

基于多种机器学习算法的霜预报模型对比

尹丁一*, 曹馨元

中国民用航空东北地区空中交通管理局空管中心, 辽宁 沈阳

收稿日期: 2026年9月11日; 录用日期: 2026年10月13日; 发布日期: 2026年10月26日

摘要

本研究针对机场霜预报问题, 采用沈阳桃仙国际机场2019~2024年冬半年地面观测资料, 基于霜形成的物理机制及皮尔逊相关系数筛选预报因子, 构建了逻辑回归、SVM、随机森林、XGBoost、LightGBM和MLP共六种机器学习模型及两种集成模型。研究采用时间序列划分数据集, 以TS评分最大化为目标进行分类阈值优化。结果表明SVM与随机森林表现最优, 集成学习增益有限。阈值优化可大幅提升预报性能, 各模型TS评分相对提升17.4%~51.9%。01时气温为核心预报因子, 各因子重要性排序符合辐射冷却物理机制。本研究构建的随机森林模型综合性能最优, 可为民航气象部门开展霜预报业务提供技术参考。

关键词

航空气象, 霜预报, 机器学习, 随机森林, 分类阈值优化

Comparison of Frost Forecasting Models Based on Multiple Machine Learning Algorithms

Dingyi Yin*, Xinyuan Cao

Air Traffic Management Center, Northeast Regional Air Traffic Management Bureau of CAAC, Shenyang Liaoning

Received: September 11, 2026; accepted: October 13, 2026; published: October 26, 2026

Abstract

Focusing on the airport frost forecasting task, this study utilizes ground meteorological observation data from Shenyang Taoxian International Airport collected during the winter half-years of 2019 to 2024. Based on the physical mechanism of frost formation and Pearson correlation coefficient

*通讯作者。

文章引用: 尹丁一, 曹馨元. 基于多种机器学习算法的霜预报模型对比[J]. 计算机科学与应用, 2026, 16(10): 64-73.
DOI: 10.12677/csa.2026.1610311

analysis, representative forecasting predictors are screened out. Six machine learning models—logistic regression, support vector machine (SVM), random forest, XGBoost, LightGBM, and multilayer perceptron (MLP)—together with two ensemble models are established for frost prediction. The dataset is divided in chronological order to ensure evaluation objectivity, and classification threshold optimization is conducted with the goal of maximizing the Threat Score (TS). The results indicate that SVM and random forest deliver the optimal forecasting performance, whereas ensemble learning brings limited performance improvement. Threshold optimization can significantly enhance forecasting capability, with the relative TS improvement of each model ranging from 17.4% to 51.9%. The air temperature at 01:00 is recognized as the core predictor, and the importance ranking of all factors conforms to the physical mechanism of radiative cooling. Overall, the random forest model constructed in this study achieves the best comprehensive performance, and can provide technical reference for civil aviation meteorological departments in operational frost forecasting.

Keywords

Aviation Meteorology, Frost Forecasting, Machine Learning, Random Forest, Classification Threshold Optimization

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

霜是近地面空气中的水汽在温度低于 0°C 的物体表面直接凝华形成的白色冰晶, 是秋末至初春常见的天气现象[1][2]。在航空气象中, 机场霜的出现会降低跑道摩擦系数, 影响飞机起降安全; 航空器表面结霜则会改变机翼气动外形、增加飞行阻力, 严重时甚至会引发飞行事故[3]。因此, 准确的霜预报对于保障民航冬季运行安全、优化资源调度具有重要意义。

目前, 机场霜的预报主要基于经验, 通过设定关键气象要素的阈值进行判别, 客观性不足、对预报员经验依赖度高。这类方法物理意义明确、操作简便, 但多依赖单一阈值或简单线性组合, 难以充分刻画各气象要素之间的非线性交互作用。随着机器学习的快速发展, 其在处理非线性高维问题上的优势逐渐显现, 为霜的定量化预报提供了新思路[4]。

国内外学者在霜预报领域已开展了较多研究。国外研究[5][6]较早将机器学习方法应用于机场结冰预报, 预报效果优于传统经验方法; 国内研究则更多面向农业领域[7], 直接针对机场霜现象的机器学习预报研究相对较少。而现有研究[8][9]多采用单一算法进行建模, 缺乏多种机器学习方法的系统横向对比。

针对上述问题, 本文以沈阳桃仙机场为研究站点, 选取 2019~2024 年冬半年的地面气象观测资料, 构建涵盖线性模型、核方法、树集成模型和神经网络四类的六种机器学习模型及两种集成模型, 采用时序划分保证评估真实性, 重点开展阈值优化研究, 结合特征重要性验证模型物理可解释性, 系统对比不同算法在机场霜预报场景的适用性。本研究旨在为民航气象部门开展霜的客观化、定量化预报提供一套完整的技术方案和模型选型参考。

