

分类号：
学号：20222112078

密级：公开
单位代码：10759

石河子大学

硕士学位论文



设施番茄和辣椒苗期立枯病、猝倒病预测 模型研究

学位申请人	胡馨悦
指导教师	刁明 教授 樊庆鲁 研究员
申请学位类别	专业硕士
专业名称	农业
研究领域	农业工程与信息技术
所在学院	农学院

中国·新疆·石河子

2024年6月

分类号：
学 号：20222112078

密 级：公开
单位代码：10759

石河子大学

硕士学位论文



设施番茄和辣椒苗期立枯病、猝倒病预测 模型研究

学 位 申 请 人	胡馨悦
指 导 教 师	刁明 教授 樊庆鲁 研究员
申 请 学 位 类 别	专业硕士
专 业 名 称	农业
研 究 领 域	农业工程与信息技术
所 在 学 院	农学院

中国·新疆·石河子

2024 年 6 月

**Research on predictive models for sheath blight and damping-off
of tomato and pepper in seedling stage under protected culture**

A Dissertation Submitted to
Shihezi University
In Partial Fulfillment of the Requirements
for the Degree of
Master of Agriculture

By

Hu xin-yue
(Agricultural Engineering and Information Technology)

Dissertation Supervisor: Prof. Diao ming
and Researcher Fan qing-lu

May,2024

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所呈交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：胡馨悦

时间：2024年5月20日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：胡馨悦

时间：2024年5月20日

导师签名：刁明

时间：2024年5月20日

摘要

【目的】设施集约化育苗因其环境可控、管理便捷等特点，是蔬菜育苗阶段采用的主要方法。然而，由于苗期生长对环境条件要求极高，设施小气候需长期保持在温暖潮湿的环境下，可能导致苗期病害频发，进而影响成苗质量。立枯病和猝倒病是番茄和辣椒苗期高发且具有较大危害的病害，严重影响了优质成品苗的培育，生产上常采取药剂浸种、基质消毒和喷施杀菌剂等措施，尽管这些防治策略能一定程度上控制病害的发生，但高昂的管理成本、直接的幼苗损失，以及药害和伤苗等问题都与我国现阶段“减施增效”的绿色农业发展理念相悖。为提高设施育苗质量，及时预测番茄和辣椒苗期猝倒病、立枯病的发生，便于育苗人员提前采取干预措施，培育壮苗，亟需开展苗期病害预测模型的构建工作。

【方法】研究于 2023 年 3 月-8 月在位于疏勒县的新疆喀什（山东水发）现代蔬菜产业园进行，以番茄和辣椒的立枯病、猝倒病为监测对象，通过在设施内部署环境传感器，采集苗期温、湿度等环境数据，结合人工调查记录的病害发生种类和时间，经过数据整理和预处理后，运用机器学习算法，分别构建番茄和辣椒立枯病、猝倒病的二项逻辑斯蒂（Binary Logistic Regression, BLR）、支持向量机（Support Vector Machine, SVM）、决策树（Decision Tree, DT）、随机森林（Random Forest, RF）、K 最近邻（K-Nearest Neighbors, KNN）病害发生预测模型，并进行指标评价和性能对比，及对不同播期、不同设施类型的独立验证。

【结果】（1）试验表明太阳辐射和基质含水量加入番茄和辣椒立枯病、猝倒病预测模型的输入特征变量后，与环境温度和相对湿度结合大大提升了模型的预测效果，AUC 从平均在 0.75 提升到 0.92。

（2）综合模型预测效果，RF 模型最适用于番茄和辣椒立枯病、猝倒病的预测。RF 模型在番茄和辣椒立枯病、猝倒病预测中表现最佳，显示了高召回率、高精确率和高 F1 分数。在番茄立枯病预测任务中，RF 模型的 AUC 达到了 0.9799，精确度、准确率为 0.92，召回率、F1 分数是 0.91；在番茄猝倒病中，其 AUC 为 0.9765，准确率和精确度为 0.92，F1 分数是 0.91，召回率为 0.90；在辣椒立枯病中，其 AUC 为 0.9884，准确率、精确度为 0.95，F1 分数是 0.94，召回率为 0.93；在辣椒猝倒病中，其 AUC 达到 0.9799，准确率、精确度、F1 分数、召回率都高达 0.95，这表明 RF 模型具有优秀的分类能力和预测准确性。

