农业物联网大数据平台构建及病虫害预测模型开发

唐仁宗[1] 刘显富[2（通讯作者）]

1.铭创信息技术（莆田）有限公司 福建 莆田 351100

2.内蒙古汇鑫软件股份有限公司 内蒙古 呼和浩特 010040

【摘 要】：针对设施番茄生产中温室微环境波动快、病虫害发生滞后及农事记录分散等问题，以冀中平原A县50栋日光温室为对象，构建农业物联网大数据平台，并开展病虫害风险预测模型开发与应用验证。平台集成空气温湿度、土壤温湿度、光照强度、CO_2 浓度、叶面湿度、虫情灯、人工巡检和气象预报等多源数据，形成约180万条环境数据、1.2万条虫情数据和3000条病害标注数据。通过数据清洗、特征构建和标签生成，建立以XGBoost为核心的病虫害风险预测模型。试运行结果表明，平台传感器在线率达96.8%，数据缺失率控制在2.4%以内，平均入库延迟小于8 s。XGBoost模型准确率为89.6%，召回率为87.8%，F1值为88.5%，AUC为0.93，平均提前预警2.6 d。应用后主要病虫害发生面积下降21.3%，化学防治次数下降15.8%，亩均管理成本降低9.6%，说明该平台可有效提升设施番茄病虫害预警与生产管理水平。

【关键词】：农业物联网；大数据平台；设施番茄；病虫害预测；XGBoost

0 引言

设施番茄生产高度依赖温室环境调控，温湿度、光照、CO_2 浓度及叶面湿润状态的变化会直接影响病虫害发生风险。传统管理方式主要依靠人工巡检和经验判断，存在数据分散、响应滞后和预警精度不足等问题。随着农业物联网、边缘通信和机器学习技术的发展，多源数据连续采集与智能分析为设施作物病虫害精准防控提供了新路径。因此，构建面向生产现场的农业物联网大数据平台，对提高风险识别效率、减少化学防治投入和推动设施农业数字化管理具有重要意义。

1 农业物联网大数据平台构建需求与案例数据基础

鉴于设施番茄生产中温室微环境波动快、病虫害发生具有滞后性以及农事干预记录分散等现实情境，本研究把冀中平原A县设施番茄示范基地当作统一案例来开展平台需求界定。研究对象覆盖50栋日光温室，种植面积约80亩，数据采集周期为2023年9月至2024年6月，业务需求集中在棚室环境连续感知、虫情动态汇聚、病害巡检留痕以及预警信息联动推送等方面。平台数据来源囊括空气温湿度、土壤温湿度、光照强度、二氧化碳浓度、叶面湿度、虫情灯诱捕数量、人工巡检病斑率、农事记录以及气象预报数据，并把温室编号、时间戳、传感器位置、作物生育期以及数据单位纳入统一标准。环境传感器按每10 min采集1次进行配置，虫情灯每日汇总1次，人工巡检每周开展2次，最终形成约180万条环境数据、1.2万条虫情数据以及3000条病害标注数据，为后续平台构建以及病虫害风险预测模型训练提供口径一致的数据基础。

2 农业物联网大数据平台架构设计与病虫害预测模型开发

2.1 多源农业物联网数据采集与平台总体架构设计

鉴于冀中平原A县50栋日光温室的连续生产场景，平台把设施番茄病虫害风险识别需求嵌入感知层、传输层、数据层、模型层以及应用层的纵向架构中进行组织，见图1。感知层围绕棚室微环境以及病虫害发生前兆进行设备部署，空气温湿度、土壤墒情、光照强度、二氧化碳浓度以及叶面湿度传感器承担10 min级环境采集任务，虫情灯负责对蚜虫、白粉虱等诱捕数量进行日尺度汇聚，摄像设备用于留存植株长势、叶片病斑以及棚内作业图像。传输层选用4G以及LoRa网关混合链路，把棚间距离、供电条件以及数据量差异纳入通信方式选择依据，其中低频环境数据经LoRa汇入边缘网关，图像以及预警回执数据经4G网络上传至平台服务端。数据层建设时序数据库、图像文件库以及病虫害标签库，分别承接结构化监测序列、非结构化图像资料以及人工巡检标注结果。模型层完成特征计算、样本构建、模型训练以及风险推理，应用层面向基地管理人员提供病虫害风险预警、温室

