

基于地理栅格变量与机器学习的松材线虫病扩散风险分析

刘婷婷¹, 杨晋帆¹, 周汝良¹, 刘琳²

(1. 西南林业大学 地理与生态旅游学院, 云南 昆明 650224; 2. 西南林业大学 理学院, 云南 昆明 650224)

摘要: 【目的】松材线虫 *Bursaphelenchus xylophilus* 病是威胁中国森林生态系统安全最为严重的病害。本研究利用地理栅格模型及地图代数运算来模拟和表达驱动变量, 构建松材线虫病测报系统, 并以县域与地理栅格单元双尺度形成空间连续化测报。【方法】集成对松材线虫病扩散传播有影响的地形地貌、气象、寄主、人类活动和土地利用等地理栅格空间数据集, 基于随机森林和支持向量机的机器学习方法构建模型, 将测报得出的扩散风险概率与松科 Pinaceae 植物易感性叠加进行感染概率地图代数运算, 基于地理栅格表达分析松材线虫病在中国范围内扩散的风险等级。【结果】①随机森林模型预测精度为 83.95%, 支持向量机模型预测精度为 77.97%。②海拔、年均最低降水量、年均降水量、年均低温对模型构建的贡献率分别为 0.151、0.303、0.258、0.194, 是影响松材线虫病发生的主要因素; 人类活动变量的贡献率为 0.194, 是影响扩散的决定性变量。③潜在扩散区位于人类活动密集的低海拔地区、道路通达的林区、城市城镇分布区和人工林分布区。其中, 极高风险分布地区主要位于华东地区的浙江、江西、福建, 华南地区的广西、广东, 华中地区的湖南。【结论】利用空间模拟与机器学习方法构建了松材线虫病空间化测报模型, 将松材线虫病的扩散风险测报到地理栅格单元, 可为中国林草灾害精准监测提供方法借鉴, 对中国松材线虫病的疫情防控攻坚行动具有重要指导意义。图 2 表 7 参 36

关键词: 松材线虫病; 地理栅格; 机器学习; 风险分析

中图分类号: S763.105 文献标志码: A 文章编号: 2095-0756(2023)03-0617-10

Prediction of spread risk of pine wilt disease based on geographic raster variables and machine learning models

LIU Tingting¹, YANG Jinfan¹, ZHOU Ruliang¹, LIU Lin²

(1. College of Geography and Ecotourism, Southwest Forestry University, Kunming 650224, Yunnan, China; 2. College of Science, Southwest Forestry University, Kunming 650224, Yunnan, China)

Abstract: [Objective] Considering the fact that pine wilt disease (PWD) has been the most serious disease threatening forest ecosystem in China. This study, with the simulation and expression of the driving variables by using geographic grid model and map algebra operation, is aimed to construct PWD measurement and forecasting system and form the spatial continuous measurement and prediction by using the dual scale of county area and geographic grid unit. [Method] First, datasets were formed by integrating geographic raster spatial data such as topography, meteorology, host, human activities and land use that affect PWD dispersal. Then, the model was constructed using the machine learning method of random forest and support vector machine before predicted results were superimposed with the susceptibility map of Pinaceae plants to perform the map algebraic operation of infection probability. Finally, an analysis was conducted of the risk level of

收稿日期: 2022-07-19; 修回日期: 2023-01-12

基金项目: 国家自然科学基金资助项目 (31760212); 云南省教育厅基金项目 (2021J0154); 云南省科技厅重大科技专项 (202002AA100007)

作者简介: 刘婷婷 (ORCID: 0000-0003-4377-1289), 从事生物灾害预警研究。E-mail: 1693931917@qq.com。通信作者: 周汝良 (ORCID: 0000-0001-6740-3873), 教授, 从事自然灾害及森林火灾的监测预警研究。E-mail: zhou_ruliang@163.com

PWD spread in the whole country based on geographic grid unit. [Result] (1) The prediction accuracy was 83.95% for the random forest model and 77.97% for the support vector machine model; (2) Altitude, average annual minimum precipitation, average annual precipitation and average annual low temperature were the main factors affecting the occurrence of PWD, with their contribution rates to the model construction being 0.151, 0.303, 0.258 and 0.194 respectively whereas human activity variables were the decisive variable affecting the diffusion of PWD with their contribution rate to the model construction being 0.194; (3) The potential dispersal areas were located in low altitude areas with dense human activities, forest areas adjacent to roads, urban distribution areas and plantation distribution areas while the highest risk areas were mainly Zhejiang, Jiangxi and Fujian in East China, Guangxi and Guangdong in South China as well as Hunan in central China. [Conclusion] With the employment of spatial simulation and machine learning methods, a mapping model was constructed to predict the spatial transmission pattern of PWD with diffusion risk of PWD predicted of specific geographical grid cells. It provides a reference for the accurate supervision of forest and grass disasters with practical guiding significance for the prevention and control of PWD epidemic in China. [Ch, 2 fig. 7 tab. 36 ref.]
Key words: pine wilt disease; geographical raster; machine learning; risk prediction

由松材线虫 *Bursaphelenchus xylophilus* 引起的松材线虫病是目前松树上最具破坏性的病害之一。自1905年日本首次报道发生松材线虫病以来, 亚欧其他国家相继出现松材线虫病疫区^[1-2]。中国的松材线虫病最早报道于1982年, 出现在南京中山陵的黑松 *Pinus thunbergii* 上, 之后该病害快速扩散, 据国家林业和草原局统计, 中国现共有松材线虫病县级疫区726个、乡镇级疫点5479个, 发生面积达180.9万 hm^2 , 包括轻型疫区206个(含2021年初新增疫情发生区)、重型疫区518个^[3-4]。目前松材线虫病已成为中国最为严重的林业病害之一^[5-6]。为实现《全国松材线虫病疫情防控五年攻坚行动计划(2021—2025)》中新发疫情“早发现、早报告、早除治、早拔除”的目的, 达到控制增量的总体目标, 就需要把握疫情的总体趋势, 精准定位和定量各疫区的疫情, 提高防疫针对性^[7]。

