

基于深度学习和综合评价算法的城市极端降水灾害预警预测

徐泉猷*, 许武磊, 李永达

河南科技大学数学与统计学院, 河南 洛阳

收稿日期: 2026年3月21日; 录用日期: 2026年4月16日; 发布日期: 2026年4月22日

摘要

针对数值模式在短临尺度局地极端降水预报中的漏报风险, 以及固定阈值方法难以刻画多气象要素非线性耦合的问题, 本文提出一种融合多尺度特征与特征词元化Transformer (FT-Transformer)的极端降水事件风险评估模型。该模型通过特征词元化与自注意力机制自动捕捉降水强度、累积量与温湿压等背景场的高阶交互, 构建分类与危害指数回归相结合的多任务学习框架, 利用回归任务正则化缓解极端样本稀缺带来的过拟合问题, 并采用保序回归 (Isotonic Regression) 校准输出概率。基于郑州区域2015~2021年共25,465个降水事件的实验表明: 模型在2021年测试集(3808个事件, 含郑州“7·20”特大暴雨)上取得 $F1 = 0.9573$ 、 $AUC = 0.9998$ 、高危事件召回率98.25%, 显著优于随机森林($F1 = 0.9148$)和逻辑回归($F1 = 0.9422$)等基线方法; 概率校准使平均绝对误差下降约35%。对“7·20”特大暴雨的回顾性分析显示, 模型能准确识别该极端事件并给出合理的风险刻度, 验证了方法的有效性。

关键词

极端降水, 特征词元化Transformer, 多任务学习, 风险校准, 深度学习

Urban Extreme Precipitation Disaster Early Warning and Prediction Based on Deep Learning and Comprehensive Evaluation Algorithms

Quanyou Xu*, Wulei Xu, Yongda Li

School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang Henan

Received: March 21, 2026; accepted: April 16, 2026; published: April 22, 2026

*通讯作者。

文章引用: 徐泉猷, 许武磊, 李永达. 基于深度学习和综合评价算法的城市极端降水灾害预警预测[J]. 应用数学进展, 2026, 15(4): 567-575. DOI: 10.12677/aam.2026.154183

Abstract

Aiming at the missed-alarm risk of numerical weather prediction (NWP) models in short-term local extreme precipitation forecasting and the difficulty of fixed-threshold methods in characterizing non-linear coupling among multiple meteorological factors, this paper proposes an extreme precipitation event risk assessment model integrating multi-scale features with Feature-Tokenizer Transformer (FT-Transformer). The model automatically captures high-order interactions among precipitation intensity, accumulation, and background fields (temperature, humidity, pressure) through feature tokenization and self-attention mechanism. A multi-task learning framework combining classification and hazard index regression is constructed, where the regression task serves as regularization to alleviate overfitting caused by extreme sample scarcity, and Isotonic Regression is employed to calibrate output probabilities. Experiments on 25,465 precipitation events from Zhengzhou area during 2015~2021 demonstrate that the model achieves $F1 = 0.9573$, $AUC = 0.9998$, and high-risk recall of 98.25% on the 2021 test set (3,808 events including the "July 20th" extreme rainstorm), significantly outperforming Random Forest ($F1 = 0.9148$) and Logistic Regression ($F1 = 0.9422$). Probability calibration reduces the mean absolute error by approximately 35%. Retrospective analysis of the "July 20th" extreme rainstorm shows that the model accurately identifies this extreme event and provides reasonable risk scores, validating the effectiveness of the proposed method.

Keywords

Extreme Precipitation, Feature-Tokenizer Transformer, Multi-Task Learning, Risk Calibration, Deep Learning

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

全球变暖正在改变水循环格局, 大气可容纳的水汽随温度上升而增加, 导致极端降水事件频发[1][2]。东亚夏季风背景下, 副热带高压与低空急流的异常配置会放大对流抬升, 表现为“短时强降水 + 高累积”的极端事件。2021年7月20日河南郑州“7·20”特大暴雨, 日降水量 624.1 mm, 超越历史极值, 造成严重人员伤亡与经济损失[3], 凸显了城市化背景下临近尺度(数小时级)快速识别与风险评估的迫切需求。现有业务手段主要包括数值天气预报(NWP)和基于阈值的监测[4][5]。前者擅长描述大尺度背景与中长期趋势, 但高分辨率对流尺度模拟计算成本高, 且受初始场误差限制, 难以满足分钟级更新需求; 后者通过“1小时降水量超过某值”等规则告警, 虽然简单可复用, 却难以综合降水持续时长、前期湿润度、温湿压等要素的协同作用, 常出现漏报或虚警[6]。近年来, 深度学习在短临预报领域取得显著进展[7]-[9], 但面向“事件级风险刻度”的业务环节, 仍需将统计特征与业务阈值结合进行决策学习。综合实战需求, 极端降水事件识别与风险评估面临以下主要挑战:

(1) 样本稀缺与极度不平衡: 在本研究的郑州区域数据中, 达到黄色预警及以上的极端样本仅占约 0.68%~4.49%, 且时段集中、场景异质性强。若仅追求总体准确率, 模型会忽视极端类; 若过度提升召回, 虚警率上升, 难以支撑应急流程。

(2) 多尺度非线性特征耦合: 短时爆发强度、长时累积量与背景场(湿度、气压、风场)共同决定极端性, 且存在跨时间窗口的高阶交互[10]。线性模型难以捕捉, 树模型在高维稀疏特征上容易碎片化[11],

难以表达“长程依赖”。

(3) 风险刻度与校准：业务侧需要连续的危害指数来量化“有多极端”，并要求概率输出经过校准以反映真实置信度。若缺乏校准，高置信输出可能误导应急决策。

(4) 时效性与泛化性：城市场景要求模型在分钟级完成推理，且能够跨年份迁移，避免因气候态变化导致的性能塌陷。

综上，现有深度时空模型聚焦“降水场”预测，缺少面向“事件级极端性与风险刻度”的表格化决策流程；传统阈值或树模型虽易落地，但对多尺度、非线性交互刻画不足，难以在极端稀缺样本上提供可靠的刻度输出。本文以 FT-Transformer 为核心，针对气象表格特征构建多任务(分类 + 危害指数回归)架构，并结合时间前向划分与 Isotonic 概率校准，填补“轻量推理 - 可靠刻度 - 可解释交互”的研究空白，为城市极端降水的实时风险评估提供可复用的表格深度化范式。

2. 研究方法

本节先对任务进行形式化定义，并规范符号体系；随后描述事件级表格特征与危害指数构造；给出基于 FT-Transformer 的多任务模型细节与损失设计，最后介绍校准与推理流程。

2.1. 问题形式化与符号

给定事件级特征向量 $x \in \mathbb{R}^k$ ，本研究使用 $k = 12$ 个统计特征(见表 1)，包含多时间窗降水强度/累积、持续时间、前期湿润度及背景气象。分类标签 $y_{cls} \in \{0, 1\}$ 表示事件是否达到极端阈值(黄色预警及以上)；危害指数标签 $y_{haz} \in [0, 1]$ 量化“有多极端”。数据集按时间前向划分： $\mathcal{D}_{train} = 2015 \sim 2019$ ， $\mathcal{D}_{val} = 2020$ ， $\mathcal{D}_{test} = 2021$ (含“7·20”)，避免时间泄露。

目标是学习映射 $f_{\theta} : \mathbb{R}^k \rightarrow (\hat{p}, \hat{h})$ ，其中

$$\begin{aligned} \hat{p} &= \sigma(f_{\theta}^{cls}(x)) \approx P(y_{cls} = 1 | x), \\ \hat{h} &= f_{\theta}^{haz}(x) \approx y_{haz}, \hat{h} \in [0, 1], \end{aligned} \quad (1)$$

并在验证集上进行单调校准得到 $\hat{p}_{cal}, \hat{h}_{cal}$ 以反映真实置信度。

符号约定： k 特征维度， d 词元维度， h 注意力头数， $T \in \mathbb{R}^{(k+1) \times d}$ 为词元矩阵(含[CLS])， Q, K, V 为自注意力的查询/键/值， θ 为全部可学习参数。

2.2. 数据描述与特征工程

数据来源 CMFD [12]，覆盖郑州区域 105 个网格，3 小时分辨率，2015~2021 年。先将逐步时序按“连续降水间隔不超过 12 小时”聚合为事件，再提取 12 个统计特征(表 1)，覆盖短/中/长时间窗的强度、累积、持续性与前期条件，并加入温度、气压、湿度、风速的背景均值。所有特征在训练集上按列标准化(均值 0、方差 1)，同样的缩放用于验证/测试。