环境看板以及农事建议，从而把数据接入、清洗、存储、分析以及预警推送衔接为闭环流程。

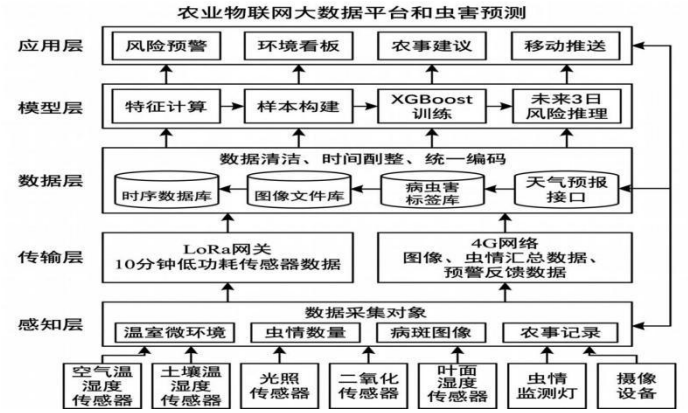

图 1 农业物联网大数据平台与病虫害预测流程图

2.2 数据清洗、特征构建与样本标签生成

设施番茄棚室数据在连续采集过程中容易受到传感器漂移、短时断电以及人工巡检间隔差异影响，数据治理环节把缺测识别、异常剔除以及时间粒度统一作为样本生成的基础工作。环境数据依据番茄适生阈值以及设备量程开展阈值剔除，对明显超出物理范围的温度、湿度、光照以及二氧化碳记录予以标记；对 10 min 级序列中的短时缺测值，运用相邻时段滑动均值进行补全，连续缺测时间较长的数据不进入模型训练样本。虫情灯日汇总数据、人工巡检周记录以及气象预报数据在 24 h 预测窗口内进行时间对齐，使环境波动、虫口变化以及病斑发展能够落入同一分析单元。特征构建围绕病虫害发生机理展开，主要囊括日均温、昼夜温差、相对湿度持续高值时长、叶面湿润时长、土壤含水率变化率、近 3 日虫口增长率以及近 7 日病斑率变化趋势。标签生成把人工巡检记录中的灰霉病、白粉病、蚜虫以及白粉虱发生等级转换为低风险、中风险、高风险 3 类，其中高风险样本占比约 18%，中风险约 32%，低风险约 50%，以保持训练样本结构与基地真实生产风险分布相近。

2.3 病虫害风险预测模型构建与预警规则设定

在完成多源特征矩阵构建后，模型开发把环境特征、虫情特征以及历史病害特征共同作为输入变量，输出未来 3 天设施番茄主要病虫害发生风险概率。样本集按照 7:2:1 划分为训练集、验证集以及测试集，训练集用于参数学习，验证集用于调节树深、学习率以及子采样比例，测试集保留为独立评估数据。为便于不同量纲特征进入同一风险判别尺度，连续型风险因子进入模型前进行归一化处理，公式如下：

$$R_i = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}}$$

其中，R_i 表示第 i 个样本归一化后的病虫害风险指数，S_i 表示第 i 个样本的原始风险因子取值，$S_{\min}$ 表示该类风险因子的最小观测值，$S_{\max}$ 表示该类风险因子的最大观测值。

主模型选用 XGBoost，并把逻辑回归以及随机森林当作对照模型来开展比较。XGBoost 借助多棵回归树逐步拟合残差，适宜处理温室环境变量之间的非线性关系以及特征交互关系，其二分类目标函数可表示为：

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t), \quad \Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

其中，$\mathcal{L}^{(t)}$ 表示第 t 轮迭代的目标函数，n 表示样本数量，l 表示二分类逻辑损失函数，y_i 表示真实风险标签，$\hat{y}_i^{(t-1)}$ 表示前 $t-1$ 轮模型预测值，$f_t(x_i)$ 表示第 t 棵树对样本 x_i 的输出，$\Omega(f_t)$ 表示复杂度惩罚项，T 表示叶节点数量，w_j 表示第 j 个叶节点权重，γ 以及 λ 为正则化参数。

平台把输出概率划分为 3 级风险区间，0 至 0.4 为低风险，0.4 至 0.7 为中风险，0.7 至 1.0 为高风险。当同一温室连续 2 天输出高风险时，系统触发红色预警，并向管理端推送通风降湿、黄板增设、虫源监测加密以及药剂轮换等处置建议，使模型推理结果能够进入基地日常农事决策流程。