已有研究发现: 全球变暖环境下, 气温增高会影响松材线虫病的地理分布^[8]; 人类活动是影响松材线虫病远距离、跳跃式、梅花式扩散传播的重要因素^[9-11]; 年均气温、年均降水量等是影响松材线虫定殖的主要气候因素^[12]; 中国松材线虫病的发生受到寄主植物分布的强烈影响^[13-14]; 媒介昆虫对寄主的选择性也影响松材线虫病的发生^[15-17]。当前对松材线虫病传播风险的研究多集中于采用传统的地统计学方法来分析松材线虫病在省、市范围的风险格局^[18-20], 而将机器学习引入病害风险测报研究较少见。机器学习方法中随机森林、支持向量机具有需要样本量少, 能够同时利用多变量进行非线性回归、分类预测并且结果精度高等优点, 使风险预测精确化分析成为可能^[21-23]。

本研究综合考虑松材线虫病扩散的影响因素, 利用地理环境、气候环境、卫星遥感、道路网络等构成地理栅格数据集。对比随机森林和支持向量机2种算法建立松材线虫病风险扩散模型, 探究机器学习算法在松材线虫病扩散传播预测研究中的可行性和准确性, 寻求最佳预测方案, 将潜在疫区根据等级进行风险区划分, 并进行栅格化分析, 将全国松材线虫病测报落实到“山头地块”上, 以期为林业生态系统中松科 Pinaceae 植物的动态监测提供有效的手段。

1 数据与方法

1.1 疫区样本数据

以中国县区为行政单元, 以国家林业和草原局2018—2021年关于松材线虫病疫区公告为基础^[4], 收集整理了中国松材线虫病发生区731个以及未发生区2131个, 以松材线虫病发生(1)或未发生(0)作为模型响应变量, 由点及面构建了2862个样点。其中随机选取30%样本点(954个)作为模型独立检验样本。总样本点分布见表1。

1.2 测报变量空间模拟

1.2.1 气候环境数据 根据以往研究^[8, 12], 选取1970—2000年共30a逐月气象数据获得年均气温、年均

降水量以及年均最高气温、年均最低气温、年均最低降水量与年均最高降水量数据。空间分辨率为 250 m 的气候数据集来源于 <https://www.worldclim.org/>。

1.2.2 传播媒介数据 选取影响昆虫媒介墨天牛属 *Monochamus* 的地理因子，包括高程数据，以及根据高程数据计算提取的坡度与坡向指数。选取海拔、坡度、植被覆盖度以及温度构建阻力面，表征其自然传播的能力。根据全国高速公路、国道、省道以及一级水系等交通网络，利用欧氏距离来模拟人类活动导致的有害生物空间传播度量^[9]。道路、水系矢量数据等人为影响数据来源于 Bigemap 平台，利用 ArcGIS 处理形成综合交通便利性数据；土地利用类型数据来源于 GLOBELAND 30 平台。

1.2.3 寄主植物数据 全国 1:100 万植被分布数据来源于资源环境数据云平台 <http://www.Resdc.cn/>，从中提取出易受松材线虫病感染的针叶林植被。根据前人研究结果^[24]将不同树种的寄主易感指数赋分为 0~100，数值越高表示寄主易感性越高，划分标准见表 2。

表 1 松材线虫病疫区

Table 1 Pine wilt disease endemic area

省级行政区	县域总数/个	疫区数/个	占比/%	省级行政区	县域总数/个	疫区数/个	占比/%
重庆	38	35	92.1	陕西	107	25	23.4
江西	100	85	85.0	四川	183	42	23.0
浙江	89	71	79.8	辽宁	100	20	20.0
湖北	103	77	74.8	贵州	88	16	18.2
福建	85	57	67.1	山东	137	24	17.5
湖南	122	76	62.3	天津	16	1	6.2
广东	125	76	60.8	河南	158	9	5.7
安徽	105	50	47.6	吉林	60	2	3.3
广西	111	39	35.1	云南	129	2	1.5
江苏	97	23	23.7	甘肃	87	1	1.1