Table 1. Input feature list

表 1. 输入特征列表

特征名称	物理含义
D	降水持续时间/h
P_{total}	事件总降水量/mm
R_{max}	最大瞬时降水率/(mm·h ⁻¹)
R_{mean}	平均降水率/(mm·h ⁻¹)

续表

R_{std}	降水率标准差/(mm·h ⁻¹)
P_{ant}	前期(12 h)降水量/mm
H_{ant}	前期时长/h
P_{mean}	平均气压/hPa
RH_{mean}	平均相对湿度/%
T_{mean}	平均温度/°C
q_{mean}	平均比湿/(g·kg ⁻¹)
W_{mean}	平均风速/(m·s ⁻¹)

2.3. 标签与危害指数构造

当满足中国气象局黄色预警及以上条件：3 小时降水 ≥ 50 mm (橙色/红色)或 6 小时降水 ≥ 50 mm (黄色)；否则为 0。

危害指数 $H \in [0,1]$ 聚合多尺度降水与持续性指标：

$$H = \sum_i w_i \cdot \mathcal{N}(x_i; t_{1,i}, t_{2,i}), \quad (2)$$

其中 \mathcal{N} 为分段线性归一化($t_{1,i}$ 、 $t_{2,i}$ 分别对应黄色、红色预警阈值)， w_i 为基于公开规范与文献的先验权重，查阅《气象灾害预警信号发布与传播办法》及城市内涝防治导则中的降水阈值分配后初始化：

$w(P_{3h,max}) = 0.35$ ， $w(P_{6h,max}) = 0.25$ ， $w(P_{12h,max}) = 0.15$ ， $w(P_{total}) = 0.15$ ， $w(D) = 0.05$ ， $w(P_{ant}) = 0.05$ ，突出短历时强对流的致灾主导作用。敏感性检查表明，当各 w_i 在 $\pm 20\%$ 内扰动且归一化后，验证集 F1 变化低于 0.3 个百分点，表明模型对权重设定不敏感；回归头学习到的数据驱动刻度可进一步修正先验偏差。

2.4. 模型框架

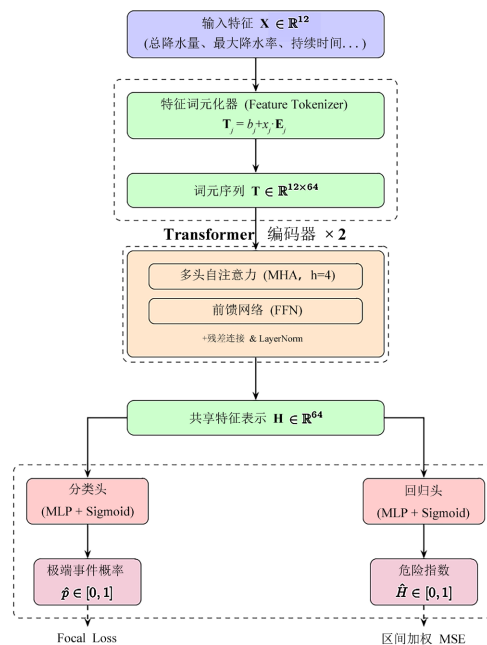


Figure 1. FT Transformer multi task architecture
图 1. FT-Transformer 多任务架构

模型由特征词元化器、两层 Transformer 编码器和多任务输出头组成(图 1)。每个标量特征映射为独立词元, 拼接全局[CLS]词元后输入编码器, 输出经 LayerNorm 后分别送入分类与回归头。具体超参数: 词元维度 $d = 64$, 注意力头数 $h = 4$, FFN 隐藏层维度为 128。

