

DB5104

四川省（攀枝花市）地方标准

DB5104/T 162—2026

气象敏感性疾病发病风险等级预测指南

2026-03-09 发布

2026-04-09 实施

目 次

前 言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 模型构建	2
5 等级划分及描述	3
6 预警发布	4
附录 A（规范性）模型性能评估方法	5
参考文献	7

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由攀枝花市中心医院提出。

本文件由攀枝花市卫生健康委员会归口。

本文件起草单位：攀枝花市中心医院、成都信息工程大学、攀枝花市气象局、四川财经职业学院。

本文件主要起草人：尹立、李咸志、刘顺金、朱茜、王式功、刘金科、雷金莉、苏秋芳、王嘉鑫、张祥健、惠富斐、李永军、李玄、何科、李琳、李成林。

本文件首次制定发布。

气象敏感性疾病发病风险等级预测指南

1 范围

本文件规定了气象敏感性疾病发病风险等级预测的模型构建、等级划分及描述、预警发布的方法。本文件适用于气象敏感性疾病发病风险等级的预警服务。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

T/CI 189—2023 疾病预测模型技术规范
ICD—10 国际疾病分类标准第十次修订版

3 术语和定义

下列术语和定义适用于本文件。

3.1

气象敏感性疾病 meteorological-sensitive diseases

因气象要素变化而诱发或加重的相关疾病，主要包括呼吸系统疾病和循环系统疾病两大类。

注：其中呼吸系统疾病主要包括慢性阻塞性肺疾病（COPD）、支气管哮喘、肺炎、慢性支气管炎、过敏性鼻炎、鼻窦炎等；循环系统疾病主要包括高血压、冠状动脉粥样硬化性心脏病（冠心病）、心力衰竭、心律失常、脑卒中（中风）、心肌梗死等其他经流行病学证实的与气象要素变化相关的疾病。

3.2

气象敏感性疾病发病风险等级 meteorological-sensitive disease onset risk level

因气象要素变化导致气象敏感性疾病潜在发病危险增加的可能性等级。

3.3

就诊人数百分位数 percentile of patient visits

将逐日就诊人数按从小到大顺序排列后，处于特定百分位置的数值。

注：第 p 就诊人数百分位数表示在该时间序列中，至少有 $p\%$ 的观测日就诊人数小于或等于该值，且至少有 $(100-p)\%$ 的观测日就诊人数大于或等于该值。

3.4

疾病预测模型 disease prediction model

一种使用统计或机器学习方法构建的模型，旨在根据一组输入变量（如基因型、生活方式、生物标记、环境因素等）来预测个体在未来某一时期内患某种疾病的概率。

[来源：T/CI 189—2023, 3.2]

3.5

分级准确率 classification accuracy rate

模型预测的风险等级与根据实际就诊人数百分位数划分的真实风险等级之间的一致性程度，即预测正确的天数占测试样本总天数的百分比。

4 模型构建

4.1 数据准备

气象敏感性疾病发病风险等级预测模型的构建需要两类数据作为基础：疾病数据和气象数据。

a) 疾病数据

- 1) 来源：至少 1 家三级医院或当地疾病预防控制中心；
- 2) 内容：逐日就诊人数（按 ICD—10 提取呼吸系统或循环系统疾病就诊数据）；
- 3) 时效：连续 3 年及以上的日就诊人数时间序列。

