量, FN (false negative) 代表被错误分类为尖孢镰刀菌活性非抑制剂的数量。

天然药物分子的筛选 使用基于分子描述符的 3 个模型对 ZINC 虚拟筛选数据库 (<https://zinc.docking.org/>) 中的 4 114 个天然产物进行虚拟筛选。首先计算这些化合物的 208 个 RDKit 分子描述符, 然后进行同

样的数据预处理与特征选择步骤。根据模型的平均得分, 筛选出预测得分前 20 的化合物作为候选化合物集。考虑到化合物的来源、提取难度以及成本, 最终选择 5 种天然药物进一步进行实验验证。本研究的具体流程如图 1 所示。

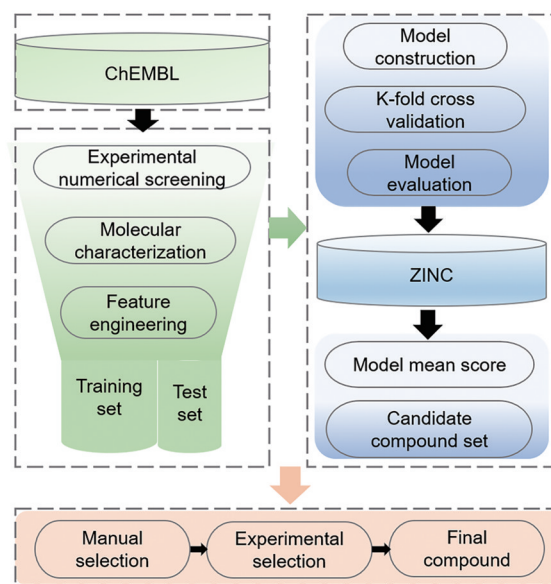


Figure 1 Screening process of active drugs against *Fusarium oxysporum*

抑菌活性验证实验 将 5 种筛选出的候选药物分别溶于二甲亚砜 (DMSO), 配制成 $10\text{ mg}\cdot\text{mL}^{-1}$ 的母液常温保存。阳性对照为多菌灵 (无菌水配制成 $2\text{ mg}\cdot\text{mL}^{-1}$ 溶液), 阴性对照为 DMSO 溶液。取生长 7 天的尖孢镰刀菌, 在培养皿中加入无菌水, 使用 9 层无菌纱布去除菌丝和培养基, 制备浓度为每毫升 5.78×10^6 CFU 的孢子悬浊液。

MIC 测定: 采用微量液基稀释法^[31]测定各候选药物对尖孢镰刀菌菌株的 MIC 值。将 96 孔培养板置于 $25\text{ }^{\circ}\text{C}$ 恒温箱中培养 48 h。孔板中每孔包括 $100\text{ }\mu\text{L}$ PD 培养液、 $90\text{ }\mu\text{L}$ 药液、 $10\text{ }\mu\text{L}$ 菌液, 将每个药物的初始浓度设为 $5\text{ mg}\cdot\text{mL}^{-1}$, 经 2 倍梯度稀释 9 次后最终浓度为 $9.766\text{ }\mu\text{g}\cdot\text{mL}^{-1}$ 。培养 48 h 后在黑色背景下肉眼观察, 阳性对照孔溶液清晰透亮, 阴性对照孔病菌生长良好, 此时溶液完全透明清澈的最低浓度孔中的药物浓度即 MIC。

将 96 孔板中的溶液继续 $25\text{ }^{\circ}\text{C}$ 恒温、黑暗培养 6 天, 挑取边缘菌丝制成玻片, 在光学显微镜下观察记录菌丝形态。

相对孢子数目: 使用 1 mL 药液和 19 mL PDA 培养基制成 PDA 平板, 在平板中央打取直径为 0.6 mm 的圆孔, 放入直径为 0.6 mm 的圆形菌块。 $25\text{ }^{\circ}\text{C}$ 、黑暗环境

下生长7天后,将PDA平板使用无菌水冲洗、过滤,配成10 mL孢子悬浊液。用血球计数板统计真菌孢子在对照组(DMSO)和不同药物处理组中的个数,以相同条件下DMSO培养基的孢子数目作为100%,计算相对孢子数目。相对孢子数目 = (处理组孢子数目 / 阴性对照组孢子数目) × 100%

人参根部尖孢镰刀菌回接实验:结合以上条件筛选出最佳药物,设置该候选药物的5个具有抑菌效果的浓度,将不同浓度的药物和尖孢镰刀菌孢子悬浮液混匀后接种于人参根部,使用 $2\text{ mg}\cdot\text{mL}^{-1}$ 的多菌灵作为阳性对照,DMSO作为阴性对照。将采挖后的人参去除地上部分,使用无菌水冲掉人参根部泥土,吸干根表面水分。在无菌环境下,放入75%乙醇消毒3 min,再放入5%次氯酸钠溶液中灭菌15 s,无菌水洗净后,使用无菌刀片在人参根部以十字形划伤,将10 μL 药液和20 μL 菌液混合接入到人参划伤处。每个处理重复6次,密闭黑暗、25 $^{\circ}\text{C}$ 培养25天,期间加入无菌水保持环境湿润,观察并记录病斑面积。