2. 资料与方法

2.1. 研究资料

沈阳桃仙机场作为研究站点, 地处辽宁省沈阳市南郊, 冬季寒冷漫长, 秋末至初春期间辐射冷却强

烈, 霜天气频发, 在民航冬季运行中具有典型代表性[10]。研究采用 2019 年 10 月至 2024 年 4 月共五个冬半年的地面气象观测资料, 有效样本共 1062 天, 其中霜日 202 天, 霜日发生率 19.0%, 属于典型的不平衡二分类数据集。从月际分布(图 1)来看, 霜日数呈明显的单峰型特征, 12 月为霜高发期, 与北方冬季辐射冷却的气候规律一致。

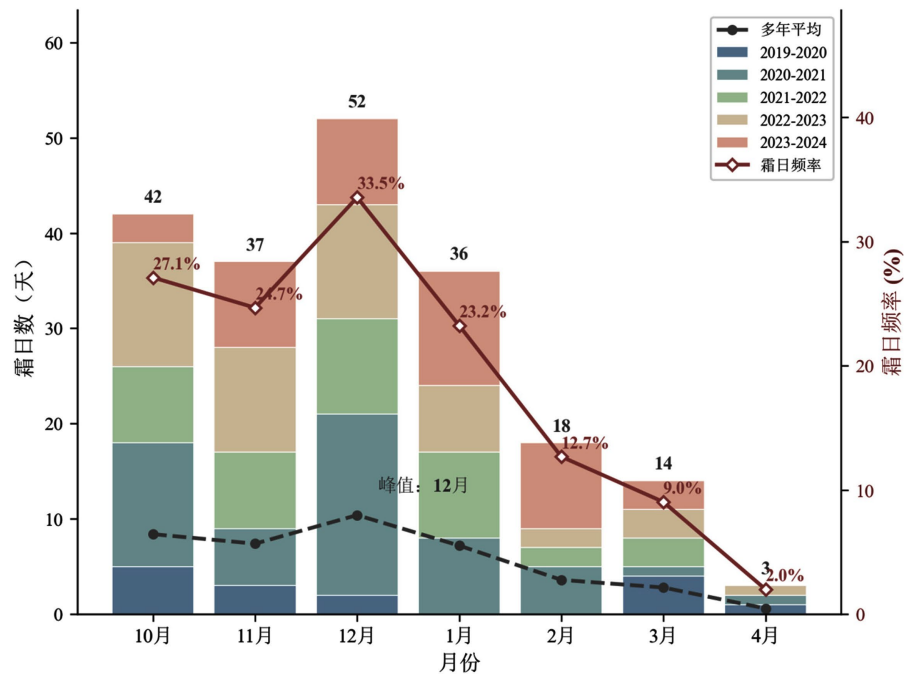


Figure 1. Distribution of the number of frost days by month during the winter half-year at Shenyang Taoxian Airport (2019~2024)

图 1. 沈阳桃仙机场冬半年各月霜日数分布(2019~2024 年)

基于霜形成的物理机制与现有研究成果[11][12], 结合机场观测业务实际, 通过计算各要素与霜发生的皮尔逊相关系数(表 1), 最终确定 01 时气温、06 时气温露点差、01 时风速和 06 时总云量四个因子作为模型输入。以常规地面观测记录中是否出现霜作为二分类标签, 有霜记为 1, 无霜记为 0。

Table 1. Pearson correlation coefficients between hourly observations of selected meteorological elements and the occurrence of frost

表 1. 部分气象要素逐时观测值与霜发生的皮尔逊相关系数

	1 时	2 时	3 时	4 时	5 时	6 时
rT-FR	-0.6265	-0.6263	-0.6255	-0.6199	-0.6226	-0.6214
rTg-FR	-0.5584	-0.5435	-0.5422	-0.5399	-0.5518	-0.5656
rT-Tg-FR	-0.2080	-0.2151	-0.2561	-0.2490	-0.2553	-0.2142
rRH-FR	0.1850	0.2228	0.2252	0.2195	0.2459	0.2810
rwind-FR	-0.4782	-0.4662	-0.4038	-0.4410	-0.3725	-0.3285
rCloudall-FR	-0.4588	-0.4681	-0.4701	-0.4528	-0.4927	-0.5435
rCloudlow-FR	-0.3014	-0.3638	-0.3701	-0.3431	-0.3955	-0.4228

