

气象因素驱动的布地奈德高用药日预测集成学习模型的构建与比较^Δ

陈祺焄^{1,2*}, 周悦², 张晓俊², 倪璟雯², 孙国强², 高分飞^{1#}, 夏丽珍², 李梓豪¹ (1. 汕头大学医学院, 广东 汕头 515041; 2. 三明市中西医结合医院药学部, 福建 三明 365000)

中图分类号 R952 文献标志码 A 文章编号 1001-0408(2025)21-2723-04
DOI 10.6039/j.issn.1001-0408.2025.21.18



摘要 **目的** 通过气象因素构建预测布地奈德高用药日的集成学习模型,为医院药房管理提供参考。**方法** 基于三明市中西医结合医院主管区域2024年气象数据及同期该院的布地奈德门诊使用数据,将布地奈德门诊使用数据的第75百分位数定义为高用药日,并相应形成高用药日数据集,将预测问题转化为分类问题,构建随机森林模型、极端梯度提升模型、直方图梯度提升分类模型3种集成学习模型。以准确率、精确率、召回率、F1分数、对数损失函数为指标评估模型的性能,通过沙普利可加性特征解释(SHAP)方法分析模型的可解释性。**结果** 直方图梯度提升分类模型的性能最佳(准确率=0.75, F1分数=0.48),其次为极端梯度提升模型(准确率=0.74, F1分数=0.43)和随机森林模型(准确率=0.72, F1分数=0.22);SHAP结果提示后2个模型的预测结果相关性最高。**结论** 集成学习模型可有效预测布地奈德的高用药日,其中直方图梯度提升分类模型的预测能力最优;低温、高湿和低气压对布地奈德日用药量预测存在显著正向影响。

关键词 布地奈德;气象因素;集成学习;可解释性人工智能

Meteorological factor-driven prediction of high-use days of budesonide: construction and comparison of ensemble learning models

CHEN Qitao^{1,2}, ZHOU Yue², ZHANG Xiaojun², NI Jingwen², SUN Guoqiang², GAO Fenfei¹, XIA Lizhen², LI Zihao¹ (1. Medical College, Shantou University, Guangdong Shantou 515041, China; 2. Dept. of Pharmacy, Sanming Hospital of Integrated Traditional Chinese and Western Medicine, Fujian Sanming 365000, China)

ABSTRACT **OBJECTIVE** To construct ensemble learning models for predicting high-use days of budesonide based on meteorological factors, thereby providing reference for hospital pharmacy management. **METHODS** Meteorological data for 2024 and outpatient budesonide usage data from the jurisdiction of Sanming Hospital of Integrated Traditional Chinese and Western Medicine were collected. High-use days were defined as the 75th percentile of outpatient budesonide usage, and a corresponding dataset was established. The prediction task was formulated as a classification problem, and three ensemble learning models were developed: Random Forest, Extreme Gradient Boosting (XGBoost), and Histogram-based Gradient Boosting Classifier. Model performance was evaluated using accuracy, precision, recall, F1-score, and log-loss. Model interpretability was analyzed using Shapley Additive Explanations (SHAP). **RESULTS** The Histogram-based Gradient Boosting Classifier achieved the best performance (accuracy=0.75, F1-score=0.48), followed by XGBoost (accuracy=0.74, F1-score=0.43) and Random Forest (accuracy=0.72, F1-score=0.22). SHAP results suggested that the prediction results of the last two models have the highest correction. **CONCLUSIONS** Ensemble learning models can effectively predict high-use days of budesonide, with the Histogram-based Gradient Boosting Classifier demonstrating the best predictive performance. Low temperature, high humidity, and low atmospheric pressure show significant positive impacts on the prediction of daily budesonide usage.