数据分析 使用SPSS20.0 (IBM Corp., Armonk, NY, USA)对数据进行单因素方差分析,使用Origin (version 2018, Origin Lab Corporation, Northampton, MA, USA)和Adobe Photoshop CC (version 2019, Adobe Systems Incorporated, San Jose, CA, USA)完成图像的绘制和拼接。

结果与分析

1 数据预处理

本研究将ChEMBL数据库中获取的与尖孢镰刀菌相关的668个化合物分成了380个抑制剂和288个非抑制剂,针对收集到的用于机器学习的数据集,计算每个化合物分子的AlogP和分子量,进行化学空间分析(图2)。建模数据集的AlogP范围在-2~8,分子质量范围在0~700 Da,表明基于此数据集构建的模型化学空间较大,稳定性较好。

2 机器学习模型的生成与性能评估

本研究选用了RF、SVM、ANN 3种机器学习算法对抑制尖孢镰刀菌活性的相关化合物构建了模型。基于化合物的分子描述符,利用生成的机器学习模型在训练集上的五折交叉验证和测试集结果分别见表1和表2。其中,在训练集中,RF的准确率最高,为82.96%,但是在测试集中,SVM的准确率高于RF和ANN,为83.03%。3种模型在测试集上的准确率为0.775 8~0.830 3,AUC为0.899 8~0.912 9(表2),说明模型在测试集上也表现出了优异的性能。采用3种模型的ROC曲线和混淆矩阵评估二元分类模型。3种模型的ROC

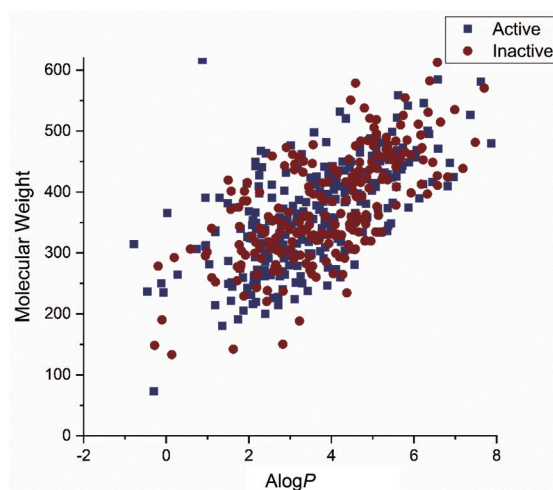


Figure 2 Two-dimensional spatial distribution of AlogP and molecular weight in the data set of compounds associated with the activity inhibition of *Fusarium oxysporum*

曲线较为接近,AUC值均在0.900左右(图3),说明RF、SVM和ANN的预测效果均较为理想。从混淆矩阵图中可以看出,模型对抑制剂(“0”)的预测效果均强于非抑制剂(“1”),其中,SVM和ANN模型的F1得分(表示准确率和召回率平衡关系)均 $\geq 85.00\%$,说明3种模型均能在一定程度上实现对抑菌活性物质的准确预测。考虑到每种算法的偏好性,以及单一模型可能会降低其泛化能力,因此将3种模型准确率平均值作为预测药物是尖孢镰刀菌抑制剂的准确率。

Table 1 Prediction results of the three machine learning models in the training set. RF: Random forest; SVM: Support vector machine; ANN: Artificial neural network

Model	Accuracy	Precision	Recall	F1 score	AUC
RF	0.829 6	0.809 7	0.798 9	0.801 0	0.900 1
SVM	0.789 0	0.762 0	0.765 9	0.758 0	0.878 9
ANN	0.821 4	0.807 7	0.780 2	0.789 7	0.886 6

Table 2 Prediction results of the three machine learning models in the test set

Model	Accuracy	Precision	Recall	F1 score	AUC
RF	0.775 8	0.745 5	0.640 6	0.689 1	0.901 3
SVM	0.830 3	0.790 3	0.765 6	0.777 8	0.912 9
ANN	0.812 1	0.770 5	0.734 4	0.752 0	0.899 8

3 基于机器学习模型的虚拟筛选及候选化合物的确定

在生成机器学习模型并评估其预测性能之后,将ZINC数据库中4 114个天然产物作为预测集,基于3种机器学习模型进行虚拟筛选。对预测集的化合物均做了同样的数据预处理、分子表征和特征工程。根据3个机器学习模型将化合物预测为抑制剂的准确率平均值进行筛选,发现178个化合物的平均准确率 $\geq 90.00\%$ 。