2.2. 数据预处理

为消除量纲差异, 本文对输入特征进行 Z-score 标准化处理。标准化参数仅由训练集计算, 避免数据泄露。

在数据集划分上, 考虑到霜预报属于时间序列预报问题, 随机划分会引入未来信息, 导致模型性能虚高, 因此本文采用时间序列划分方式, 按时间顺序将前 80% 的样本作为训练集, 用于模型训练与超参数寻优, 后 20% 作为独立测试集, 仅用于最终性能评估, 不参与任何训练过程, 以保证评估结果的客观性, 符合实际预报业务场景。

2.3. 机器学习模型

本文选取六种具有代表性的机器学习算法, 涵盖线性模型、核方法、树集成模型和神经网络四类, 同时构建两种集成模型。

逻辑回归是经典线性分类模型, 通过 Sigmoid 函数输出概率, 结构简单、可解释性强、训练速度快。支持向量机(SVM)利用核函数将原始特征映射到高维空间, 通过寻找最大间隔分类超平面实现分类, 在小样本、非线性问题上表现出色。本文采用 RBF 径向基核函数:

$$K(x, y) = \exp(-\gamma x - y^2) \quad (1)$$

随机森林是基于 Bagging 框架的集成学习方法, 随机抽取样本和特征构建多棵决策树投票输出, 抗过拟合能力强, 能提供特征重要性评估。XGBoost 和 LightGBM 是梯度提升树的高效实现, 通过串行迭代拟合残差, 并内置正则化项控制模型复杂度, LightGBM 在此基础上优化训练效率。

多层感知机(MLP)是最基础的前馈神经网络, 本文采用单层隐藏层的简化 MLP 结构, 用于验证小样本场景下深度学习适用性。

为探索多模型融合的预报潜力, 本文构建了两种集成学习模型: 对各基模型输出的类别标签进行多数投票的硬投票集成; 以各模型 AUC 为权重融合预测概率的软投票集成。

2.4. 超参数调优

本文采用网格搜索结合 5 折交叉验证的方法进行超参数寻优。交叉验证同样在训练集内按时间序列划分, 确保调参过程不引入测试集信息。调优参数涵盖各模型正则系数、树数量、最大深度、神经网络神经元数量等核心超参数。

2.5. 评价指标与阈值优化

本文采用准确率、召回率、特异性及 AUC 等多维度评估模型。针对民航预报业务特点, 重点采用 TS 评分(威胁评分)作为核心评价指标, 综合平衡漏报与空报, 其计算公式为:

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

其中, TP 为正确预报的霜日数, FN 为漏报霜日数, FP 为空报霜日数。

常规二分类模型通常默认以 0.5 作为分类阈值, 模型普遍偏保守、空报较多, 本文以 TS 评分最大化为目标, 在 0.1 至 0.95 区间内遍历搜索各模型的最优分类阈值, 以获得漏报率与空报率的最佳平衡。

3. 模型性能分析

3.1. 分类阈值优化分析

首先采用默认阈值 0.5 完成模型测试, 非线性模型输出概率普遍偏高, 空报数量多; 逻辑回归概率分

布均匀，虽空报相对较少，但 TS 评分偏低。直接使用默认阈值无法适配霜预报场景，需要开展阈值优化以挖掘模型真实预报潜力。

以随机森林为例，分析其阈值从 0.1 变化至 0.95 过程中各指标的变化规律(图 2)，阈值提升过程中，召回率单调下降、特异性单调上升，TS 评分呈先升后降的单峰形势；在阈值 0.83 处 TS 达到峰值 0.885，为最优平衡点。低阈值下模型漏报少但空报激增；超过最优阈值后，漏报快速增加，TS 评分持续回落。