【结论】设施环境（包括温、湿度）极大地影响着番茄和辣椒立枯病、猝倒病的发生，可作为病害预测模型的输入因子；而当太阳辐射和基质含水量也被纳入模型输入端时，可大大提升模型的预测效果，AUC 从平均在 0.75 提升到 0.92。综合分析比较上述五种模型的预测效果，RF 模型的评价指标相对最高，AUC 可达到 0.98 以上，最适用于番茄和辣椒立枯病、猝倒病的预测。

关键词：番茄辣椒；苗期；立枯病；猝倒病；预测模型

Abstract

【 Objective 】 Intensive seedling nursery in facilities is the main method used in the vegetable seedling stage due to its controlled environment and easy management. However, due to the high environmental requirements for seedling growth, the microclimate of the facility needs to be maintained in a warm and humid environment for a long period of time, which may lead to the frequent occurrence of seedling diseases, thus affecting the quality of the finished seedlings. sheath blight and damping-off are highly prevalent and harmful diseases in the seedling stage of tomato and pepper, which seriously affect the cultivation of high-quality finished seedlings, and the production often adopts measures such as pharmaceutical seed dipping, substrate disinfection and fungicide spraying, etc. Although these control strategies can control the occurrence of diseases, the high management cost, direct seedling loss, as well as the problems of drug damage and seedling injury are all contrary to the green agricultural development concept of "reducing application and increasing efficiency" at the present stage of China. In order to improve the quality of seedling nurseries in facilities, predict the occurrence of tomato and pepper sheath blight and damping-off in time, and facilitate the nursery personnel to take intervention measures in advance to cultivate strong seedlings, it is urgent to carry out the construction of seedling disease prediction models.

【 Methods 】 The study was conducted in the Xinjiang Kashgar (Shandong Shufa) Modern Vegetable Industry Garden located in Shule County from March to August 2023, with tomato and pepper sheath blight and damping-off as the monitoring objects, through the deployment of environmental sensors in the facility to collect environmental data such as temperature and humidity during the seedling period, combined with the type and time of disease occurrence recorded by manual investigation, after data collation and pre-processing, machine learning and deep learning were applied to construct the Binary Logistic Regression (BLR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest(RF), K-Nearest Neighbors(KNN), Disease occurrence prediction models with metrics evaluation and performance comparison, and independent validation for different sowing periods and facility types.

【 Results 】 (1) The experiments showed that solar radiation and substrate moisture content were added to the input characteristic variables of the tomato and pepper sheath blight and sudden downing-off

prediction model, and combined with ambient temperature and relative humidity greatly improved the prediction effect of the model, AUC increased from an average of 0.75 to 0.92.

(2) Combining the model prediction effects, the RF model was most suitable for tomato and pepper sheath blight and sudden blight prediction. The RF model performed best in tomato and pepper damping-off prediction, showing high recall, high precision, and high F1 scores. In the tomato sheath blight prediction task, the RF model achieved an AUC value of 0.9799, precision, accuracy of 0.92, recall, and F1 score of 0.91; in tomato sudden damping-off, it had an AUC value of 0.9765, accuracy and precision of 0.92, F1 score of 0.91, and recall of 0.90; and in pepper sheath damping-off, it had an AUC value of 0.9884, accuracy and precision were 0.95, F1 score was 0.94, and recall was 0.93; in pepper damping-off disease, its AUC value reached 0.9799, and accuracy, precision, F1 score, and recall were as high as 0.95, which indicated that the RF model had excellent classification ability and prediction accuracy.