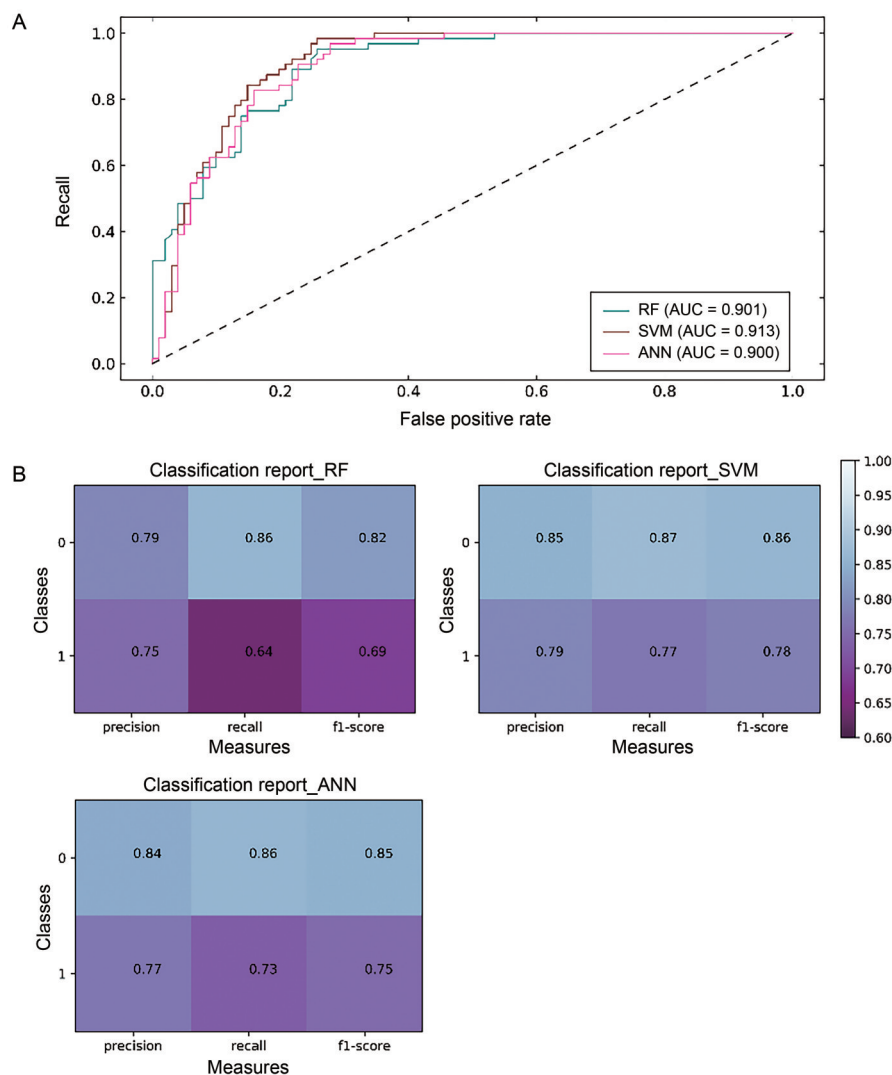


Figure 3 ROC curve (A) and confusion matrix diagram (B) of three machine learning models

在平均准确率排名靠前的20个化合物(表3, 平均准确率 $\geq 93.77\%$)中, 综合药物的提取难度、购买成本等因素, 重点选择了5种化合物宝藜苣(i-cariside II)、淫羊藜苣(i-cariin)、依托泊苣(etoposide)、硫秋水仙苣(thiocolchicoside)、芒柄花苣(ononin)进行了实验验证。

4 候选药物的实验验证

5种候选化合物的化学结构和MIC结果如表4所示, ononin的MIC值最小, 为 $0.3125 \text{ mg}\cdot\text{mL}^{-1}$, 其次为i-cariside II ($0.625 \text{ mg}\cdot\text{mL}^{-1}$), i-cariin、etoposide、thiocolchicoside的MIC值均 $> 1 \text{ mg}\cdot\text{mL}^{-1}$, 需要药物浓度较高时才能对尖孢镰刀菌起到抑制作用, 因此其抑制效果相对较弱。

根据MIC值结果, 本研究选取ononin作为最终化合物进行进一步的验证。阴性对照DMSO处理下对尖孢镰刀菌的生长没有影响, 其菌丝体正常发育(图4)。而在不同浓度ononin处理下的尖孢镰刀菌菌丝体