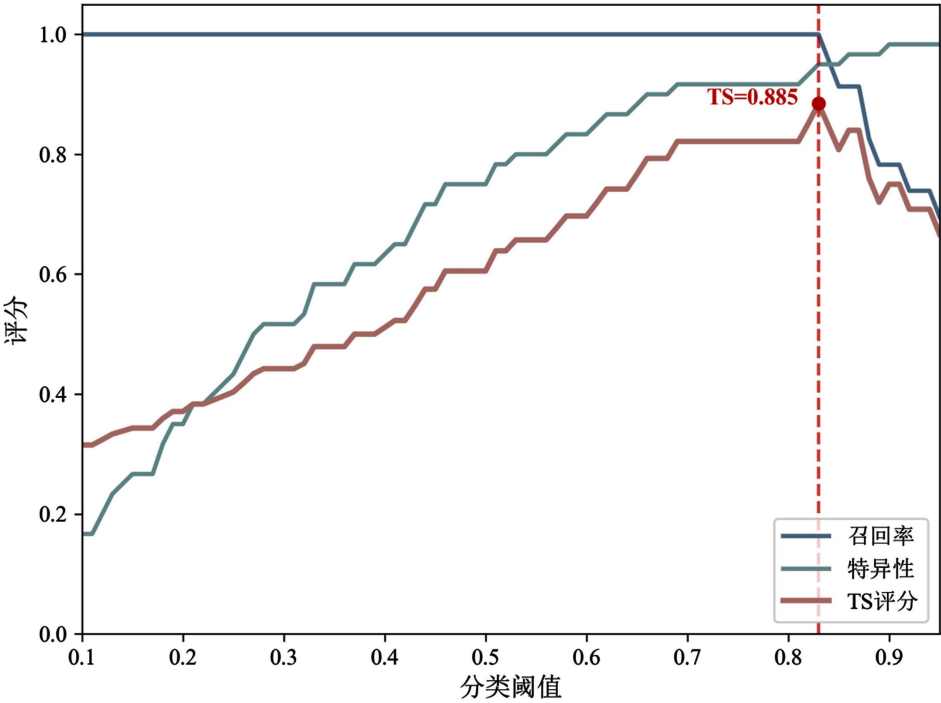


Figure 2. Changes in key performance indicators during threshold variation in a random forest model
图 2. 随机森林模型阈值变化过程中关键性能指标变化

各模型的最优阈值存在明显差异(表 2)，主要原因是不同算法的概率输出校准程度不同，SVM 和随机森林输出的概率普遍偏高，因此需要更高的阈值才能达到最优平衡；而逻辑回归的概率分布相对居中，最优阈值更接近 0.5。

Table 2. Comparison of TS scores for each model before and after classification threshold optimization
表 2. 分类阈值优化前后各模型 TS 评分对比

	最优阈值	阈值优化前 TS 评分	阈值优化后 TS 评分	TS 评分相对提升百分比
逻辑回归	0.68	0.605	0.710	17.4%
SVM	0.88	0.657	0.885	34.7%
随机森林	0.83	0.605	0.885	46.2%
XGBoost	0.70	0.535	0.767	43.4%
LightGBM	0.71	0.605	0.821	35.7%
MLP	0.69	0.460	0.579	25.8%
软投票集成	0.75	0.561	0.852	51.9%

经阈值优化后, 各模型的 TS 评分较默认阈值均有显著提升, 提升幅度在 17.4%~51.9%, 其中软投票集成提升幅度最大, 逻辑回归提升幅度最小, 这表明非线性模型和软投票集成模型的概率输出在默认阈值下普遍偏保守, 空报较多, 通过调高阈值获得的性能增益更为显著; 而逻辑回归作为线性模型, 其概率分布相对均匀, 默认阈值的适配性更好, 优化空间相对有限。阈值优化作为一种无需重新训练模型的后处理手段, 能够以极低的计算成本显著提升预报性能, 在业务应用中具有很强的实用价值。

3.2. 最优阈值下各模型综合性能对比

各模型经阈值优化后的综合性能指标如图 3 所示。非线性模型, 尤其是 SVM 与随机森林, 全面优于线性模型, 二者的 TS 评分均达到 0.885, 准确率为 96.4%, 召回率均达 100%, 特异性保持在 95.0%, 实现了漏报与空报之间的最优平衡。而逻辑回归的 TS 评分为 0.710, 非线性模型较其提升幅度普遍在 15% 以上, 这表明霜的形成确实强烈依赖于气象要素间的非线性交互作用。

MLP 的表现最弱, 各指标均为最低, TS 评分仅 0.579, 准确率为 80.7%, 特异性为 75.0%。这印证了在小样本条件下, 结构相对简单的神经网络难以学习到足够的泛化特征, 易陷入过拟合, 其性能远不如树模型与核方法。

ROC 曲线及其对应的 AUC 值衡量的是模型对正负样本的整体排序能力, 从 ROC 曲线对比(图 4)来看, 随机森林的 AUC 值高达 0.987, SVM 为 0.986, 这两类模型对有无霜的区分能力最强。AUC 排序与 TS 评分排序基本一致, 验证了性能评估结果的稳定性。

然而, XGBoost 与 LightGBM 出现了“高 AUC、低 TS”的现象。两者 AUC 值均为 0.978, 与最优模型相近, 但其 TS 评分却显著落后。这说明在极端不平衡的分类任务中, AUC 所反映的全局排序能力可能高估模型的实用价值, 模型优秀的排序能力并不完全能够转化为优秀的分类能力。因此, 对于本研究的霜预报场景, 以 TS 评分和召回率为核心的评估体系更能反映模型真实的应用潜力。