b) 气象数据

- 1) 来源：当地气象主管机构；
- 2) 要素：逐日气温（平均/最高/最低）、相对湿度、气压、降水量、风速等；
- 3) 时效：与疾病数据同期的连续观测值。

4.2 工作流程

模型构建工作流程包括模型因子选取、建模训练、模型输出、模型性能评估、调整模型参数，具体见图 1。

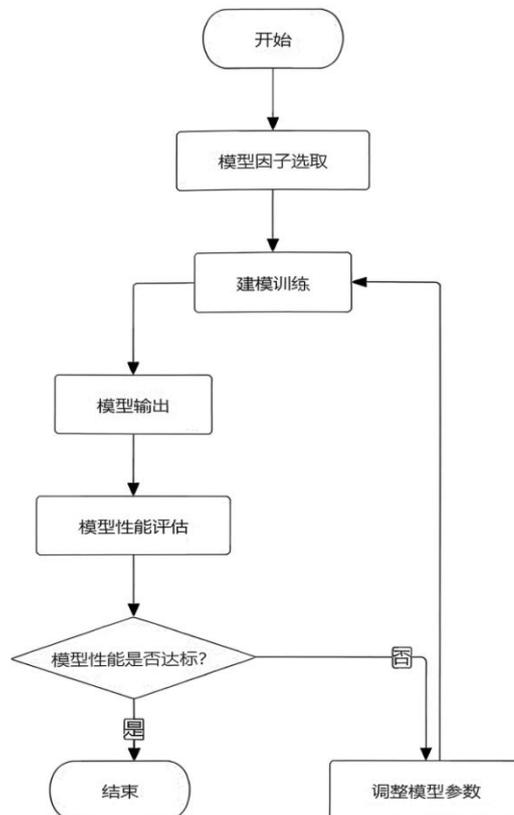


图 1 模型构建工作流程

- a) 模型因子选取方法
 - 1) 统计显著性：通过 Spearman 相关性分析（ $\alpha = 0.05$ 显著性水平），筛选与疾病就诊人数显著相关的因子；
 - 2) 医学合理性：优先选择具有明确病理机制关联的气象要素（如气温突变对心脑血管疾病的影响）；
 - 3) 时序特性：同时考虑即时效应（1 日变化）和累积效应（3 日变化）；
 - 4) 潜在关联性：保留部分未达显著水平但具有潜在医学意义的要素，增强模型泛化能力。
- b) 建模训练方法
 - 1) 数据划分：采用时间序列交叉验证方法，按照时间顺序将数据集划分为训练集、验证集和测试集，确保训练集时间早于验证集，验证集时间早于测试集，以模拟真实业务场景中的预测过程；
 - 2) 模型选择：根据数据特征选用适宜模型，可采用广义相加模型（GAM）、自回归移动平均季节乘积模型（SARIMA）等统计模型处理非线性及滞后效应，或采用随机森林、梯度提升机、长短期记忆神经网络等机器学习模型捕捉复杂交互关系；
 - 3) 训练目标：以最小化预测误差为主要目标，同时兼顾分级准确率，确保模型对极端事件的预测能力；
 - 4) 特征输入：将筛选出的气象因子、星期效应、季节趋势等混杂变量作为输入特征，并进行标准化处理。
- c) 模型输出：模型输出的结果为预测的就诊人数。
- d) 模型性能评估方法
 - 1) 采用均方误差（MSE）、平均绝对误差（MAE）、决定系数（ R^2 分数）指标对模型结果进行量化评估，具体计算方法见附录 A。
 - 2) 在模型性能评估阶段，依据第 5 章定义的等级划分方法将预测的逐日就诊人数映射至预设风险等级（高风险、中风险、低风险、正常四个等级），并要求模型在独立测试集上实现分级准确率 $\geq 70\%$ 的核心目标，分级准确率计算方法见附录 A。
- e) 模型参数调整方法
 - 1) 超参数优化：采用网格搜索、随机搜索等方法，在验证集上对模型超参数进行优化，提升模型泛化能力；
 - 2) 交叉验证：结合时间序列交叉验证，评估不同参数组合下的模型稳定性，避免过拟合；
 - 3) 目标导向：以分级准确率 $\geq 70\%$ 为核心约束，优先优化影响分级性能的关键参数，同时兼顾均方误差（MSE）、平均绝对误差（MAE）等误差指标；
 - 4) 迭代优化：根据模型性能评估结果，动态调整参数搜索范围，直至模型在独立测试集上达到预设性能要求。

5 等级划分及描述

气象敏感性疾病发病风险等级按照预测人数划分为高风险、中风险、低风险、正常四个等级，分别用红色、橙色、黄色和绿色标示，具体见表 1。

表 1 气象敏感性疾病发病风险等级划分方法及描述

风险等级	潜在疾病风险描述	颜色标示	颜色值	划分标准
高风险	气象条件极易诱发或加重该类疾病。	红色	(255, 0, 0)	预测就诊人数 \geq 逐日就诊人数时间序列第 70 百分位数
中风险	气象条件较易诱发或加重该类疾病。	橙色	(255, 165, 0)	逐日就诊人数时间序列第 50 百分位数 \leq 预测就诊人数 $<$ 逐日就诊人数时间序列 70 百分位数
低风险	气象条件对该类疾病略有影响。	黄色	(255, 255, 0)	逐日就诊人数时间序列第 30 百分位数 \leq 预测就诊人数 $<$ 逐日就诊人数时间序列 50 百分位数
正常	气象条件对该类疾病无明显影响。	绿色	(0, 128, 0)	预测就诊人数 $<$ 逐日就诊人数时间序列第 30 百分位数