数量明显减少, 形态上菌丝体均断裂、变细, 甚至无菌丝生长(图4)。由此可见, ononin可以通过抑制尖孢镰刀菌菌丝的生长, 抑制尖孢镰刀菌的活性。用无菌水洗脱不同浓度ononin处理后的尖孢镰刀菌菌丝的孢子, 可见随着ononin浓度的增加, 其孢子数目显著降低(图5C), 表明ononin抑制了尖孢镰刀菌的孢子量。

为了进一步验证活体植株中ononin对尖孢镰刀菌的抑制作用, 本研究对人参根部进行了尖孢镰刀菌回接实验。发现接菌到人参根部后, 病斑面积的大小与ononin浓度的高低呈负相关(图5A、B)。接种DMSO的人参根部病变面积最大, 尖孢镰刀菌除了在划伤处侵染了人参根部, 其他位置处也发生了病变(图5A)。同时接入ononin或者多菌灵的人参根部, 均与未接入病菌抑制剂的人参病变面积有显著性差异。当药物浓度 $\geq 0.625 \text{ mg}\cdot\text{mL}^{-1}$ 时, ononin的抑菌效果强于多菌灵。

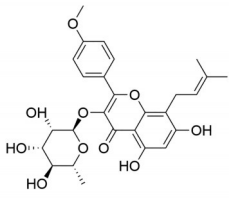
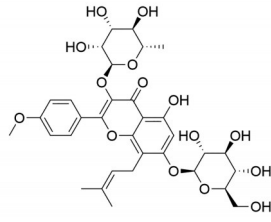
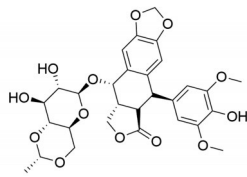
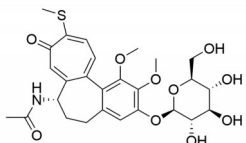
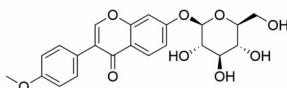
Table 3 Top 20 drugs with high average accuracy selected based on machine learning

Compound	Accuracy/%			Average accuracy/%
	RF	SVM	ANN	
Phyllanthostatin A	93.58	95.70	99.83	96.37
Icariside II	89.30	98.26	100	95.86
Icariin	91.44	94.11	100	95.18
Homaloside D	88.24	96.71	99.93	94.96
Etoposide	93.58	91.29	100	94.96
Dehydrodeguelin	87.11	96.72	100	94.61
Pierreione B	84.99	98.78	100	94.59
Lethedoside A	84.99	98.54	99.99	94.51
Tremulacin	91.44	92.92	99.05	94.47
Luzonoside D	90.91	93.03	99.46	94.47
Cypellocarpa C	87.17	96.20	99.98	94.45
Nigrolineaxanthone K	86.04	97.00	100	94.35
Pongapinone B	88.71	93.40	100	94.04
Curculigoside A	85.52	97.19	99.25	93.99
Prunose II	90.37	91.41	100	93.93
Thiocolchicoside	86.63	95.05	100	93.89
Paederoside	84.50	97.00	99.99	93.83
Ligstroside	88.24	93.39	99.71	93.78
Ononin	84.99	96.33	100	93.77
Styraxlignolide B	93.58	87.93	99.80	93.77

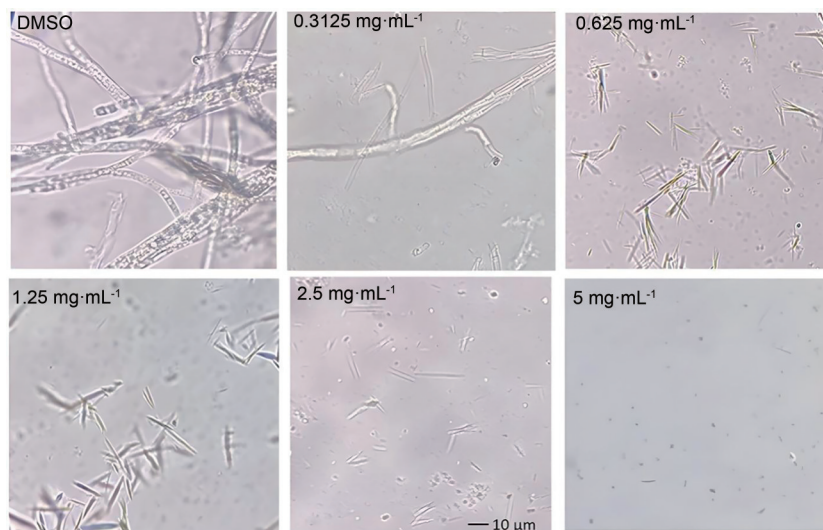
讨论

在我国, 28.9%的药用植物以地下部分入药^[32], 根部病害直接影响中药材的产量和质量^[33]。尖孢镰刀菌作为土传真菌, 侵染植物根部使其腐烂、变质^[34], 如丹参根部病害发病率高达60%~70%^[6], 大蒜根腐病使其减产20%~30%^[35], 栽培人参根腐病的发病率高达50%^[36]。基于机器学习算法的药物筛选和计算机辅助药物设计在医药领域已被广泛应用于新型冠状病毒^[37]、阿尔茨海默病、癌症^[38]等疾病防治药物开发, 且具有多个可用于药物靶点选择、建模和筛选的数据库, 但这些数据库中与农药或者植物病害相关的信息较少^[16], 限制了植物源农药的开发。本研究针对药用植物根腐病的主要