表 2 松材线虫病寄主量化

Table 2 Quantification of pine wilt disease hosts

科属	植被类型	易感指数
松科 Pinaceae 松属 <i>Pinus</i>	赤松 <i>P. densiflora</i> , 黑松 <i>P. thunbergii</i> , 马尾松 <i>P. massoniana</i> , 思茅松 <i>P. kesiya</i> var. <i>langbianensis</i>	100
松科松属	云南松 <i>P. yunnanensis</i> , 华山松 <i>P. armandii</i> , 油松 <i>P. tabulaeformis</i> , 高山松 <i>P. densata</i> , 台湾松 <i>P. taiwanensis</i>	80
松科松属	白皮松 <i>P. bungeana</i> , 巴山松 <i>P. henryi</i> , 长叶松 <i>P. palustris</i> , 乔松 <i>P. wallichiana</i> , 海南五针松 <i>P. fenzeliana</i> , 偃松 <i>P. pumila</i> , 樟子松 <i>P. sylvestris</i> var. <i>mongolica</i> , 西伯利亚红松 <i>P. sibirica</i>	70
松科落叶松属 <i>Larix</i>	落叶松 <i>L. gmelinii</i> , 新疆落叶松 <i>L. sibirica</i> , 黄花落叶松 <i>L. olgensis</i> , 日本落叶松 <i>L. kaempferi</i> , 华北落叶松 <i>L. gmelinii</i> var. <i>principis-rupprechtii</i> , 秦岭红杉 <i>L. potaninii</i> var. <i>chinensis</i> , 红杉 <i>L. potaninii</i> , 大果红杉 <i>L. var. macrocarpa</i>	60
松科冷杉属 <i>Abies</i>	冷杉 <i>A. firma</i> , 长苞冷杉 <i>A. georgei</i> , 川滇冷杉 <i>A. forrestii</i> , 臭冷杉 <i>A. nephrolepis</i> , 台湾冷杉 <i>A. kawakamii</i> , 巴山冷杉 <i>A. fargesii</i> , 紫果冷杉 <i>A. recurvata</i> , 鳞皮冷杉 <i>A. squamata</i> , 苍山冷杉 <i>A. delavayi</i> , 墨脱冷杉 <i>A. delavayi</i> var. <i>motuoensis</i> , 西藏冷杉 <i>A. spectabilis</i>	50
柏科 Cupressaceae	柏木 <i>Cupressus funebris</i> , 杉木 <i>Cunninghamia lanceolata</i> , 侧柏 <i>Platycladus orientalis</i> , 杜松 <i>Juniperus rigida</i> , 祁连圆柏 <i>J. przewalskii</i> , 塔枝圆柏 <i>J. komarovii</i> , 玉山圆柏 <i>J. morrisonicola</i> , 大果圆柏 <i>J. tibetica</i> , 方枝柏 <i>J. saltuaria</i> , 密枝圆柏 <i>J. convallium</i>	40
松科云杉属 <i>Picea</i> 和铁杉属 <i>Tsuga</i>	云杉 <i>P. asperata</i> , 鱼鳞云杉 <i>P. jezoensis</i> , 白扦 <i>P. meyeri</i> , 青扦 <i>P. wilsonii</i> , 青海云杉 <i>P. crassifolia</i> , 雪岭杉 <i>P. schrenkiana</i> , 紫果云杉 <i>P. purpurea</i> , 麦吊云杉 <i>P. brachytyla</i> , 丽江云杉 <i>P. likiangensis</i> , 川西云杉 <i>P. likiangensis</i> var. <i>rubescens</i> , 林芝云杉 <i>P. linzhiensis</i> , 台湾云杉 <i>P. morrisonicola</i> , 台湾铁杉 <i>T. formosana</i> , 云南铁杉 <i>T. dumosa</i>	20
	针阔混交林	10
	其他非植被类型	0

综上，选择 16 个环境变量因子作为主要环境数据 (表 3)。环境数据集的空间分辨率通过重采样统一为 250 m。地理数据为 1:320 万的中国国界、省界以及县界行政区划图，来自于自然资源部标准地图服务网 <http://bzdt.ch.mnr.gov.cn/>。

2 扩散风险模型构建

2.1 模型

随机森林 (Random Forest Classifier, RF) 是一种监督学习的机器学习算法，即将多个决策树分类器

表3 松材线虫病预测模型自变量表

Table 3 Independent variables of the prediction model of pine wilt disease

变量类型	变量代号	变量名称	变量类型	变量代号	变量名称
生物气候变量	Bio1	年平均气温	人为扩散因子	Wspct	水系影响力
	Bio2	年均高温		Hwpct	高速公路影响力
	Bio3	年均低温		Nrpct	国道影响力
	Bio4	年均降水量		Pvrpct	省道影响力
	Bio5	年均最高降水量		Rawpct	铁路影响力
	Bio6	年均最低降水量			
生物类因子	Hst	寄主分布	地理环境因子	Ele	海拔
	LU	土地利用		Slp	坡度
				Asp	坡向指数

集合在一起进行数据的分类和回归^[25-27]。模型从原始样本抽取的随机性, 以及利用小于样本集特征的数据建立训练集, 增加了不同决策树结果的差异性, 中和了过拟合现象, 使得模型的泛化能力提高使其具有较高的分类准确性^[28-29]。支持向量机(Support Vector Machine, SVM)算法能够解决机器学习中样本量小、非线性和高维模式识别中“维数灾难”和“过学习”等方面的问题^[30-31]。该方法选择适当的核函数通过非线性变换将输入空间变换到一个高维空间, 然后在这个高维空间求取最优分类面, 找到输入变量和输出变量之间的一种非线性关系。

2.2 精度评价

ROC-AUC 特征曲线分析方法用来评价二值分类器的优劣^[32]。ROC 曲线反映了输出概率的准确性, 曲线上每个点反映对同一信号刺激的感受性。横轴为负正类率(false positive rate, FPR) 特异度, 纵轴为真正类率(true positive rate, TPR), 灵敏度均为 [0, 1]。ROC 曲线在斜对角线以下, 表示该分类器效果差于随机分类器, 反之, 效果好于随机分类器^[33]。ROC 曲线下的面积(area under curve, AUC) 作为数值可以直观地评价分类器的好坏。

2.3 扩散风险模拟

RF 和 SVM 模型均以松材线虫病在全国县域尺度中存在(1) 或不存在(0) 作为响应变量, 对所有环境变量在输入前进行归一化处理, 以消除由于纲量导致的差异。数据集以 7:3 的比例随机分成两部分进行模型训练和测试。在建模过程中 16 个环境变量根据模型重要性及其影响, 筛除部分低贡献度变量, 并将针叶林分布变量提出作为重要变量与预测风险区叠加分析。模型预测精度和 ROC-AUC 曲线用来评价每个模型的性能, 利用预测结果与独立检验样本通过混淆矩阵进行精确性检验。通过 Python 转为 0~1 之间的发生概率, 输入 ArcGIS 中进行风险可视化处理。除此之外, 全国栅格尺度更能够体现空间连续变化, 在进行运算操作、叠加分析方面也比县级矢量数据更具优势^[34]。利用从全国 1:100 万植被分布类型中提取出来的寄主松科植物, 根据松科植物不同的易感性重分类, 栅格化后与模型预测的松材线虫病潜在发生区域叠加分析。