【 Conclusion 】 The facility environment (including temperature and humidity) greatly affects the occurrence of tomato and pepper sheath blight and damping-off, and can be used as input factors for disease prediction models; When solar radiation and substrate moisture content are also included in the model input, the predictive performance of the model can be greatly improved, with the AUC increasing from an average of 0.75 to 0.92. After comprehensive analysis and comparison of the predictive performance of the five models mentioned above, the RF model has the highest evaluation index, with the AUC of over 0.98, making it the most suitable for predicting tomato and pepper sheath blight and damping-off.

Key words: tomato and pepper; seedling; sheath blight; damping-off; prediction models

目录

摘要.....	I
Abstract.....	II
第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 番茄和辣椒立枯病和猝倒病研究进展.....	2
1.2.2 病害预测模型研究现状.....	3
1.2.3 机器学习在病害预测模型的研究现状.....	4
1.3 研究内容.....	6
1.4 技术路线图.....	6
第 2 章 材料与方法.....	8
2.1 试验设计.....	8
2.2 数据采集与数据集构建.....	10
2.2.1 环境数据的采集.....	10
2.2.2 发病率的调查.....	12
2.2.3 病情指数的调查.....	12
2.2.4 环境与病情发展的数据处理.....	13
2.2.5 数据集构建.....	14
2.3 模型构建.....	14
2.3.1 数据预处理.....	14
2.3.2 五种模型及原理介绍.....	17
2.3.3 模型验证.....	20
2.3.4 建模环境.....	20
2.4 模型评价指标介绍.....	20
第 3 章 结果与分析.....	22
3.1 环境因素与病情的发展.....	22
3.1.1 环境与病发情况.....	22
3.1.2 环境与病情发展的关系.....	25
3.2 番茄立枯病模型构建及验证.....	30
3.2.1 番茄立枯病模型构建.....	30

3.2.2 番茄立枯病模型验证.....	32
3.3 番茄猝倒病模型构建及验证.....	35
3.3.1 番茄猝倒病模型构建.....	35
3.3.2 番茄猝倒病模型验证.....	36
3.4 辣椒立枯病模型构建及验证.....	39
3.4.1 辣椒立枯病模型构建.....	39
3.4.2 辣椒立枯病模型验证.....	40
3.5 辣椒猝倒病模型构建及验证.....	42
3.5.1 辣椒猝倒病模型构建.....	42
3.5.2 辣椒猝倒病模型验证.....	44
3.6 日光温室夏播独立验证建模.....	46
3.6.1 番茄立枯病.....	46
3.6.2 番茄猝倒病.....	49
3.6.3 辣椒立枯病.....	51
3.6.4 辣椒猝倒病.....	53
第4章 讨论与结论.....	55
4.1 讨论.....	55
4.1.1 预测模型特征类别分析.....	55
4.1.2 模型预测效果分析.....	55
4.2 结论与展望.....	56
4.2.1 结论.....	56
4.2.2 展望.....	56
参考文献.....	58
致谢.....	63

第1章 绪论

1.1 研究背景及意义

设施番茄和辣椒是我国设施生产的重要蔬菜作物，在苗期常常受到立枯病和猝倒病等病害的侵袭^[1]。立枯病和猝倒病是番茄和辣椒设施育苗苗期的主要病害，危害茄果和瓜类蔬菜的幼苗^[2]，影响番茄和辣椒种苗的生长和质量。立枯病在番茄和辣椒幼苗期和大苗期均可发病^[3]，病原菌以菌丝体、菌核在土中或病残体中越冬，可以通过水流、雨水、农具等途径进行传播，症状表现为在病苗茎基部产生暗褐色病斑，严重情况下基部萎缩，茎叶萎蔫，病苗直立死亡^[4]。猝倒病俗称“卡脖子”、“倒苗”、“小脚瘟”等^[5]，是一种常见的苗期土传病害，其具有广泛寄主，能引起茄果类、瓜类和十字花科等蔬菜的苗期病害，并能引起果实腐烂，是一类危害性极强的病害^[6]。