KEYWORDS budesonide; meteorological factors; ensemble learning; explainable artificial intelligence

^Δ **基金项目** 国家中医药管理局项目(No.国中医药人教函[2022] 75号);2025年度福建省自然科学基金项目(No.2025J011566);福建中医药大学校管课题临床专项(No.XB2024075)

* **第一作者** 执业药师。研究方向:药理学、临床药学。电话: 0598-8033609。E-mail: cqt1945@163.com

通信作者 教授,博士。研究方向:药理学、临床药学。电话: 0598-8033609。E-mail: 869220642@qq.com

呼吸系统疾病的发生与发展受多种因素影响,气象因素作为重要的环境因素之一,对患者的疾病进展及药物使用需求具有潜在影响^[1-2]。布地奈德是一种广泛用

于哮喘和慢性阻塞性肺疾病治疗的吸入性糖皮质激素^[3],其用药需求在不同气象因素下可能存在波动。尽管部分研究显示,气温、湿度、气压等气象因素可能影响呼吸系统疾病发作与药物使用^[4-5],但相关性研究仍较为零散,尚未就其与具体药物(如布地奈德)使用量之间的关系形成系统性结论。

传统线性方法在处理高维、多重共线性的气象数据时存在一定局限^[6]。近年来,集成学习模型在临床辅助决策、疾病预测、药物需求估算等领域应用广泛^[7-9],具备识别复杂模式与提高预测精度的优势,适用于构建基于气象数据的用药预测模型。为此,本研究基于2024年三明市中西医结合医院门诊使用布地奈德的用药数据及同期气象数据,建立高用药日数据集,以此构建3种高用药日集成学习模型并对模型的准确率(accuracy)、精确率(precision)、召回率(recall)、F1分数(F1-score)、对数损失函数(log-Loss)进行对比,同时结合沙普利可加性特征解释(Shapley additive explanations, SHAP)法分析模型的可解释性,旨在为医院药房管理提供参考。

1 资料与方法

1.1 数据来源与整合

本研究数据由以下两部分构成:(1)药品使用数据——三明市中西医结合医院2024年1月1日—12月31日布地奈德的每日使用量(单位:支);(2)气象数据——源自中国气象局,收集2024年1月1日—12月31日三明市中西医结合医院主管区域逐日气象数据,剔除未完整记录参数的无效日期数共计23天后,共纳入有效观测记录343条。为便于后续进一步研究世界各地气象数据,本研究统一采用美国国家海洋和大气管理局的气象因素单位^[10]。

本研究在Python 3.10环境下,调用Pandas库对2024年三明地区每日气象数据与布地奈德门诊使用数据进行处理。通过DATE字段采用内连接方式合并两个数据源,形成初始数据集。考虑到气象因素对药物的使用存在滞后效应,因此采用shift函数为关键气象因素(如温度、湿度、气压、风速)构建1~3天的滞后特征(Lag 1-3),滞后特征的引入旨在增强集成学习模型对气象影响的时间敏感性。清洗缺失值后,形成高用药日数据集,作为后续模型构建的基础。

1.2 高用药日定义

为实现对布地奈德使用高峰期的精准识别与预测,本研究将连续变量“日用药量”离散化为分类变量。结合统计学与临床药事管理的双重考量,第75百分位常用于标识数据中的上四分位区域(top quartile),代表相对

较高的水平^[9],可在公共卫生、资源调度、流行病学等领域广泛用于定义“高负荷”或“高风险”状态。本研究将2024年三明市中西医结合医院布地奈德日用药量的第75百分位数(即上四分位数)作为高用药日划分阈值。

1.3 高用药日集成学习模型构建与性能评估

1.3.1 模型构建

为实现对高用药日的精准预测,本研究基于分类建模方法构建了3种集成学习模型:随机森林模型(Random Forest)、极端梯度提升模型(Extreme Gradient Boosting, XGBoost)和直方图梯度提升分类模型(Histogram-based Gradient Boosting Classifier),在模型构建过程中,将高用药日数据集按照7:3的比例随机划分为训练集与测试集,使用训练集进行模型构建与训练。