3 结果与分析

3.1 模型精度验证

通过模型精度和 ROC-AUC 检验结果(表 4) 可知: 2 个模型均有较好的预测能力, 其预测精度和模型稳定性均在 75% 以上。模型精度是通过 RF 和 SVM 模型对独立检验样本数据进行松材线虫病的发生风险概率值预测, 设定临界值为 0.5, 将大于此值县域视为松材线虫病“存在”(y=1), 小于此值归为“不存在”(y=0), 得出分类结果与独立检验样本的实际值来验证模型预测的精度。结果表明: 构建模型的测试集预测精度在 RF 和 SVM 模型中总精度分别为 83.95% 和 77.97%。

RF 和 SVM 模型的 ROC-AUC 曲线值(图 1) 分别为 0.89 和 0.77 (平均标准误差为 0.01)。这证明 2 个模型在预测松材线虫病县级风险分布过程中具有逻辑可行性, 预测结果具有可信性, 但 RF 的效果、模型稳定性、模型分类结果泛化能力均高于 SVM 模型。

表 4 模型精度验证数据表

Table 4 Model accuracy verification data sheet

观测 y	RF 预测y		正确率/%	总体百分比/%	SVM 预测y		正确率/%	总体百分比/%
	0	1			0	1		
0	384	59	86.68	83.95	360	106	77.25	77.97
1	96	415	81.21		104	384	78.68	

3.2 变量重要性评价

根据 RF 模型对数据贡献度 (图 2) 结果得出, 重要性排序中位于前列的为气候变量, 包括年均最低降水量 (0.302 59)、年均降水量 (0.170 33)、年均低温 (0.102 98); 其次是海拔变量 (0.150 84), 这 2 类是影响松材线虫病的重要环境变量。松材线虫病发生概率在低海拔地区较高, 随着年均低温的升高风险概率增加, 而高温、干旱、低湿气候直接影响松材线虫病的爆发。在传播媒介影响中, 人为因素中各级交通网络影响力重要性更为突出 (0.067 58), 说明密集的人流和物流是松材线虫病远距离传播的重要媒介。在重要性排序中松科植物类型数据贡献度与实际影响松材线虫病的情况不符, 寄主是影响松材线虫病存在和发展的必然条件, 故将针叶林分布单独提出栅格化, 与预测风险区叠加分析其影响。

3.3 松材线虫病县域风险测报

利用 Python 将预测结果提取为 0~1 之间的概率数据, 依据离散程度划分为 5 个等级, 由高到低依次为极高风险、高风险、中等风险、低风险、极低风险。因样本原因, 不涉及中国香港、台湾、澳门行政区。对比 RF 与 SVM 模型的分析结果 (表 5 和表 6), SVM 对重要变量的模糊表达导致其预测结果等级差距小、重点疫区不突出、预测精确性低、全国大部分地区呈同一等级, 这使得对松材线虫病疫区的监测与防治难度增大。相比而言, RF 预测结果具有更高的准确性, 其潜在疫区分布与原有疫区重合度高, 危险等级一致, 并能够明显表达城市、道路以及地形对潜在疫区分布的影响。

3.4 全国栅格尺度松材线虫病风险测报

寄主植物根据其易感性划分等级, 松材线虫病易感性松科植物多分布于浙江、福建、湖南、广西等省, 另外, 云南、贵州、四川以及东北林区也有大量寄主植物分布。利用 RF 模型叠加寄主易感性分布数据对全国 31 个省进行地理栅格尺度的疫区划分与分析, 其中概率大于 0.8 以上的栅格点为极易感染区域。结果 (表 7) 表明: 松材线虫病的极高风险区为华东地区的浙江、江西、福建, 华南地区的广西、广东以及华中地区的湖南, 高易感性寄主分布与极高风险区预测结果高度重合, 疫区内多为赤松、黑松、马尾松; 高风险区包括华东地区的安徽, 华中地区的湖北, 以及西南地区的重庆、贵州, 高风险区寄主也呈高易感性, 但其分布分散且多于其他植被混生, 森林具有类型多样化; 中等风险区包括华中的河南、江苏, 东北地区辽宁, 华东地区山东, 寄主植物易感性降低, 多呈中等易感性且面积小, 多为散布状态, 其中很多省份为粮食大省, 以种植农作物为主, 林业分布较少; 无风险区包括西北地区甘肃、宁夏、青海、新疆, 华北地区的内蒙古、山西和西南地区的西藏, 缺少寄主植物的分布, 人流、物流相较于其他地区也较少, 难以形成松材线虫病潜在风险区; 其余省份划分为低风险区, 其中需注意的是云