培育健壮的种苗是蔬菜生产的重要环节，种苗的质量直接影响到蔬菜作物的生长、产量和品质^[7]。随着农业产业结构的不断调整，设施蔬菜产业迅速发展，蔬菜育苗技术也在发生新的变革^[8]。穴盘育苗作为设施育苗的重要技术，有着省工、省力、节种、节能的特点，正在逐步替代传统的土壤畦面育苗和营养钵育苗^[9]。蔬菜集约化育苗技术是在相对可控的环境条件下，充分利用自然资源，采用科学、标准化的技术措施，运用机械、自动化手段，使作物秧苗生产达到快速、优质、高效、成批而又稳定的一种育苗方式，代表蔬菜育苗发展的方向^[10]。番茄和辣椒作为蔬菜设施育苗的重要作物，随着设施栽培的大力推广，番茄和辣椒的设施育苗面积不断扩大。然而，由于育苗设施内相对密闭高湿的环境，常导致苗床出现低温高湿、通风不良和光照不足的情况^[11]，给立枯病和猝倒病的发生提供了有利的条件，导致病害频繁暴发，种苗产量大幅度降低，给农户带来严重的经济损失，因此急需提供预测种苗病害的安全、有效、经济的预测办法^[12]。

随着计算机科学与人工智能的兴起，机器学习作为实现人工智能的重要方法，被广泛关注；运用机器学习方法，研究者可在计算机平台利用现有数据进行模型训练，寻找数据间的关系与规律，并对研究目标进行分类、回归等预测工作，为研究提供新思路，还可人为调整或利用算法自主循环上述过程以不断完善模型，提高精度以满足研究需要^[13]。机器学习正以强劲的势头向各个行业蔓延，已经成为植物病害预测模型的解决方案^[14]，机器学习使得对蔬菜病害进行及时有效且准确率高的预测变成现实，并成为了培育优质种苗的重要研究部分^[15]。通过传感器及物联网技术可以精确获取蔬菜生长环境数据，

结合机器学习算法发现植物发病规律，在一定程度上指导蔬菜种植与生产^[16]。同时，植物病害的预测也能最大程度地降低种苗病害带来的危害以及可能造成的损失，病害预测模型的研究正不断深入，但目前针对番茄和辣椒苗期发生病害进行预测的研究相对较少，因此，针对番茄和辣椒苗期易发高发病害——立枯病和猝倒病，运用机器学习方法构建病害预测模型，以降低病害给种苗带来的危害及损失。

本文探索了对于常规机器学习分类效果较好的二分类模型在植物病害预测上的效果，运用二项逻辑斯蒂（BLR）、支持向量机（SVM）、随机森林（RF）、决策树（DT）和 K 最近邻（KNN）这五种机器学习模型，作为番茄和辣椒立枯病和猝倒病病害发生预测的机器学习分类预测模型，并对五种模型进行比较与评价，从而为选用适宜的番茄和辣椒苗期病害的预测模型算法提供参考。

1.2 国内外研究现状

1.2.1 番茄和辣椒立枯病和猝倒病研究进展

1.2.1.1 立枯病的发生及危害

立枯病亦称作菜苗立枯病、烂根病或枯苗病，是危害番茄、辣椒、马铃薯、黄瓜等蔬菜生产上最普遍的危害^[17]，该病是一种常见的世界性病害，危害严重时可造成植株全株枯死，给农户造成巨大的经济损失。该疾病在湿润的条件下更易发病，有时与猝倒病混合发生，这种病害对植物苗期生长构成了极大的威胁，不论是发育成熟的大苗还是刚刚出土不久的小苗，都有可能受到影响而感染此病，成株期甚至番茄整个生长期也可发生^[18]。但是，该病害普遍发生于苗期，特别是在苗期的中、晚期^[19]。立枯病主要发生在植物幼苗的双子叶至 1 片真叶期，病程进展慢，发病初期，地上部分在白天会呈现出萎靡状态，但是在晚上这种状态会恢复^[20]，在感病苗的茎基部会产生椭圆形不规则的暗褐色病斑，向四周扩展，后期形成菌核^[12]，病斑逐渐凹陷，病发后在茎上蔓延一周，病部缢缩，有的木质部暴露在外^[19]，干枯，病苗萎蔫，趋势会不断加重，最后茎叶萎垂枯死^[21]，且幼苗在枯死后不会呈现倒伏现象。在潮湿的环境下，拔起病苗，就会发现一种类似蛛丝的褐色霉菌^[22]。