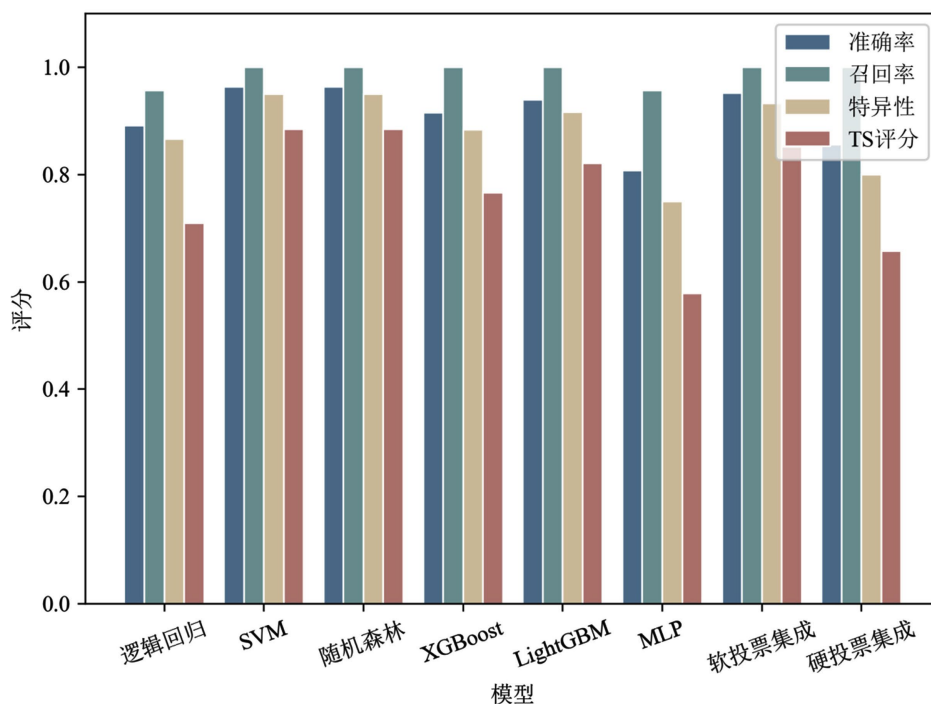


Figure 3. Comparison of comprehensive performance metrics for each model after threshold optimization
图 3. 阈值优化后各模型综合性能指标对比

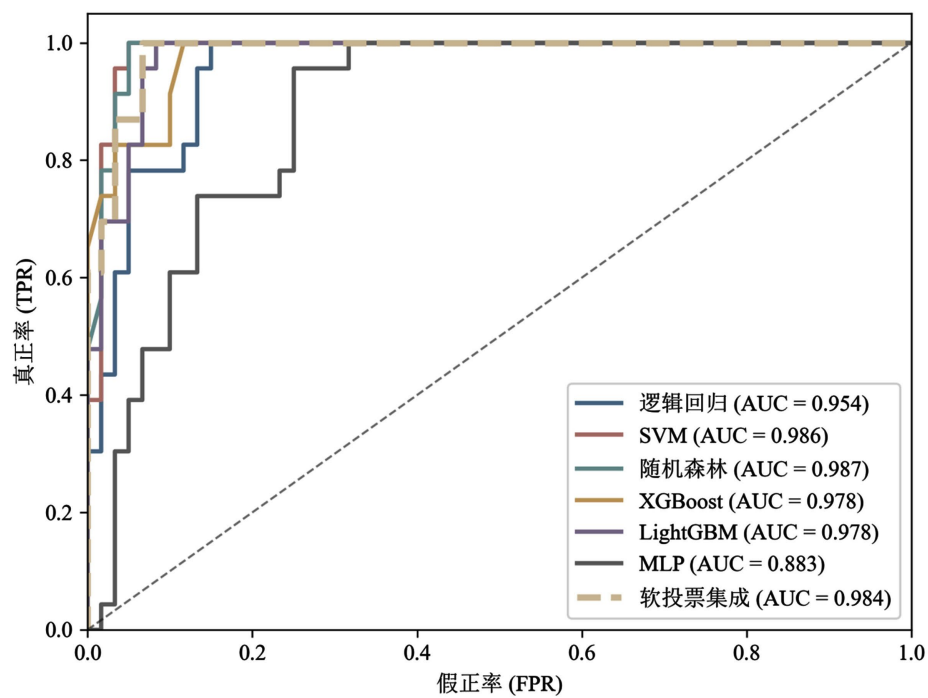


Figure 4. Comparison of ROC curves for the models after threshold optimization
图 4. 阈值优化后各模型 ROC 曲线对比