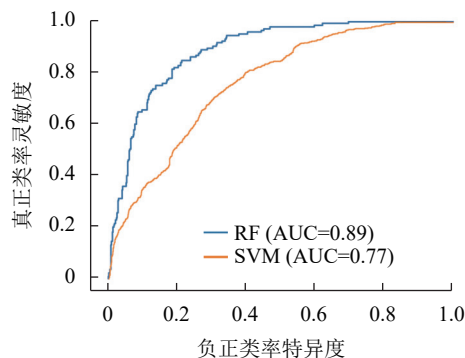


图 1 ROC-AUC 曲线

Figure 1 ROC-AUC curve

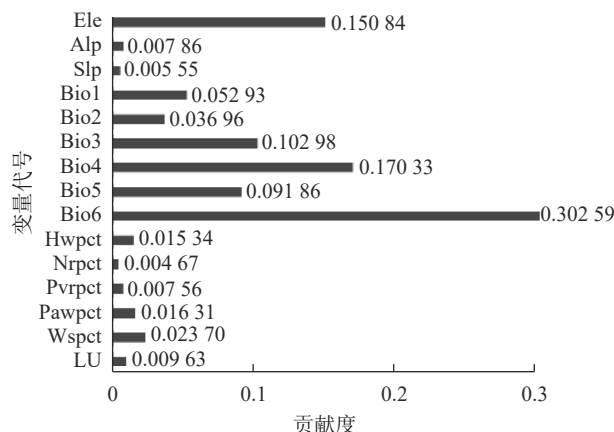


图 2 RF 模型变量贡献度

Figure 2 Contribution of RF model variables

表5 随机模型预测松材线虫病县域潜在风险等级

Table 5 Potential risk levels of pine wilt disease in counties based on RF

省级行政区	县域数	不同等级风险区占比/%					省级行政区	县域数	不同等级风险区占比/%				
		极高	高	中	低	极低			极高	高	中	低	极低
北京	16	12.50	18.75	12.50	25.00	31.25	湖南	122	82.79	13.93	2.46	0.82	0
天津	16	0	6.25	25.00	43.75	25.00	广东	125	82.40	5.60	1.60	0.80	9.60
河北	168	4.76	7.74	27.38	29.76	30.36	海南	28	50.00	21.42	14.29	0	14.29
山西	117	0	0	0.85	3.42	95.73	重庆	38	36.84	39.48	18.42	5.26	0
辽宁	100	8.00	37.00	23.00	25.00	7.00	四川	183	7.10	13.66	20.22	14.76	44.26
吉林	60	0	10.00	41.67	35.00	13.33	贵州	88	6.82	13.64	30.68	40.91	7.95
黑龙江	133	0	6.77	10.52	36.09	46.62	云南	129	0	0.78	12.40	27.13	59.69
上海	16	12.50	25.00	18.75	6.25	37.50	陕西	107	0	7.48	10.28	23.36	58.88
江苏	97	27.84	38.14	10.31	3.09	20.62	甘肃	87	0	0	0	5.75	94.25
浙江	89	74.16	20.22	2.25	1.12	2.25	青海	45	0	0	0	6.67	93.33
安徽	105	67.62	16.19	8.57	0	7.62	宁夏	22	0	0	0	4.55	95.45
福建	85	63.53	21.18	11.76	2.35	1.18	新疆	105	0	0	0	100.00	0
江西	100	95.00	5.00	0	0	0	内蒙古	103	0	0	0	13.59	86.41
山东	137	27.00	18.98	23.36	22.63	8.03	广西	111	78.38	9.91	8.11	2.70	0.90
河南	158	13.93	37.97	22.78	13.93	11.39	西藏	74	0	0	0	0	100.00
湖北	103	65.05	17.48	10.68	2.91	3.88							

表6 SVM 松材线虫病县域潜在风险等级

Table 6 SVM potential risk levels of pine wilt disease in counties

省级行政区	县域数	不同等级风险区占比/%					省级行政区	县域数	不同等级风险区占比/%				
		极高	高	中	低	极低			极高	高	中	低	极低
北京	16	0	0	56.00	44.00	0	湖南	122	7.38	20.49	50.82	21.31	0
天津	16	0	0	13.00	68.00	19.00	广东	125	8.00	28.80	48.80	13.60	0.80
河北	168	0	0.60	48.80	44.05	6.55	海南	28	3.57	14.29	42.85	25.00	14.29
山西	117	0	0	13.68	86.32	0	重庆	38	2.63	2.63	94.74	0	0
辽宁	100	2.00	12.00	72.00	14.00	0	四川	183	0	1.09	50.82	45.36	2.73
吉林	60	0	0	78.00	22.00	0	贵州	88	0	0	40.91	59.09	0
黑龙江	133	0	2.26	51.88	45.86	0	云南	129	0	0	19.38	71.32	9.30
上海	16	0	0	81.25	18.75	0	陕西	107	0	0	32.71	66.36	0.93
江苏	97	0	7.22	82.47	10.31	0	甘肃	87	0	0	9.19	83.91	6.90
浙江	89	10.11	13.48	68.54	7.87	0	青海	45	0	0	4.44	75.56	20.00
安徽	105	5.71	18.10	62.86	13.33	0	宁夏	22	0	0	9.09	81.82	9.09
福建	85	7.06	15.29	63.53	14.12	0	新疆	105	0	0	18.00	80.00	2.00
江西	100	10.00	30.00	48.00	12.00	0	内蒙古	103	0	0.97	32.04	66.02	0.97
山东	137	0	5.11	61.31	31.39	2.19	广西	111	7.21	11.71	64.87	14.41	1.80
河南	158	0	6.33	72.78	20.89	0	西藏	74	0	0	0	100	0
湖北	103	0.97	22.33	63.11	13.59	0							

南、四川大部分地区存在高易感性的思茅松、云南松等优势种，风险等级容易上升。

4 讨论

相比于县域行政单元尺度的传统地学统计方法，利用地理栅格信息描述驱动变量，以RF模型建模预报风险的方法，实现了测报变量的数字矩阵描述，解决了测报结果的空间分异性与连续化描述的难题。全国栅格尺度的松材线虫病风险测报能够更好地反映人类活动、地理环境变化、寄主易感性对松材线虫病传播轨迹的影响，对基层单位开展检疫、预防和治理等工作有指示和支持作用。

表 7 RF 与寄主易感性的松材线虫病地理栅格潜在风险预测

Table 7 Prediction of the potential risk of geographic grids for RF and host susceptibility to pine wilt disease