Table 4 Chemical structures and MIC values of candidate compounds

Compound	Chemical structure	MIC/mg·mL ⁻¹
Icariside II		0.625
Icariin		1.25
Etoposide		2.5
Thiocolchicoside		1.25
Ononin		0.312 5

致病菌提出抑制尖孢镰刀菌活性的天然药物筛选方法, 通过构建RF、SVM、ANN模型对天然药物分子进行预测, 缩小了候选药物范围, 简化了传统药物发现方法繁琐的工作流程, 提高了抑菌药物的筛选效率。此外, 尖孢镰刀菌的致病机制已得到广泛关注, 在香蕉中

**Figure 4** Morphology of *Fusarium oxysporum* under different ononin concentrations

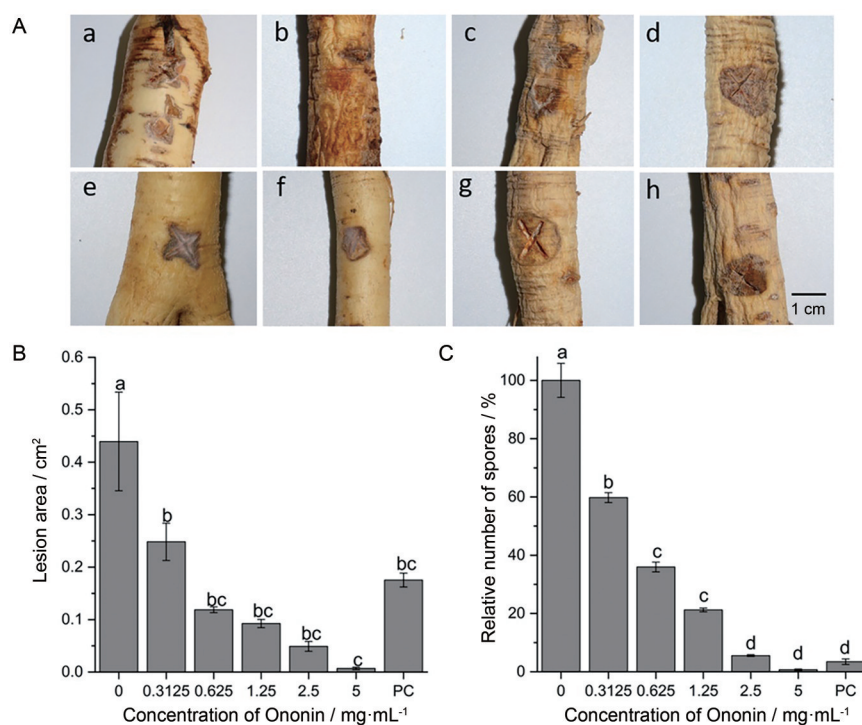


Figure 5 Root morphology (A), spot area (B) and relative spore number (C) of ginseng under different ononin concentrations. In Figure A, a and b were negative controls, c, d, e, f, g was 0.312 5, 0.625, 1.25, 2.5, 5 mg·mL⁻¹, and h were positive controls, respectively. In figure B and C, PC represents the positive control, and different lowercase letters above the bar chart indicate significant differences ($P \leq 0.05$)

发现转录因子 ZafA 可以增强尖孢镰刀菌早期的生长和致病性^[39], 在番茄中细胞壁结构的调控因子 rho1 也对真菌的侵染起到抑制作用^[40]。但是目前尖孢镰刀菌对人参、三七等药用植物的侵染机制研究报道较少^[5,41]。由于不同宿主植物感染尖孢镰刀菌的发病机制存在特异性, 因此针对药用植物的发病规律和栽培过程进行抑菌药物的筛选十分迫切。随着尖孢镰刀菌致病机制的研究不断深入, 未来基于药物靶点结合分子对接、分子动力学模型等方法的虚拟筛选也将会更为准确、有效^[42]。