立枯病的病原是立枯丝核菌（*Rhizoctonia solani*），以菌丝体、菌核等方式在土中越冬，可在土中腐生 2—3 年，菌丝能直接侵入寄主，通过雨水、灌溉水、粪肥、农具进行传播、蔓延，属于土传病害。病菌发育适温 20—24℃，最高温度 40—42℃，最低温度 13—15℃^[23]，在 12℃ 以下或 30℃ 以上病菌生长受到抑制^[19]。苗期床温高、土壤水

分多、施用未腐熟肥料、播种过密、间苗不及时、徒长等均易诱发本病^[24]。高温高湿利于病菌生长，忽高忽低的温湿度会加重病情。当幼苗生长过密、间苗不及时、老化衰弱、温度偏高、通风透光条件差时，易引发此病^[19]。该病主要为害苗床，育苗期阴雨天气多、光照少的年份发病严重。在设施农业中，带病菌的土壤是病害初侵染的重要来源^[12]。立枯病不仅在苗床期可引起成片死苗现象，还可以在大田期造成茎基腐病，严重影响产量和品质^[25]。由立枯丝核菌引起的加工番茄苗期立枯病的发生，在各地造成了不同程度的经济损失，发病严重时常造成幼苗成片死亡，受害加工番茄一般减产 10%—20%，严重年份减产 40%—50%^[26]。

1.2.1.2 猝倒病的发生及危害

猝倒病俗称“卡脖子”、“倒苗”、“小脚瘟”，是一种常见的发生在苗期的土传病害^[6]，多发生在早春育苗床或育苗盘上，常见的症状有烂种、死苗和猝倒三种^[17]。猝倒病发病初期，幼苗茎基部呈浅黄绿色水渍状，似被开水烫过，很快转为黄褐色并发展至绕茎一周，逐渐缢缩呈细线状，致使幼苗上部失去支撑^[27]，幼苗地上部因失去支撑能力而倒伏地面，子叶未凋萎幼苗即倒伏于地，即秧苗折倒时，叶片仍为鲜绿色。幼苗倒地后萎蔫失水、干枯死亡^[28]。病菌易从果脐部或伤口处侵入，造成烂果。在空气湿度大时，果实病部表面见有白色棉絮状物，有别于立枯病，即是菌丝体^[17]。猝倒病在适宜条件下传播迅速，个别苗发病后一两天内即可造成周围幼苗成片倒伏死亡，因发病迅速，造成难以防治，常引起缺苗或毁种，损失很大^[29]。

猝倒病的病原菌多样，有腐霉 (*Pythium*)、镰刀菌 (*Fusarium*)、丝核菌 (*Rhizoctonia*) 等，喜高湿环境。猝倒病病菌随病残体在土壤中越冬，条件适宜时卵孢子萌发产生芽管直接侵入幼芽，或芽管顶端膨大后形成孢子囊^[28]，继而产生游动孢子侵入寄主，病菌腐生能力很强，可在土壤中长期存活，特别是富含有机质的土壤；病菌主要借雨水、灌溉水、带菌的堆肥、农具和农事活动传播^[30-31]，从秧苗茎基部侵入，潜育期 1—2 d。病菌虽喜 34—36 °C 高温，但在 8—9 °C 低温条件下也可生长，故苗床较长时间保持低温高湿的条件都会引起猝倒病的发生^[32]，使感病期拉长，很易发生猝倒病，尤其当苗期幼苗子叶中养分快耗尽而新根尚未扎实前，幼苗营养供应不足，抗病力最弱，如遇低温高湿环境，幼苗即刻发病^[28]。育苗期如遇寒流侵袭，保护棚不注意放风则会加剧猝倒病的发生。当幼苗皮层木栓化后，真叶长出，则逐步进入抗病阶段^[33]。猝倒病多在幼苗长出 1—2 片真叶前发生，3 片真叶后发病较少^[19]。