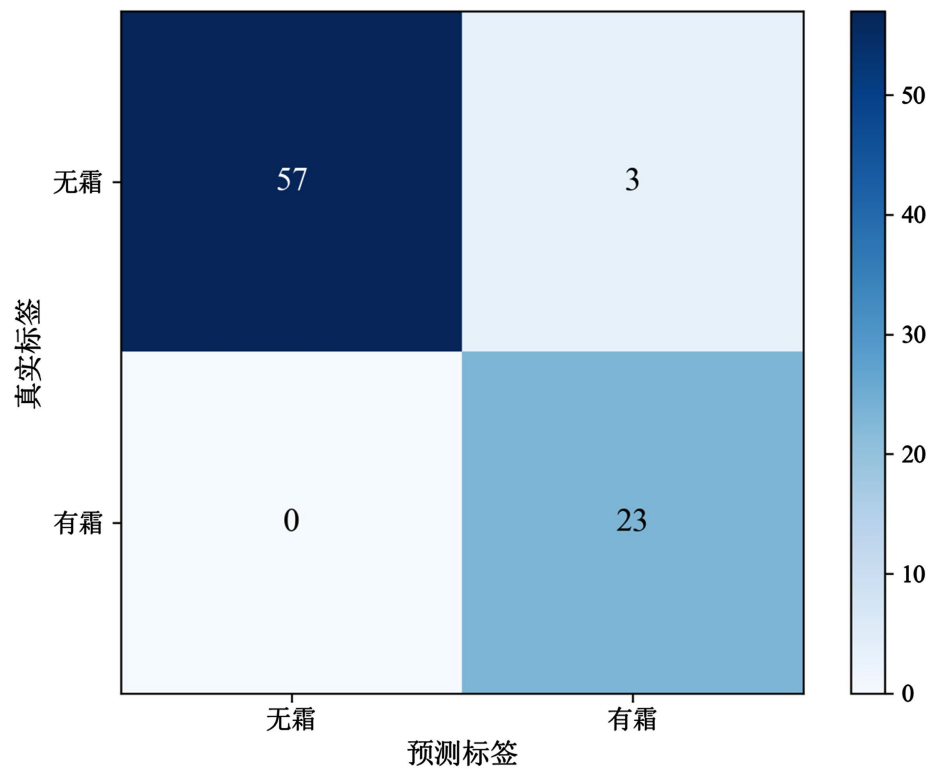


Figure 5. Confusion matrix for the random forest model at the optimal threshold
图 5. 随机森林模型最优阈值下的混淆矩阵

以最优模型随机森林为例, 其最优阈值下的混淆矩阵(图 5)显示, 测试集中 23 个霜日全部被正确预报, 无漏报发生, 60 个无霜日中仅有 3 个被误报为有霜。对于机场运行而言, 漏报霜天气可能导致跑道摩擦系数不足、机身结霜等安全隐患, 其代价远高于空报带来的除冰成本增加, 因此随机森林零漏报的特性具有重要的业务价值。

3.3. 特征重要性对比分析

为探究各预报因子对霜形成的贡献程度并验证模型的物理可解释性, 本文对比了不同算法的特征重要性排序。从特征重要性热力图(图 6)可以看出, 不同算法在核心因子的判断上具有高度共识, 但在次要因子的排序上存在一定差异。

01 时气温在所有算法中均被识别为最重要的预报因子, 其重要性占比在 SVM、随机森林、XGBoost 和 MLP 中均超过 50%, 在 MLP 中更是高达 90%以上。这一结果与霜形成的辐射冷却机制高度吻合: 夜间地表通过长波辐射持续降温, 气温越低越有利于凝华, 是决定霜能否形成的最核心条件。

06 时总云量普遍位列第二或第三重要因子, 反映了云量对夜间辐射冷却的显著调控作用: 云量越少, 地表长波辐射越易散失, 降温越剧烈, 越有利于霜的形成。06 时气温露点差反映近地层的水汽饱和程度, 气温露点差越小, 空气越接近饱和, 在温度降至冰点时更容易凝华成霜, 其重要性与总云量相当。01 时风速的重要性相对最低, 风速增大会加强近地层湍流混合, 抑制辐射冷却的强度。