6 预警发布

- 6.1 当预测发病风险等级为高风险（红色）时发布预警。
- 6.2 预警信息需包含疾病类型、风险时段、风险等级及防护建议。
- 6.3 发病风险等级降至低风险（黄色）及以下且持续 24 小时后解除预警。

附 录 A
(规范性)
模型性能评估方法

A.1 均方误差 (Mean Squared Error, MSE)

均方误差是衡量预测值与真实值之间差异的平方的平均值，是评估模型预测精度的常用指标。均方误差MSE按式(1)计算：

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \dots\dots\dots (1)$$

式中：

MSE —— 均方误差；

n —— 测试样本总数；

y_i —— 第 i 个样本的真实值（通常为就诊人数的历史百分位）；

\hat{y}_i —— 第 i 个样本的模型预测值。

其中，MSE的值越小，表明模型的预测值与真实值之间的总体偏差越小，预测精度越高。该指标对异常值（较大误差）极为敏感，在模型优化过程中，应以降低MSE值为首要目标。

A.2 平均绝对误差 (Mean Absolute Error, MAE)

平均绝对误差是衡量预测值与真实值之间绝对差异的平均值，能直观反映预测误差的平均大小。平均绝对误差MAE按式(2)计算：

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \dots\dots\dots (2)$$

式中：

MAE —— 平均绝对误差。

MAE值越小，表示模型的平均预测误差越小。与MSE相比，MAE对异常值（较大误差）的敏感性较低，其表示的是误差的线性度量。

A.3 决定系数 (R-squared, R^2)

决定系数用于衡量预测模型对目标变量变动的解释程度，即模型拟合优度。决定系数 R^2 按式(3)计算：

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \dots\dots\dots (3)$$

式中：

R^2 —— 决定系数；

\bar{y} —— 样本真实值的平均值。

决定系数 R^2 的取值范围一般为0到1（也可能为负值，表明模型性能差于简单均值预测）。其值越接近1，表明模型对数据变动的解释能力越强，可预测性越高。原则上，用于业务化运行的模型，其 R^2 分数应不小于0.5。

A.4 分级准确率 (Classification Accuracy Rate, CAR)

分级准确率是模型验证阶段的核心性能指标,衡量模型预测的风险等级与根据真实数据划分的风险等级之间的一致程度。分级准确率*CAR*按式(4)计算:

$$CAR = \frac{N_{correct}}{N_{total}} \times 100 \% \dots\dots\dots (4)$$

式中:

CAR ——分级准确率;

N_{correct} ——正确预测的天数,即模型预测的风险等级(高风险、中风险、低风险、正常)与根据该日实际就诊人数所处的历史百分位区间所划分出的真实等级完全一致的天数;

N_{total} ——测试样本中的总天数,即用于模型独立验证的连续测试数据集的总天数,应不少于365天;

该指标直接反映了模型在实际预警应用中的分类可靠性。模型在独立测试集上的分级准确率必须≥70%,方可认为其满足业务化运行的核心性能要求。

参 考 文 献

- [1] GB/T 35221—2017 地面气象观测规范 总则
- [2] WS/T 10037—2025 热浪人群健康风险评估技术指南
- [3] T/CI 189—2023 疾病预测模型技术规范
- [4] DB12/T 1308—2024 脑卒中气象风险预警等级划分规范
- [5] 张楠, 侯斌, 乔丽, 等. 气象因素对心脑血管疾病影响的研究概况[J]. 中西医结合心脑血管病杂志, 2018, 16(09):1193-1196.
- [6] 王琰, 胥美美, 童俞嘉, 等. 基于机器学习的环境监测数据对循环系统疾病死亡影响及预测预警模型构建[J]. 数据分析与知识发现, 2022, 6(10):79-92.
- [7] 兰莉, 哈尔滨市突发极端天气事件人群健康风险预警技术研究. 黑龙江省, 哈尔滨市疾病预防控制中心, 2015-02-13.
- [8] 屈芳. 环境气象因素对呼吸系统疾病影响的研究进展[J]. 气象科技进展, 2013, 3(06).
- [9] 黄伟, 谭宏宇, 方量. 空气污染物与气象因素对慢性阻塞性肺疾病急性加重就诊人次影响的时间序列研究[J]. 实用心脑血管病杂志, 2024, 32(10).
-