3 平台应用验证与病虫害预测效果评估

3.1 平台运行稳定性与预警响应效果分析

2024 年 3 月至 6 月，平台在冀中平原 A 县设施番茄示范基地进入试运行阶段，运行对象仍为 50 栋日光温室。鉴于春季棚室通风频次增加、昼夜温湿度波动扩大以及白粉虱等虫口基数上升的生产现实，平台重点承担连续监测、风险判别以及处置提醒的衔接任务。试运行期间，平台日均接入环境、虫情、巡检以及农事记录约 6500 条，平均数据入库延迟控制在 8 s 以内，传感器在线率达到 96.8%，数据缺失率稳定在 2.4%以内。短时离线主要集中在棚间网关供电波动以及个别叶面湿度探头接触不良时段，经边缘缓存以及断点续传后，关键监测序列未出现大范围中断。预警链路方面，系统把模型输出结果同温室编号、作物生育期以及近期巡检记录进行关联，病虫害风险信息平均提前 2 至 3 d 推送至管理端。高风险预警的农户处置响应率达到 87.5%，通风降湿、

黄板增设以及虫情复查等措施能够被记录回传，说明平台已把监测、预测以及干预过程纳入相对连续的生产管理闭环。

3.2 病虫害预测精度与应用成效评估

模型效果评价以独立测试集以及试运行期人工巡检复核结果为依据，评价指标覆盖准确率、召回率、F1 值、AUC 以及平均提前预警天数。表 1 显示，逻辑回归对线性风险因子的解释较为稳定，但对连续高湿叠加虫口增长的复合情境识别不足；随机森林提高了非线性关系刻画能力，误报率有所降低；XGBoost 在各项指标上表现更优，准确率为89.6%，召回率为87.8%，F1 值为88.5%，AUC 达到 0.93，平均提前预警 2.6 d。结合基地应用前后生产记录，选用 XGBoost 的预警方案后，灰霉病、白粉病以及主要虫害发生面积下降 21.3%，化学防治次数下降 15.8%，亩均管理成本降低 9.6%。该结果表明，平台不仅提高了风险识别的时效性，还把模型输出转化为可执行的棚室管理动作。

表 1 不同病虫害预测模型性能及平台应用效果结果表

模型	准确率/%	召回率/%	F1 值/%	AUC	平均提前预警天数/d	主要病虫害发生面积变化/%	化学防治次数变化/%	亩均管理成本变化/%
逻辑回归	82.4	79.6	80.8	0.86	1.8	-14.2	-8.7	-5.1
随机森林	86.7	84.1	85.3	0.90	2.2	-18.6	-12.4	-7.8
XGBoost	89.6	87.8	88.5	0.93	2.6	-21.3	-15.8	-9.6

4 结束语

本研究围绕设施番茄病虫害监测预警需求，完成了多源数据采集、平台架构设计、样本治理、模型训练和应用验证。结果表明，农业物联网平台能够实现温室环境、虫情动态、病害巡检和农事处置的连续联动，XGBoost 模型在风险识别和提前预警方面具有较好表现。平台应用后，病虫害发生面积、防治次数和管理成本均有所下降，体现出较强的实用价值。后续研究可进一步扩大区域样本，融合图像识别与气象灾害数据，提升模型在不同季节、不同棚室条件下的泛化能力。

参考文献

[1]邓国飞.医院大数据平台构建与数据整合策略[J].中国科技信息,2025,(23):70-72.

[2]郭亚勤,王春博,郑富元,奥杰,蔡睿,杜伟.公共服务视角下地方特色知识产权大数据平台构建研究[J].科技管理研究,2025,45(04):176-184.

[3]杨丽.面向智慧校园生态体系的大数据平台研究及实践——以宁夏警官职业学院为例[J].电脑知识与技术,2024,20(33):76-78.

[4]宋姗姗,黄璜.面向公共管理学科知识服务的大数据平台构建及实践探索[J].实验室研究与探索,2025,44(11):202-208.

[5]张若宇,张建强,夏彬,吴超,常金强,王三辉.数字棉花技术与装备研究进展[J].农业机械学报,2025,56(05):1-16.

[6]陈孝照,孙维峰,吴尚璟.物联网生猪自动称重系统的设计[J].机电技术,2025,(03):60-64.

[7]王庆福.算力驱动下的智慧农业技术升级与发展模式创新[J].农业经济,2025,(06):7-9.

[8]张严,龙万圳,陈中柱,刘强,付清波.高校智慧实验室管理模式与平台构建研究[J].中国现代教育装备,2025,(13):161-164.

[9]霍小鹏,张毅,杨珍希,刘竞,林青,赵铁.油田生态环境大数据平台建设和应用[J].中国管理信息化,2024,27(11):97-100.

作者简介：唐仁宗（1987—），男，汉族，本科，研究方向为计算机工程。