1.2.2 病害预测模型研究现状

董萍^[34]等基于对 Web of Science 上 2003 至 2022 年间全球作物病害监测和预警研究

的相关论文进行分析,发现病害预警的研究目前正处于快速发展时期,在2017年之后相关发文量较之前持续大幅度增长,在收集的21609篇文献中,76%的文献发布于此阶段,可见当前作物病害监测和预警这一主题是当前的热点和前沿。据分析,模型是搜索期间该领域的最高频关键词,常用的机器学习是重要的研究办法之一,成为未来可持续发展方向。

植物病害预测模型的研究一直是农业领域的重要课题之一。近年来,随着机器学习等技术的发展,相关研究取得了很大进展^[35]。这些模型可以根据植物病害图像、环境气候数据^[36]和病害历史数据等信息,对未来可能发生的病害做出预测,从而帮助农民及时采取措施进行防治^[37],提高农作物产量和质量^[38]。机器学习模型在植物病害预测中发挥着重要作用。支持向量机(SVM)、随机森林(RF)等机器学习算法可以利用环境数据(如气象数据、土壤数据)等特征进行病害预测^[39]。这些模型在特征选择、数据集成等方面有一定优势,能够有效地对植物病害进行预测和监测^[40]。

根据数据集类型,植物病害预测模型主要可以大致分为四种:基于传感器数据的方法,利用各种类型的传感器(如温度传感器、湿度传感器等)来监测植物的生长环境^[41],结合机器学习算法对这些数据进行分析,可以预测植物病害的发生情况;基于气象数据的方法,通过分析气象数据(如温度、湿度、降水量等)^[42]与植物病害之间的关系,可以建立预测模型,帮助预测植物病害的发生时间和程度;基于图像识别的方法,利用深度学习技术,特别是卷积神经网络^[43],可以对植物叶片的图像进行分析,从而实现植物病害的自动识别和分类,这种方法已经在许多研究中取得了很好的效果;基于专家系统的方法,专家系统结合了专家知识和规则,可以帮助诊断和预测植物病害^[44]。这种方法通常需要大量的专家知识来建立规则库。

总的来说,植物病害预测模型的研究方向多样且不断发展,结合不同的技术和方法可以取得更好的预测效果^[45]。机器学习模型在植物病害预测领域有自己优势和适合的应用场景^[46]。未来,随着技术的不断进步和数据的积累,这些模型将更加普及和深入应用,为农业生产提供更多有效的技术支持。

1.2.3 机器学习在病害预测模型的研究现状

机器学习在根据环境变化预测植物病害发生的研究是一个备受关注的领域^[47]。机器学习技术通过分析环境数据和病害数据之间的关联性,可以帮助农民和研究人员更准确地预测植物病害的发生,从而采取相应的防控措施^[48]。在这一领域的研究中,常用的机器学习算法包括支持向量机(SVM)、随机森林(RF)、逻辑回归(LR)等^[49]。这些算法可以利用历史的环境和病害数据进行训练,从而建立预测模型。通过分析气象数据、