与传统线性建模方法不同,集成学习模型具有处理高维特征、非线性关系以及特征间交互效应的优势,同时对多重共线性具有更强的鲁棒性,能够自动完成特征筛选与权重评估。因此,本研究将所有构建的气象滞后特征(包括温度、湿度、风速、气压等的1~3天滞后项)作为模型输入变量,以提升模型泛化能力;模型输出变量为判断是否为高用药日。

1.3.2 模型性能评估

模型性能评估基于高用药日数据集中的测试集进行,评估指标包括准确率、精确率、召回率、F1分数及对数损失函数。模型解释性分析使用SHAP法,模型解释性用于揭示各气象因素预测结果的重要性。本研究使用Python 3.10环境下的scikit-learn库,利用classification_report和log-Loss等函数计算上述指标,综合评估模型的预测能力与稳定性。

2 结果

2.1 性能评估结果

3种模型在测试集上的性能评估结果显示,3种模型的准确率总体处于0.72~0.75水平,直方图梯度提升分类模型的F1分数、召回率明显优于其他模型,体现出更优的高用药日预测能力。结果见表1。

表1 3种模型在测试集上的性能评估结果对比

模型名称	准确率	精确率	召回率	F1分数	对数损失函数
随机森林模型	0.72	0.44	0.14	0.22	0.543 6
极端梯度提升模型	0.74	0.53	0.36	0.43	0.766 7
直方图梯度提升分类模型	0.75	0.55	0.43	0.48	0.764 1

2.2 模型解释性评估结果

利用SHAP法基于测试集对3种模型进行特征贡献解释,结果及相关热力图见图1~5(由于随机森林模型的SHAP值整体接近于0,未能形成可解释性特征排序,

因此本研究不展示其SHAP分析的解解释结果,仅展示极端梯度提升模型与直方图梯度提升分类模型的解解释结果)。由SHAP分析图及特征贡献方向性散点图可知,温度、相对湿度和气压在极端梯度提升模型和直方图梯度提升分类模型中的贡献权重最高,且低温、高湿和低气压表现出显著的正向影响。3种模型预测结果相关性的热力图显示,极端梯度提升模型与直方图梯度提升分类模型的预测结果相关性最高。