省级行政区	1 000 m 栅格数	平均值	标准差	变异系数/%	预测值 ≥0.8	极易感染区占比/%	省级行政区	1 000 m 栅格数	平均值	标准差	变异系数/%	预测值 ≥0.8	极易感染区占比/%
福建	132 413	0.72	0.25	34.72	62 349	47.09	云南	383 006	0.16	0.16	100.00	35	0.01
浙江	113 295	0.64	0.25	39.06	37 881	33.44	吉林	214 242	0.18	0.11	61.11	0	0
湖南	220 312	0.62	0.23	37.10	61 780	28.04	黑龙江	505 985	0.12	0.09	75.00	0	0
广西	242 297	0.52	0.23	44.23	46 207	19.07	上海	7 464	0.24	0.04	16.67	0	0
江西	178 486	0.58	0.19	32.76	30 787	17.25	河北	197 881	0.04	0.06	150.00	0	0
广东	186 705	0.55	0.18	32.73	27 318	14.63	海南	35 107	0.39	0.06	15.38	0	0
贵州	178 808	0.40	0.22	55.00	18 343	10.26	北京	17 322	0.05	0.03	60.00	0	0
安徽	150 885	0.45	0.20	44.44	14 622	9.69	天津	12 079	0.06	0.03	50.00	0	0
湖北	193 550	0.47	0.16	34.04	16 639	8.60	甘肃	427 332	0.02	0.06	300.00	0	0
重庆	83 845	0.43	0.15	34.88	5 520	6.58	青海	699 262	0	0.02	0	0	0
山东	166 885	0.2	0.19	95.00	4 726	2.83	宁夏	52 440	0	0.01	0	0	0
河南	172 746	0.24	0.19	79.17	3 699	2.14	新疆	1 693 157	0	0.02	0	0	0
江苏	110 980	0.31	0.11	35.48	1 358	1.22	内蒙古	1 190 105	0.03	0.08	266.67	0	0
四川	488 113	0.16	0.16	100.00	2 394	0.49	山西	162 173	0.03	0.09	300.00	0	0
陕西	209 615	0.13	0.17	130.77	648	0.31	西藏	1 237 077	0.02	0.07	350.00	0	0
辽宁	159 784	0.27	0.15	55.56	438	0.27							

降水、气温和海拔是影响松材线虫病生存的主要因素，统计分析得出松材线虫病最低生存年均降水量范围为 110~121 mm。影响松材线虫病发生的年均低温范围为 10~13 ℃，这与叶建仁^[35]以年均气温 10~14 ℃ 区域为松材线虫病可以发生区，大于 14 ℃ 为流行区划分的预测区域基本一致。影响松材线虫病发生的海拔范围为 4~247 m，这与 LEE 等^[21]预测的松材线虫病最适生存海拔小于 200 m 基本一致。在传播影响中，除了媒介昆虫通过自然传播，人类活动对松材线虫病的扩散也具有重要的影响。松材线虫病早期发生地点多位于城市化活跃、人类活动强烈的区域内，其原因为城市用地、道路用地增加以及人工育林替代原始植被降低了生态稳定性^[36]，这为松材线虫病爆发与传播埋下巨大隐患。

全国地理栅格扩散风险格局分析表明：松材线虫病在华北平原和长江中下游平原持续扩散，并且存在 3 条传播廊道：一是沿着长江水系向西南地区及四川盆地扩散，现以水热条件优越、高城市化的重庆潜在风险等级为最高；二是沿着黄渤海临海城市向东北平原的扩散廊道，这个廊道中北京和天津的潜在风险等级突出，可能是受到高城市化发展和密集交通网络的影响；三是沿金沙江、珠江流域形成了一条由广东、海南、广西向云贵高原扩散的通道，虽西南地区海拔较高，但其丰富的松科植物仍为松材线虫病的生存和传播提供基础。

5 结论

基于多类型数据建立的 RF 与 SVM 模型均有较好的预测性能 (ROC-AUC>75%)。RF 模型具有更高的精度为 83.95%，并且预测的极高风险区与寄主高易感性分布高度重合，证明 RF 模型分类结果更符合实际。

影响松材线虫病发生的敏感变量是年均降水量、年均低温和海拔，而坡度小、人类活动强度高是导致松材线虫病快速传播的重要因素。潜在扩散区位于人类活动密集的低海拔地区、道路通达的林区、城市城镇分布区和人工林分布区。其中，极高风险分布地区主要位于华东地区的浙江、江西及福建，华南地区的广西、广东以及华中地区的湖南。松材线虫扩散过程中存在明显传播廊道与地理阻隔。松材线虫在华北平原的扩散传播受到秦岭和太行山脉的阻隔，并未侵入西北地区的黄土高原，而高风险疫区多集中在黄河水系以南地区。

本研究仅使用了 RF 及 SVM 共 2 种机器学习算法，样点数据仅为国家林业和草原局 2018—2021 年松材线虫病疫区数据，也还未考虑环境动态变化导致的松材线虫病疫区发生的风险变化。未来的研究应

以松材线虫病实际病害发生位置为基础,通过更多的机器学习算法或深度学习算法进行对比研究,充分考虑与研究区相适应的自然环境、人类活动等时空数据,以期对松材线虫病疫区的预测提供更加准确、精细的风险测报。