土壤数据、植物生长数据等环境因素，这些模型可以预测植物病害的可能发生时间和地点^[50]。

赵睿等^[51]为解决温室植物病害预警、防治不及时的问题，设计了一种基于粒子群优化的支持向量机（PSO-SVR）模型的温室物联网预警系统，系统通过对观测数据进行分析训练，进而建立起植物病害预警模型。通过 PSO-SVR 模型对温室温度、湿度参数进行预测，预测准确率分别为 97.6%、96.8%。马丽丽等^[52]运用 BP 神经网络理论和方法，建立了根据环境因子预测黄瓜常见流行病，为解决植物病害预测预报提供简便的方法。杨矫捷^[53]基于多元回归分析、BP 神经网络、支持向量机多分类算法构建了安徽省小麦赤霉病预测模型并进行了对比分析，其中支持向量机多分类构建的模型准确率较高。姚志凤^[54]基于对生境信息的分析，建立了反向传播网络和支持向量机的小麦条锈病预测模型，用于小麦条锈病流程度度的中长期预测。李磊福等^[55]用孢子捕捉器对温室中甜瓜白粉病菌的孢子量进行监测，分析环境因子、孢子量和病情指数之间的关系，并采用逐步回归分析法构建温室甜瓜白粉病的流行预测模型。毕春光等^[56]利用贝叶斯算法改进传统 BP 神经网络构建玉米病害预警模型，预测的准确率达到 94.04%，相较传统 BP 神经网络模型准确率提高了 5.49%，得到了更好的预警效果。Khalili^[57]对 2000 株健康和染病植株研究了几种机器学习技术在大豆炭疽病预测中的应用，一组混合的生理和形态特征作为机器学习模型的输入，得到梯度树增强（GBT）分类器的灵敏度和特异度，研究结果证明了 GBT 在真实环境中预测木炭腐病的适用性。Mehta^[58]等利用随机森林（RF）和决策树（DT）建立了棉花病害预测模型，准确率分别达到 95.30%和 96.73%；Binyamin^[59]为识别促进疾病发展的环境因素并建立疾病预测模型，经线性回归分析，得到包括相对湿度、风速、降雨量、次优温度和最佳温度在内的 5 个环境变量多元回归模型能够正确地解释 83%的疾病生长变化；Singh^[60]利用热成像和可见光成像结合机器学习技术来估计田间条件下植物病害的严重程度，最终得出将机器学习（ML）和模型组合（MC）技术相结合，能够提高田间条件下植物病害严重程度的预测精度；Bhardwaj^[61]建立了一个评估普通燕麦白粉病高风险的逻辑回归模型，将温度、相对湿度和日照作为预测因子纳入模型，作为白粉病严重程度高的关键预测因素，采用交叉验证技术对模型的准确性进行了验证，可以提前喷洒杀菌剂，减少损失。

同时，一些研究也探索了机器学习模型在特征选择和数据集成方面的应用^[62]。通过选择最相关的环境特征和病害特征，可以提高模型的预测准确性^[63]。此外，将不同来源的环境数据进行融合和整合，可以更全面地分析植物病害发生的规律和趋势^[64]。总的来说，机器学习模型在根据环境变化预测植物病害发生的研究中发挥着重要作用^[48]。随着技术的不断进步和数据的不断积累，机器学习模型在这一领域的应用将更加广泛和深入，为植物保护和农业生产提供更有效的支持^[65]。

1.3 研究内容

考虑机器学习具有运行迅速、无需对特定病理系统进行复杂的推导，且更容易以计算机语言实现自动化运行的优势，本研究以番茄和辣椒苗期立枯病、猝倒病作为研究对象，结合物联网采集技术，运用二项逻辑斯蒂（BLR）、支持向量机（SVM）、决策树（DT）、K最近邻（KNN）、随机森林（RF）五种在植物病害预测、预报领域最常见的模型架构，开发了番茄和辣椒设施育苗病害预测模型，并对五种模型进行评价与对比，以此选出作物苗期两种常发病害的最适模型。

（1）环境温、湿度等影响因素对病害发生的调查与分析。通过环境温、湿度与发病之间的相关性分析，得到环境温、湿度会对病害产生影响，为后续模型的构建提供支撑。

（2）番茄和辣椒立枯病、猝倒病的模型构建和验证。春季在喀什的双膜双拱塑料大棚播了两批苗，记录了温、湿度数据，将大棚的第一批苗（试验1）数据构建二分类模型预测发病，第二批苗（试验2）数据用作试验1的模型验证，对比模型进行评估。

（3）番茄和辣椒立枯病、猝倒病预测模型变量选取改进。夏季在喀什的日光温室进行了育苗试验（试验3），记录了温、湿度及基质含水量和太阳辐射的数据，该部分数据独立进行二分类建模，并对比模型进行评估。