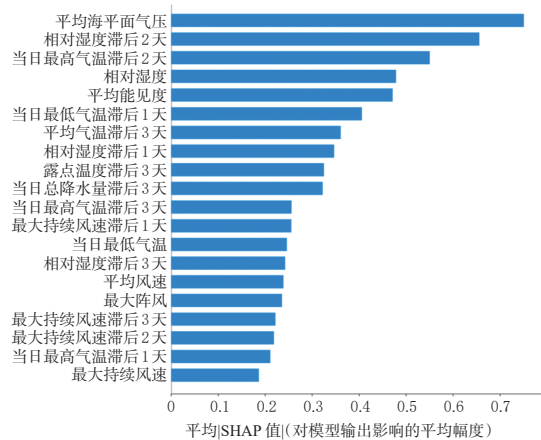


图1 极端梯度提升模型的SHAP分析图

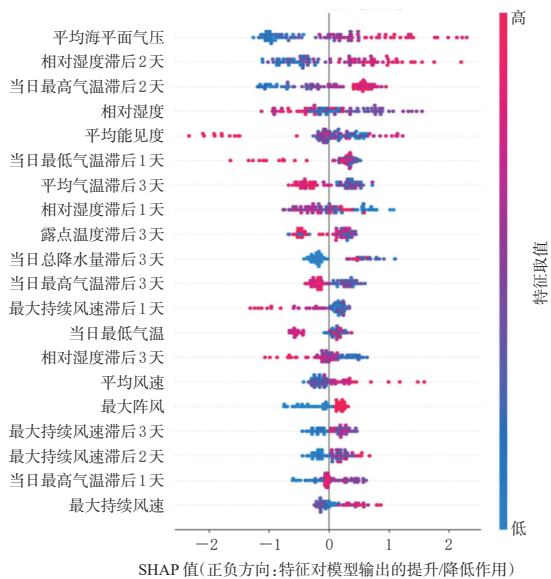


图2 极端梯度提升模型的SHAP特征贡献方向性散点图

3 讨论

3.1 高用药日集成学习模型的性能比较

从模型整体性能来看,3种集成学习模型均表现出一定的预测能力,且整体准确率 ≥ 0.72 。直方图梯度提升分类模型的准确率最高,达到0.75,F1分数也达到0.48,表示该模型对高用药日的识别能力相对较优;其次为极端梯度提升模型,该模型的准确率为0.74,F1分数为0.43;随机森林模型在准确率上略低,为0.72,且召回

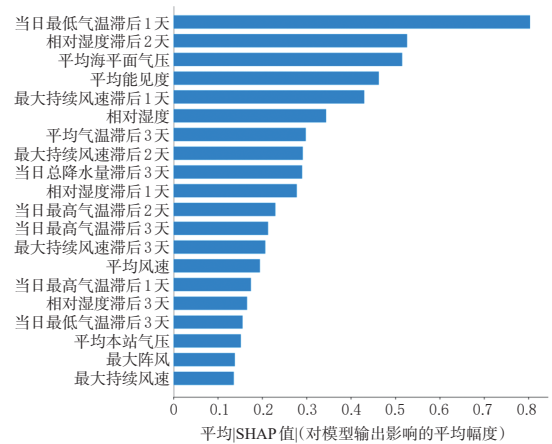


图3 直方图梯度提升分类模型SHAP分析图

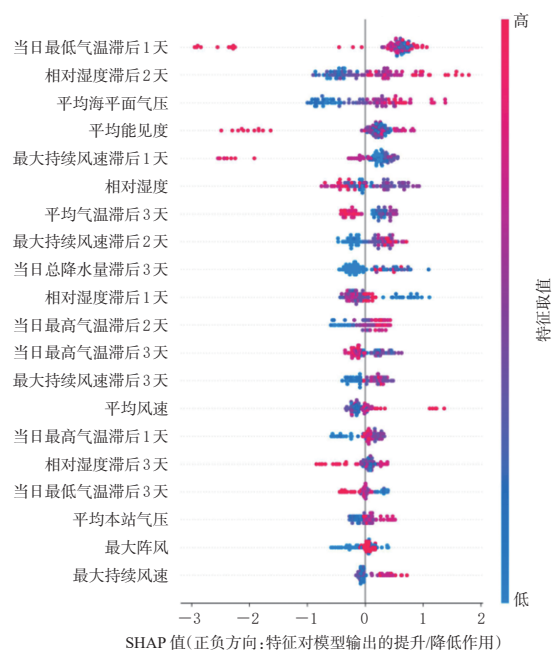
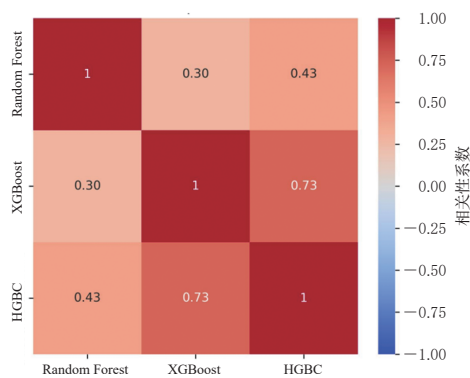