6 参考文献

- [1] MOTA M M, FUTAI K, VIEIRA P. Pine wilt disease and the pinewood nematode, *Bursaphelenchus xylophilus* [M]// CIANCIO A, MUKERJI K G. *Integrated Management of Fruit Crops Nematodes Vol 4*. Dordrecht: Springer, 2009: 253 – 274.
- [2] VICENTE C, ESPADA M, VIEIRA P, *et al*. Pine wilt disease: a threat to European forestry [J]. *European Journal of Plant Pathology*, 2012, **133**(1): 89 – 99.
- [3] SHIN S C. Pine wilt disease in Korea [M]// ZHAO B G, FUTAI K, SUTHERLAND J R, *et al*. *Pine Wilt Disease*. Tokyo: Springer, 2008: 26 – 32.
- [4] 国家林业与草原局. 全国松材线虫病疫情防控五年攻坚行动计划(2021—2025)[EB/OL]. 2021-07-07[2022-03-07]. <http://www.forestry.gov.cn>.
National Forestry and Grassland Administration. *Five-year Action Plan for Prevention and Control of Pine Wood Disease (2021–2025)* [EB/OL]. 2021-07-07[2022-03-07]. <http://www.forestry.gov.cn>.
- [5] 程瑚瑞, 林茂松, 黎伟强, 等. 南京黑松上发生的萎蔫线虫病[J]. 中国森林病虫, 1983, **2**(4): 1 – 5.
CHENG Hurui, LIN Maosong, LI Weiqiang, *et al*. Wilt nematode disease on black pine in Nanjing [J]. *Forest Pest and Disease*, 1983, **2**(4): 1 – 5.
- [6] 李计顺, 潘佳亮, 刘超, 等. 2020年全国松材线虫病疫情流行情况分析[J]. 中国森林病虫, 2021, **40**(4): 1 – 4.
LI Jishun, PAN Jialiang, LIU Chao, *et al*. Analysis of the epidemic situation of pine wilt disease in China in 2020 [J]. *Forest Pest and Disease*, 2021, **40**(4): 1 – 4.
- [7] 国家林业与草原局. 国家林业与草原局公告(2020年第5号)[EB/OL]. 2021-10-18[2022-03-07]. <http://www.forestry.gov.cn>.
National Forestry and Grassland Administration. *Announcement of National Forestry and Grassland Administration (No. 5, 2020)* [EB/OL]. 2021-10-18[2022-03-07].
- [8] 于治军, 李硕, 周艳涛, 等. 不同增温模式下我国松材线虫适生分布模拟与预测[J]. 东北林业大学学报, 2018, **46**(1): 85 – 91.
YU Zhijun, LI Shuo, ZHOU Yantao, *et al*. Spatial estimation and prediction of suitable distribution of *Bursaphelenchus xylophilus* with different warming modes in China [J]. *Journal of Northeast Forestry University*, 2018, **46**(1): 85 – 91.
- [9] LINIT M J. Nematode-vector relationships in the pine wilt disease system [J]. *Journal of Nematology*, 1988, **20**(2): 227 – 235.
- [10] IL C W, SONG H J, SOO K D, *et al*. Dispersal patterns of pine wilt disease in the early stage of its invasion in South Korea [J]. *Forests*, 2017, **8**(11): 411.
- [11] 王新荣, 朱孝伟, 胡月清, 等. 松墨天牛携带的松材线虫PCR检测技术[J]. 林业科学, 2009, **45**(7): 70 – 75.
WANG Xinrong, ZHU Xiaowei, HU Yueqing, *et al*. A PCR-based method for detecting *Bursaphelenchus xylophilus* from *Monochamus alternatus* [J]. *Scientia Silvae Sinicae*, 2009, **45**(7): 70 – 75.
- [12] 沈鹏, 李功权. 基于生态位因子模型的湖北省松材线虫病风险评估[J]. 浙江农林大学学报, 2021, **38**(3): 560 – 566.
SHEN Peng, LI Gongquan. Risk assessment of *Bursaphelenchus xylophilus* in Hubei Province based on ecological niche factor analysis model [J]. *Journal of Zhejiang A&F University*, 2021, **38**(3): 560 – 566.
- [13] 韩兵, 朴春根, 汪来发, 等. 中国松材线虫病的发生现状及治理对策[J]. 中国农学通报, 2007, **23**(2): 146 – 150.
HAN Bing, PIAO Chungun, WANG Laifa, *et al*. Development status of pinewood nematode disease and its management strategies in China [J]. *Chinese Agricultural Science Bulletin*, 2007, **23**(2): 146 – 150.
- [14] RYSS A Y, KULINICH O A, SUTHERLAND J R. Pine wilt disease: a short review of worldwide research [J]. *Forestry Studies in China*, 2011, **13**(2): 132 – 138.
- [15] 潘宏阳, 叶建仁, 吴小芹. 中国松材线虫病空间分布格局[J]. 生态学报, 2009, **29**(8): 4325 – 4331.
PAN Hongyang, YE Jianren, WU Xiaoqin. Spatial distribution patterns of pine wilt disease in China [J]. *Acta Ecologica*