1.4 技术路线图

本文研究技术路线图如图 1-1 所示。

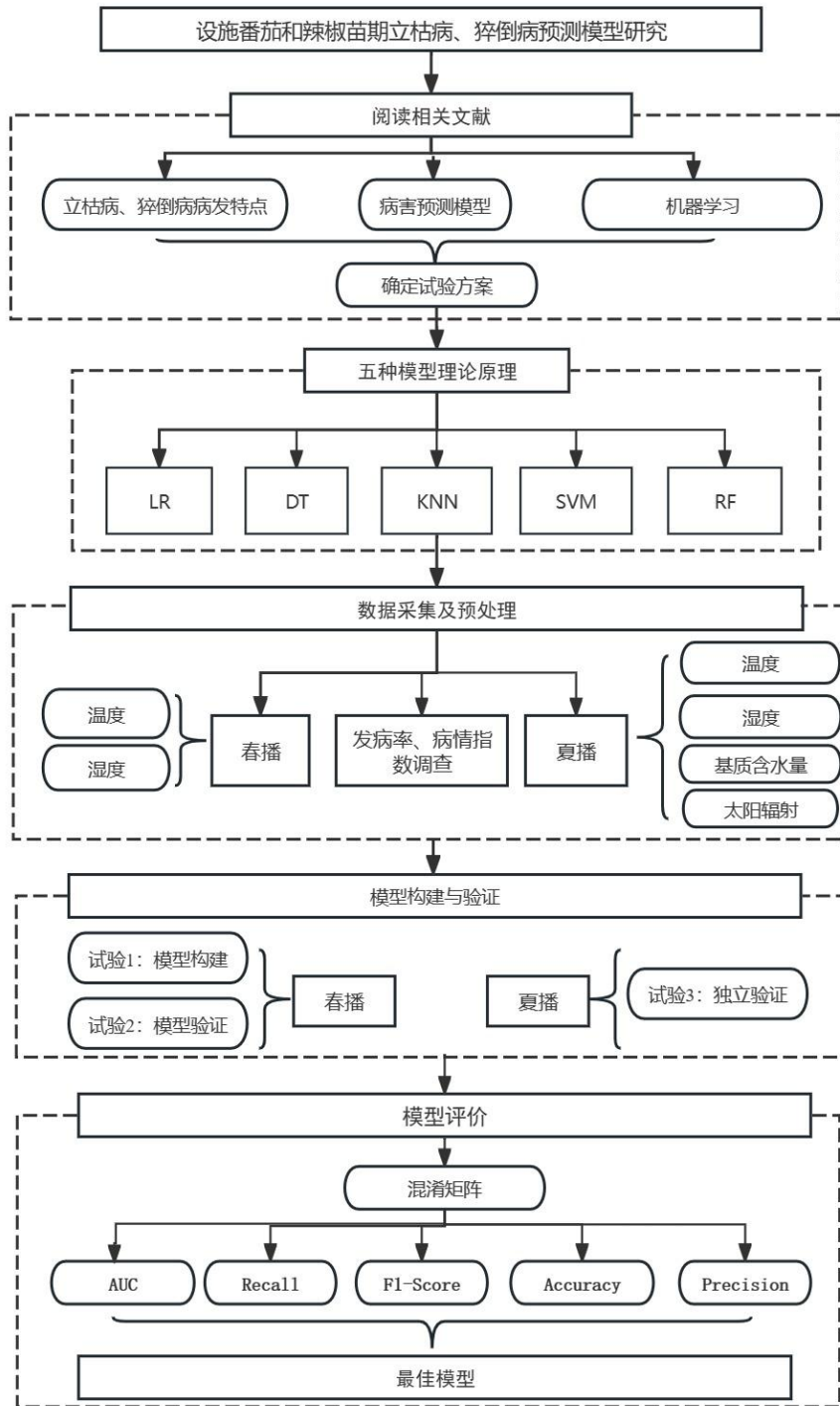


图 1-1 本文技术路线图
Figure1-1 Technical roadmap of article