图4 直方图梯度提升分类模型的SHAP特征贡献方向性散点图



注:图中HGBC为直方图梯度提升分类模型。

图5 3种模型预测结果相关性的热力图

率仅为0.14,提示存在一定的漏检风险。

从SHAP分析结果来看,温度、相对湿度、气压等因素在极端梯度提升模型与直方图梯度提升分类模型的SHAP图中的贡献权重最高,低温、高湿和低气压表现出

显著的正向影响,即这些气象因素会显著增加模型预测为高用药日的概率。

从3种模型预测结果的相关性热力图分析结果来看,极端梯度提升模型与直方图梯度提升分类模型预测结果的相关性最高(相关性系数为0.73),表明二者用于气象因素建模的模式较为接近,随机森林模型与极端梯度提升模型、直方图梯度提升分类模型的相关性最低(相关性系数分别为0.30、0.43),表明极端梯度提升模型与直方图梯度提升分类模型对于气象数据中复杂非线性关系的建模能力更强,预测模式更加接近,而随机森林模型的建模机制可能更偏向于局部特征划分,且存在一定欠拟合风险。综合准确率等指标来看,直方图梯度提升分类模型相较于另外2种模型具有较优的预测能力。

3.2 高用药日预测集成学习模型的应用前景与局限性

本研究构建的直方图梯度提升分类模型可基于实时气象数据对医院布地奈德高用药日进行动态预测,实现提前1~3天预测,为医院药房提前备货、调整库存结构、优化人力资源分配提供科学依据。未来可将该模型嵌入医院信息系统,实现药品供应智能化管理,降低库存风险、提升供应效率。

然而,本研究也存在一定的局限性:一方面,当前模型的训练样本规模相对有限,模型泛化能力尚需在更大数据集上进一步验证;另一方面,布地奈德用药量的决定因素除气象因素外,还受流感流行情况、个体健康状况、医疗行为等多重因素影响^[11],若未来引入更多非气象类变量,模型预测效果或可进一步提升。

综上,本研究初步证明了气象因素对布地奈德用药存在显著影响,低温、高湿和低气压会显著增加模型预测为高用药日的概率;集成学习模型能够有效处理这一复杂关系,具备一定的实际应用潜力。

参考文献

[1] 王晓捷,梁栩霆,宋晓青,等.广州市番禺区大气PM_{2.5}

污染与儿科门诊量关系的时间序列分析[J]. 妇儿健康导刊,2025(7):192-198.

[2] 李旭,王方怡,单冰,等. 2013—2022年济南市空气污染物、气象因素对居民呼吸及循环系统相关死亡风险的影响[J]. 预防医学论坛,2025,31(4):298-303,312.

[3] 倪宁,石伟利. 布地奈德联合氨茶碱治疗老年慢性阻塞性肺疾病的临床疗效[J]. 中国卫生标准管理,2025,16(7):166-170.

[4] 任姝煜,陈风格,关茗洋,等. 石家庄市低温寒潮天气对人群死亡影响的病例交叉研究[J]. 环境卫生学杂志,2025,15(6):501-507.

[5] 付琳. 济南市温度对老年人呼吸系统疾病死亡影响的研究[D]. 济南:山东大学,2024.

[6] 石宜衡,王琦祎,孙婧,等. 气象大模型计算资源评估技术方法[J]. 数据与计算发展前沿(中英文),2025,7(4):182-195.

[7] 吴倩,陈意,余浩凯,等. 基于集成学习的IgA肾病中医辨证智能模型的构建及应用[J/OL]. 广州中医药大学学报,2025:1-12[2025-08-19]. <https://kns.cnki.net/kcms/detail/44.1425.R.20250818.1803.002.html>.

[8] 何鑫,于芳芳. 基于堆叠集成学习的结核病诊断模型[J]. 工业控制计算机,2025,38(7):115-116.

[9] 靳帅,李静,路潜,等. 基于Super Learner的前列腺癌风险预测模型构建与验证[J]. 医学信息,2025,38(14):13-19,24.

[10] National Centers for Environmental Information. Local Climatological Data (LCD) dataset documentation[EB/OL]. [2025-05-26]. https://www.ncei.noaa.gov/pub/data/cdo/documentation/LCD_documentation.pdf.

[11] 潘相丞,倪晓凤,彭容,等. 中国儿童吸入用布地奈德混悬液使用情况的系统评价和Meta分析[J]. 药物流行病学杂志,2023,32(10):1144-1157.

(收稿日期:2025-06-13 修回日期:2025-09-21)

(编辑:刘明伟)