总体而言, 各算法特征重要性排序与气象学基本规律一致, 说明所构建的机器学习模型具有良好的可解释性, 其预报结果具备物理依据。

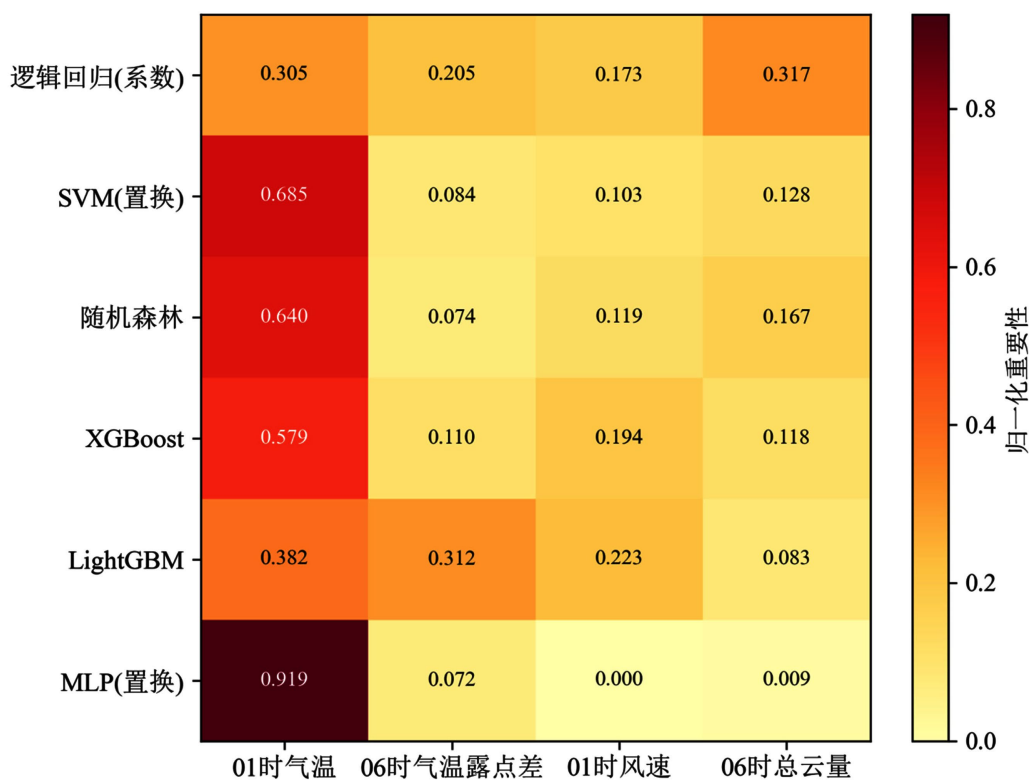


Figure 6. Feature importance heatmap
图 6. 特征重要性热力图

3.4. 集成学习效果分析

进一步构建硬投票和加权软投票两种集成模型, 以探索多模型融合对霜预报性能的提升效果。结果显示, 软投票集成的 TS 评分为 0.852, 而硬投票集成 TS 评分为 0.657。软投票集成性能显著优于硬投票, 主要原因是软投票利用了各模型的输出概率, 能捕捉分类标签之外的概率差异, 而硬投票仅依赖二值标签, 信息损失严重。

软投票集成的 TS 评分仍低于最优单模型 SVM 和随机森林, 集成学习在此任务中并未带来预期的性能增益。首先本数据集下 SVM 和随机森林的性能已经接近数据可分性的上限, 提升空间有限; 其次各基模型错分的样本高度重合, 集成投票无法有效纠正; 最后是基模型多样性不足, 六种模型输入特征完全相同, 学习到的决策边界相似度较高, 限制了集成的互补增益。

但是软投票集成的稳健性通常优于单一模型, 当数据分布发生轻微变化时, 集成模型的性能波动更小。在业务预报场景中, 可以将软投票集成作为随机森林的备份方案, 二者相互印证。

综合来看, 对于机场霜预报, 随机森林模型在性能和复杂度之间取得了最佳平衡, 是综合最优的选择; 软投票集成可作为稳健性补充方案。

4. 结论

本文利用沈阳桃仙机场 2019~2024 年冬半年地面观测资料, 选取 4 个气象预报因子, 构建了逻辑回归、SVM、随机森林、XGBoost、LightGBM 和 MLP 共 6 种机器学习模型及两种集成模型, 采用时间序列划分验证, 并以 TS 评分为目标进行分类阈值优化, 主要结论如下:

(1) 非线性机器学习模型的霜预报性能显著优于线性模型。SVM 与随机森林表现最优, 最优阈值下测试集 TS 评分均达到 0.885, 准确率为 96.4%, 且霜事件召回率为 100%, 在保证零漏报的同时将空报率控制在较低水平, 其中随机森林 AUC 值为 0.987, 综合性能略胜一筹, 是本数据集下的推荐算法。MLP 神经网络因样本量限制表现最差, 说明小样本场景下传统机器学习方法比简单神经网络更具优势。

(2) 分类阈值优化可有效提升综合预报性能。默认阈值下模型普遍偏于保守, 空报较多; 以 TS 最大化为目标优化后, 各模型 TS 评分提升 17.4%~51.9%。其中树模型和集成模型的优化增益尤为突出。随机森林的最优阈值为 0.83, 在漏报率与空报率之间取得了最佳平衡, 且阈值可根据不同业务导向灵活调整, 具有较强的实用性。

(3) 特征重要性分析表明, 01 时气温是霜预报的核心主导因子, 在多数算法中的贡献占比可达 50% 以上, 总云量、气温露点差和风速为次要影响因子, 分别从天空辐射条件、空气饱和程度和湍流混合强度三个方面影响霜的形成, 其重要性排序与辐射冷却导致霜形成的物理机制基本一致, 验证了机器学习模型的物理可解释性。

(4) 加权软投票集成模型的 TS 评分为 0.852, 略低于最优单模型随机森林, 集成增益有限, 主要原因是单模型性能已接近数据可分上限, 且各基模型的错误样本具有较高重叠度。但集成模型在稳健性上具有一定优势, 可作为业务预报的补充备份方案, 与主模型结果相互印证。

本研究构建的“预报因子筛选 - 时序验证 - 阈值优化 - 可解释性验证”完整建模框架, 验证了机器学习方法在机场霜客观化、量化预报中的有效性, 所构建的随机森林模型综合性能最优, 可为民航气象部门开展霜的客观化、量化预报提供技术参考。

参考文献

- [1] 吴艾黎, 奴开. 宁波机场 2010-2022 年霜天气特征统计和分析[J]. 科技与创新, 2024(6): 163-165.
- [2] 范强, 黄琪波. 杭州萧山机场霜的气候特征分析及预报方法初探[C]//第二十三届民航气象技术交流会论文集.

- 北京: 中国民航出版社, 2015: 91-98.
- [3] 中国民用航空局. 地面结冰条件下的运行[Z]. 北京, 2024.
https://www.caac.gov.cn/XXGK/XXGK/GFXWJ/202411/t20241107_225782.html, 2024-11-04.
- [4] 陈斌, 周冲, 刘悦. 跑道积冰特征因素提取与灰色神经网络预测[J]. 公路工程, 2022, 47(4): 192-198.
- [5] Tomkowicz, M.A. and Schmitt, A.O. (2019) Frost Prediction in Apple Orchards Based upon Time Series Models. In: Skiadas, C.H. and Bozeman, J.R., Eds., *Data Analysis and Applications 1*, Wiley, 181-94.
<https://doi.org/10.1002/9781119597568.ch13>
- [6] Jeong, Y., Chung, U. and Kim, K. (2018) Predicting Future Frost Damage Risk of Kiwifruit in Korea under Climate Change Using an Integrated Modelling Approach. *International Journal of Climatology*, **38**, 5354-5367.
<https://doi.org/10.1002/joc.5737>
- [7] 李曼. 霜与霜冻的成因及对作物的影响[J]. 现代农业科技, 2015(4): 249.
- [8] 赵京华, 张钰欣, 刘源, 等. 青岛机场结霜预测模型的构建及验证[J]. 民航学报, 2025, 9(6): 94-97, 21.
- [9] Masaki, Y. (2021) Probability of Frost Occurrence Based on Meteorological Observation in Japan. *Meteorology and Atmospheric Physics*, **133**, 693-705. <https://doi.org/10.1007/s00703-020-00775-w>
- [10] 李辑, 严晓瑜, 王颖. 辽宁省近 50 年霜的气候变化特征[J]. 气象, 2010, 36(11): 38-45.
- [11] 宋素蕊, 黄先成, 王文英, 等. 湿度小的晴夜成霜分析[J]. 气象与环境科学, 2007, 30(z1): 186-187.
- [12] Zhang, H., Yuan, N., Ma, Z. and Huang, Y. (2021) Understanding the Soil Temperature Variability at Different Depths: Effects of Surface Air Temperature, Snow Cover, and the Soil Memory. *Advances in Atmospheric Sciences*, **38**, 493-503. <https://doi.org/10.1007/s00376-020-0074-y>