本实验基于分子描述符构建了 3 种机器学习模型, 其 AUC 值均可达到 0.90, 考虑到单个模型容易导致局部偏重, 计算了 3 种模型的平均准确率并进行排序, 筛选出前 20 位的化合物作为候选化合物。综合考虑每个化合物的提取成本、获得的难易程度以及实际使用价值, 选取 icariside II、icariin、etoposide、thiocolchicoside、ononin 等 5 种候选药物的抑菌活性进行了测试。其中 icariside II 和 icariin 主要来源于中药材淫羊藿, 具有多种生物学和药理学特性, 在抗炎、抗肿瘤等方面具有明显作用^[43], 并且已有研究表明淫羊藿提取物具有明显的抑菌效果^[44,45]。Etoposide 是一种细胞周期特异性抗肿瘤药物^[46], thiocolchicoside 是一种肌肉松弛剂, 具有抗炎的作用^[47], 二者均未有抑菌活性的报

道。Ononin 是黄芪、红芪主要活性成分^[48], 具有较好的抗炎活性, 并且能够抑制癌细胞增殖^[49]。本实验中发现 ononin 可以显著抑制尖孢镰刀菌生长, 破坏其菌丝体结构, 降低孢子相对数目, 且抑菌效果随着药液浓度的增加而更加明显。将 ononin 接种到人参根部, 能够对人参根腐病的发展起到抑制作用。虽然其 MIC 值为 0.312 5 mg·mL⁻¹, 相对合成药物浓度较高, 但一方面可以对药物进行化学改造, 使其效果更为显著, 进一步发展成为农药产品; 另一方面可以考虑该药物的提取来源, 将其提取剩余的药渣及其他非药用部位等作为土壤肥料, 实现资源的可持续利用^[15], 同时又能对土壤进行改良, 抑制病菌生长。

本研究将理论计算与实验验证相结合, 筛选出针对尖孢镰刀菌的天然产物抑制剂。最终获取的候选化合物 (芒柄花苷) 有望作为设计新型尖孢镰刀菌抑制剂的先导化合物。本实验的研究结果, 一方面验证了筛选出的化合物 (芒柄花苷) 对尖孢镰刀菌活性的抑制作用; 另一方面也辅助完善了数据库中与尖孢镰刀菌活性相关的化合物的信息, 为确定农作物病虫害药物靶点和构建相应活性数据库提供数据。鉴于本研究仅对筛选结果中的 5 个化合物进行了实验验证, 存在一定的局限性。在下一步的研究中, 将在此基础上扩大筛选范围, 对准确率 > 90% 的药物均进行实验验证,

以获得效果最佳、环境友好、价格低廉的抑制尖孢镰刀菌活性的化合物,用于后续植物源农药的开发与利用。同时,将来随着药物筛选数据库中与尖孢镰刀菌抑制剂相关数据的扩增,可以进一步提高药物筛选的效率和准确性,同时,也需要进一步探究尖孢镰刀菌对植物根部侵染的发病机制和致病因子,进行更加有针对性的药物筛选和验证。