- Sinica*, 2009, **29**(8): 4325 – 4331.
- [16] 涂业苟, 喻爱林, 阙生全, 等. 不同寄主植物对松褐天牛成虫寿命和繁殖的影响[J]. *西南农业学报*, 2019, **32**(8): 1801 – 1804.
TU Yegou, YU Ailin, QUE Shengquan, *et al.* Effects of different host plants on longevity and reproduction of *Monochamus alternatus* Hope adults [J]. *Southwest China Journal of Agricultural Sciences*, 2019, **32**(8): 1801 – 1804.
- [17] 涂业苟, 李翼, 喻爱林, 等. 松褐天牛成虫对不同寄主植物的选择性[J]. *中国植保导刊*, 2019, **39**(5): 50 – 52, 57.
TU Yegou, LI Yi, YU Ailin, *et al.* Feeding and oviposition preferences of *Monochamus alternatus* adults among different host plants [J]. *China Plant Protection*, 2019, **39**(5): 50 – 52, 57.
- [18] 习妍, 牛树奎. 气候要素对松材线虫病疫情的影响研究[J]. *林业资源管理*, 2008(4): 70 – 76.
XI Yan, NIU Shukui. The effects of climatic factors on pine wilt disease [J]. *Forest Resources Management*, 2008(4): 70 – 76.
- [19] 于海英, 吴昊. 辽宁发现松材线虫新寄主植物和新传播媒介昆虫[J]. *中国森林病虫害*, 2018, **37**(5): 61.
YU Haiying, WU Hao. New host plants and new vector insects for *Bursaphelenchus xylophilus* found in Liaoning [J]. *Forest Pest and Disease*, 2018, **37**(5): 61.
- [20] 张潮. 我国松材线虫病的扩散趋势及气候对疫情的影响研究[D]. 北京: 北京林业大学, 2020.
ZHANG Chao. *Spread Trend of Pine Wilt Disease in China and the Impact of Climate on the Epidemic* [D]. Beijing: Beijing Forestry University, 2020.
- [21] LEE D S, CHOI W I, NAM Y, *et al.* Predicting potential occurrence of pine wilt disease based on environmental factors in South Korea using machine learning algorithms [J/OL]. *Ecological Informatics*, 2021, **64**: 101378[2022-08-09]. doi: 10.1016/J.ECOINF.2021.101378.
- [22] 杨红艳, 杜健民, 阮培英, 等. 基于无人机遥感与随机森林的荒漠草原植被分类方法[J]. *农业机械学报*, 2021, **52**(6): 186 – 194.
YANG Hongyan, DU Jianmin, RUAN Peiying, *et al.* Vegetation classification of desert steppe based on Unmanned Aerial Vehicle Remote Sensing and Random Forest [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, **52**(6): 186 – 194.
- [23] 李浩, 方伟泉, 李浪浪, 等. 基于深度学习的松材线虫病害松木识别[J]. *林业工程学报*, 2021, **6**(6): 142 – 147.
LI Hao, FANG Weiquan, LI Langlang, *et al.* Recognition of pine wood infected with pine nematode disease based on deep learning [J]. *Journal of Forestry Engineering*, 2021, **6**(6): 142 – 147.
- [24] 杨宝君, 王秋丽, 邹卫东, 等. 不同松树品种对松材线虫的抗性[J]. *植物病理学报*, 1987, **17**(4): 211 – 214.
YANG Baojun, WANG Qiuli, ZOU Weidong, *et al.* The resistance of species to pine wood nematode, *Bursaphelenchus xylophilus* [J]. *Acta Phytopathologica Sinica*, 1987, **17**(4): 211 – 214.
- [25] 闫云凤. 基于决策森林的回归模型方法研究及应用[D]. 杭州: 浙江大学, 2019.
YAN Yunfeng. *Research and Application of Regression Modeling Methods Based on Decision Forests* [D]. Hangzhou: Zhejiang University, 2019.
- [26] 梁慧玲, 林玉蕊, 杨光, 等. 基于气象因子的随机森林算法在塔河地区林火预测中的应用[J]. *林业科学*, 2016, **52**(1): 89 – 98.
LIANG Huiling, LIN Yurui, YANG Guang, *et al.* Application of Random Forest Algorithm on the forest fire prediction in Tahe area based on meteorological factors [J]. *Scientia Silvae Sinicae*, 2016, **52**(1): 89 – 98.
- [27] 何云, 黄翀, 李贺, 等. 基于Sentinel-2A影像特征优选的随机森林土地覆盖分类[J]. *资源科学*, 2019, **41**(5): 992 – 1001.
HE Yun, HUANG Chong, LI He, *et al.* Land-cover classification of random forest based on Sentinel-2A image feature optimization [J]. *Resources Science*, 2019, **41**(5): 992 – 1001.
- [28] 周云. 基于随机森林算法的城市人口多尺度空间化研究[D]. 重庆: 西南大学, 2021.
ZHOU Yun. *Multi-scale Spatialization of Urban Population Based on Random Forest Algorithm* [D]. Chongqing: Southwest University, 2021.
- [29] 尹华, 胡玉平. 基于随机森林的不平衡特征选择算法[J]. *中山大学学报(自然科学版)*, 2014, **53**(5): 59 – 65.
YIN Hua, HU Yuping. An imbalanced feature selection algorithm based on Random Forest [J]. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 2014, **53**(5): 59 – 65.

- [30] VAPNIK V N. *The Nature of Statistical Learning Theory* [M]. New York: Springer, 1996.
- [31] 刘方园, 王水花, 张煜东. 支持向量机模型与应用综述[J]. *计算机系统应用*, 2018, 27(4): 1-9.
LIU Fangyuan, WANG Shuihua, ZHANG Yidong. Overview on models and applications of Support Vector Machine [J]. *Computer Systems & Applications*, 2018, 27(4): 1-9.
- [32] 王彦光, 朱鸿斌, 徐维超. ROC曲线及其分析方法综述[J]. *广东工业大学学报*, 2021, 38(1): 46-53.
WANG Yanguang, ZHU Hongbin, XU Weichao. A review on ROC curve and analysis [J]. *Journal of Guangdong University of Technology*, 2021, 38(1): 46-53.
- [33] 刘会河, 徐维超, 刘舜. 基于SVM的降维方法在三类ROC分析中的应用[J]. *计算机与现代化*, 2016(7): 49-54.
LIU Huihe, XU Weichao, LIU Shun. Dimension reduction method applying in three-class ROC analysis based on SVM [J]. *Computer and Modernization*, 2016(7): 49-54.
- [34] 章莉萍. 基于栅格模式的地图图形自动综合研究[D]. 武汉: 武汉大学, 2009.
ZHANG Liping. *Automatic Synthesis Study of Map Graphics Based on Raster Patterns* [D]. Wuhan: Wuhan University, 2009.
- [35] 叶建仁. 松材线虫病在中国的流行现状、防治技术与对策分析[J]. *林业科学*, 2019, 55(9): 1-10.
YE Jianren. Epidemic status of pine wilt disease in china and its prevention and control techniques and counter measures [J]. *Scientia Silvae Sinicae*, 2019, 55(9): 1-10.
- [36] 叶江霞, 王敬文, 张明莎, 等. 基于空间矩阵模型及0~1测度的美国白蛾风险格局分析[J]. *林业科学*, 2021, 57(1): 140-152.
YE Jiangxia, WANG Jingwen, ZHANG Mingsha, et al. Risk pattern analysis of *Hyphantria cunea* based on spatial matrix model and 0-1 measure [J]. *Scientia Silvae Sinicae*, 2021, 57(1): 140-152.