作者贡献: 赵贵萍负责研究设计、结果分析及文章撰写;李洁和陈莹莹参与实验操作与数据采集;杨若琪、于大德、李西文参与实验设计和实验技术的提供;于大德和李西文负责实验监督及论文审阅。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Dean R, Van Kan JAL, Pretorius ZA, et al. The top 10 fungal pathogens in molecular plant pathology [J]. *Mol Plant Pathol*, 2012, 13: 414-430.
- [2] Wang Y, Wang L, Suo M, et al. Regulating root fungal community using *Mortierella alpina* for *Fusarium oxysporum* resistance in *Panax ginseng* [J]. *Front Microbiol*, 2022, 13: 850917.
- [3] Thatcher LF, Kidd BN, Kazan K. Belowground Defence Strategies Against *Fusarium oxysporum* [M]. Cham: Springer International Publishing, 2016: 71-98.
- [4] Huo YY, Li TT, Yang J, et al. Pathogenic factors related to *Fusarium oxysporum* and potential of essential oils to inhibit *Fusarium oxysporum*: a review [J]. *Agrochemicals (农药)*, 2022, 61: 79-86.
- [5] Wang R, Dong LL, Xu J, et al. Identification and prevention of root rot pathogen in model of ginseng cultivated in farmlands [J]. *China J Chin Mater Med (中国中药杂志)*, 2016, 41: 1787-1791.
- [6] Pu CJ, Liu S, Lu ZY, et al. Pathogenicity and induced systemic resistance of *Fusarium oxysporum* and *Verticillium dahlia* to *Salvia miltiorrhiza* [J]. *China J Chin Mater Med (中国中药杂志)*, 2022, 47: 5832-5837.
- [7] Ma YN, Chen CJ, Ma XH, et al. Influence of methanolic extract of *Artemisia annua* leaves on *Fusarium oxysporum* [J]. *Mod Chin Med (中国现代中药)*, 2018, 20: 675-679.
- [8] Wang R, Liang X, Long Z, et al. An LCI-like protein APC2 protects ginseng root from *Fusarium solani* infection [J]. *J Appl Microbiol*, 2021, 130: 165-178.
- [9] Wainwright CL, Teixeira MM, Adelson DL, et al. Future directions for the discovery of natural product-derived immunomodulating drugs: an IUPHAR positional review [J]. *Pharmacol Res*, 2022, 177: 106076.
- [10] Zhang L, Song J, Kong L, et al. The strategies and techniques of drug discovery from natural products [J]. *Pharmacol Ther (Oxford)*, 2020, 216: 107686.
- [11] Sparks TC, Hahn DR, Garizi NV. Natural products, their derivatives, mimics and synthetic equivalents: role in agrochemical discovery [J]. *Pest Manag Sci*, 2017, 73: 700-715.
- [12] Soliman SA, Hafez EE, Al-Kolaibe AMG, et al. Biochemical characterization, antifungal activity, and relative gene expression of two *Mentha* essential oils controlling *Fusarium oxysporum*, the causal agent of *Lycopersicon esculentum* root rot [J]. *Plants*, 2022, 11: 189.
- [13] Papatoti NK, Saengchan C, Daddam JR, et al. Plant systemic acquired resistance compound salicylic acid as a potent inhibitor against SCF (SKP1-CUL1-F-box protein) mediated complex in *Fusarium oxysporum* by homology modeling and molecular dynamics simulations [J]. *J Biomol Struct Dyn*, 2022, 40: 1472-1479.
- [14] Liu L, Wang X, Zhang PG, et al. Inhibition of total flavonoids in *Cirsium japonicum* DC on growth physiological indexes and control effect of *Fusarium oxysporum* f. sp. *melonis* in field [J]. *J Plant Protec (植物保护学报)*, 2020, 47: 628-636.
- [15] Li Y, Du CM. Virulence factors of pathogenic *Fusarium oxysporum*: research progress [J]. *Chin Agric Sci Bull (中国农学通报)*, 2021, 37: 92-97.
- [16] Wang JL, Zhang Y, Liu SQ, et al. Application of computer-aided drug design in the development of agrochemical compounds [J]. *Agrochemicals (农药)*, 2022, 61: 242-249.
- [17] Gupta R, Srivastava D, Sahu M, et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery [J]. *Mol Divers*, 2021, 25: 1315-1360.
- [18] Yang GF. Chemical biology-oriented molecular design of green pesticide [J]. *Bull Natl Nat Sci Found China (中国科学基金)*, 2020, 34: 495-501.
- [19] Tsou LK, Yeh SH, Ueng SH, et al. Comparative study between deep learning and QSAR classifications for TNBC inhibitors and novel GPCR agonist discovery [J]. *Sci Rep*, 2020, 10: 16771.
- [20] Jain P, Garibaldi JM, Hirst JD. Supervised machine learning algorithms for protein structure classification [J]. *Comput Biol Chem*, 2009, 33: 216-223.
- [21] Svetnik V, Liaw A, Tong C, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling [J]. *J Chem Inf Comput Sci*, 2003, 43: 1947-1958.
- [22] Mavroforakis ME, Theodoridis S. A geometric approach to support vector machine (SVM) classification [J]. *IEEE Transact Neural Networks*, 2006, 17: 671-682.
- [23] Jing Y, Bian Y, Hu Z, et al. Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era [J]. *AAPS J*, 2018, 20: 58.
- [24] Zhao X, Li H, Mo QL, et al. Toxicity prediction of pesticide to bumblebee and honey bee based on machine learning methods [J]. *Chin J Pest Sci (农药学报)*, 2020, 22: 933-941.
- [25] Sakthivel S, Habeeb S, Raman C. Screening of broad spectrum natural pesticides against conserved target arginine kinase in

- cotton pests by molecular modeling [J]. *J Biomol Struct Dyn*, 2019, 37: 1022-1042.
- [26] Islam MS, Mahmud S, Sultana R, et al. Identification and *in silico* molecular modelling study of newly isolated *Bacillus subtilis* SI-18 strain against S9 protein of *Rhizoctonia solani* [J]. *Arab J Chem*, 2020, 13: 8600-8612.
- [27] Overington J. ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr [J]. *J Comput Aided Mol Des*, 2009, 23: 195-198.
- [28] Wei G, Fu X, Wang Z. Nature identification of Chinese herbal medicine compounds based on molecular descriptors [J]. *JAOAC Int*, 2021, 104: 1754-1759.
- [29] Abraham A, Pedregosa F, Eickenberg M, et al. Machine learning for neuroimaging with scikit-learn [J]. *Front Neuroinform*, 2014, 8: 14.
- [30] Isa SM, Suwandi R, Pricilia Y. Optimizing the hyperparameter of feature extraction and machine learning classification algorithms [J]. *Int J Adv Comput Sci Appl*, 2019, 10: 69-76.
- [31] Xie XM, Xu Y, Fu YY. *In vitro* anti-aspergilli activity study of active ingredients of Chinese herbs by broth microdilution testing [J]. *Microbiology (微生物学通报)*, 2003, 30: 89-92.
- [32] Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China (中华人民共和国药典) [S]. Beijing: China Medical Science Press, 2020.
- [33] Zhao G, Pei Y, Yang R, et al. A non-destructive testing method for early detection of ginseng root diseases using machine learning technologies based on leaf hyperspectral reflectance [J]. *Front Plant Sci*, 2022, 13: 1031030..
- [34] Perveen K, Bukhari NA, Al Masoudi LM, et al. Antifungal potential, chemical composition of *Chlorella vulgaris* and SEM analysis of morphological changes in *Fusarium oxysporum* [J]. *Saudi J Biol Sci*, 2022, 29: 2501-2505.
- [35] Zhou J, Qi CD, Wu JP, et al. Biological characteristics and laboratory efficacy test for pathogen (*Fusarium oxysporum*) of garlic root rot in Enshi [J]. *J Yangtze Univ (Nat Sci) (长江大学学报 (自然科学版))*, 2022, 19: 111-116.
- [36] Wang Y, Wang CW, Gao J, et al. Toxicity test and field control effects of 24 fungicides and related proportions against *Fusarium solani*[J]. *Agrochemicals (农药)*, 2014, 53: 61-65.
- [37] Yang Y, Zhou D, Zhang X, et al. D3AI-CoV: a deep learning platform for predicting drug targets and for virtual screening against COVID-19 [J]. *Brief Bioinform*, 2022, 23: bbac147.
- [38] Urbina F, Puhl AC, Ekins S. Recent advances in drug repurposing using machine learning [J]. *Curr Opin Chem Biol*, 2021, 65: 74-84.
- [39] Lopez-Berges MS. ZafA-mediated regulation of zinc homeostasis is required for virulence in the plant pathogen *Fusarium oxysporum* [J]. *Mol Plant Pathol*, 2020, 21: 244-249.
- [40] Martinez-Rocha AL, Roncero MI, Lopez-Ramirez A, et al. Rho1 has distinct functions in morphogenesis, cell wall biosynthesis and virulence of *Fusarium oxysporum* [J]. *Cell Microbiol*, 2008, 10: 1339-1351.
- [41] Liu YL, Fan ZY, Xie SN, et al. The antifungal activity and control effect of tobacco cembranoid diterpenes against root rot pathogens of *Panax notoginseng* [J]. *Agrochemicals (农药)*, 2022, 61: 374-377.
- [42] Chen H, Ma Y, Zhou Y, et al. Pharmacophore identification, virtual screening and activity verification of pedunsaponin A on target proteins PcAdv and PenWAS of *Pomacea canaliculata* [J]. *Pest Manag Sci*, 2022, 78: 3098-3107.
- [43] Yan ZK, Qiu H. Antimicrobial tests of icariin [J]. *Chin Food Add (中国食品添加剂)*, 2005, 4: 65-67.
- [44] Liu N, Lin Y, Han LY, et al. A study on the antibacterial and osteogenic properties of icariin combined with minocycline hydrochloride [J]. *J Pract Stomatol (实用口腔医学杂志)*, 2022, 38: 595-600.
- [45] Zhang ZW, Zhang NN, Cui C, et al. *In vitro* antibacterial test on the different extracts from *Epimedium brevicornu* Maxim [J]. *Chin Pharm (中国药师)*, 2016, 19: 2247-2250.
- [46] Ohmoto A, Fujii S. Clinical feasibility of oral low-dose etoposide and sobuzoxane for conventional chemotherapy-intolerant lymphoma patients [J]. *Expert Rev Anticancer Ther*, 2021, 21: 715-722.
- [47] Mancuso G. Immediate and delayed contact hypersensitivity to thiocolchicoside [J]. *Acta Dermatovenerol Alp Pannonica Adriat*, 2018, 27: 203-205.
- [48] Dai YL, Zhang XY, Wei YX, et al. Research progress on Astragali Radix and prediction of its quality markers (Q-markers) [J]. *China J Chin Mater Med (中国中药杂志)*, 2022, 47: 1754-1764.
- [49] Huang QY, Lin QZ, Wu XF, et al. Study on the mechanism of ononin in inhibiting breast cancer growth by down-regulating cyclin D [J]. *Chin J Clin Pharm (中国临床药理学杂志)*, 2022, 38: 1999-2003.