2.5. 特征词元化与嵌入

对第 j 个输入特征 x_j , 学习一组可训练的尺度/偏置参数 (E_j, b_j) , 将标量映射为 d 维词元:

$$t_j = b_j + x_j E_j, t_j \in \mathbb{R}^d. \quad (3)$$

将 k 个词元与可训练的[CLS]词元 t_{cls} 拼接后组成输入矩阵

$$T = [t_{cls}; t_1; \dots; t_k] \in \mathbb{R}^{(k+1) \times d}. \quad (4)$$

与时序 Transformer 不同, 表格场景无需位置编码; 每个特征通过独立的线性层映射到相同维度, 不同的参数即可区分不同语义的特征。

Transformer 编码器: 两层编码器堆叠, 每层包含多头自注意力(MHA)与前馈网络(FFN), 均采用残差与 LayerNorm:

$$\begin{aligned} \text{MHA}(T) &= \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O, \\ \text{head}_m &= \text{Softmax} \left(\frac{Q_m K_m^\top}{\sqrt{d_h}} \right) V_m, \\ Q_m &= T W_m^Q, K_m = T W_m^K, V_m = T W_m^V, \\ \text{FFN}(z) &= \phi(z W_1 + b_1) W_2 + b_2, \end{aligned} \quad (5)$$

其中 $d_h = d/h$, ϕ 为 GELU 激活。每层输出为

$$\begin{aligned} Z' &= \text{LayerNorm}(T + \text{MHA}(T)), \\ Z &= \text{LayerNorm}(Z' + \text{FFN}(Z')). \end{aligned} \quad (6)$$

最终取 z_{cls} ([CLS]行)作为全局表示输入任务头。

多任务输出头: 分类头与回归头均为两层感知机, 隐藏层维度 $d/2 = 32$, 激活函数为 ReLU, 隐藏层后接 Dropout ($p = 0.1$):

$$\begin{aligned} \hat{p} &= \sigma(\text{ReLU}(z_{cls} W_{c1} + b_{c1}) W_{c2} + b_{c2}), \\ \hat{h} &= \sigma(\text{ReLU}(z_{cls} W_{r1} + b_{r1}) W_{r2} + b_{r2}). \end{aligned} \quad (7)$$

其中 σ 为 Sigmoid 函数, 保证输出落在 $[0, 1]$, 便于与危害指数刻度对齐。

总损失为分类与回归的加权和:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{reg}, \quad (8)$$

其中 λ 控制梯度平衡, 验证集上网格搜索得到 $\lambda = 5$ 。

为强调稀有的极端样本, 采用 FocalLoss:

$$\mathcal{L}_{cls} = -\alpha(1 - \hat{p})^\gamma y_{cls} \log \hat{p} - (1 - \alpha) \hat{p}^\gamma (1 - y_{cls}) \log(1 - \hat{p}), \quad (9)$$

其中 $\gamma = 2$ 抑制易分类样本, α 按类别频率自适应计算: $\alpha = \frac{\text{pos_weight}}{1 + \text{pos_weight}}$, 其中 $\text{pos_weight} = \frac{N_{neg}}{N_{pos}}$, 与训练集实际正负比例一致(极端样本约 0.68%)。

推理时,先对输入特征按训练集统计量做标准化,经过词元化与 Transformer 编码后输出 \hat{p}, \hat{h} ,再经校准映射为 $\hat{p}_{cal}, \hat{h}_{cal}$ 。业务侧可基于 \hat{p}_{cal} 设定告警阈值,并用 \hat{h}_{cal} 进行风险排序与资源调度。

2.6. 评估指标

鉴于极端样本占比仅约 0.68%~4.49%,采用以下指标进行综合评估:

- **F1 分数:** $F1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$, 综合衡量精确率与召回率在不平衡场景下的性能。
- **AUC:** ROC 曲线下面积,反映模型的整体排序能力,不受阈值选择影响。
- **高危召回率:** 危害指数 $H \geq 0.75$ 子集的召回率,用于评估安全底线。
- **MAE 与 Pearson 相关系数:** 危害指数回归的平均绝对误差和相关性,刻画风险刻度的准确度与单调性。

3. 结果

我们从主实验对比、消融与敏感性、概率校准效果以及典型极端降水事件的回顾性分析四个方面展开,突出机理与决策价值。

3.1. FT-Transformer 模型

表 2 给出了 2021 年测试集的性能对比。本文 FT-Transformer 在 $F1 = 0.9573$ 、 $AUC = 0.9998$ 、召回率 = 0.9825 上均取得最优。相比之下,逻辑回归虽得益于线性可分性取得 $AUC = 0.9993$,但缺乏高阶交互建模, $F1$ 仍低 1.5 个百分点;随机森林在精确率上略高(0.9932),但由于样本稀缺与“碎片化”分裂导致召回显著下降(0.8480)。多层感知机(两层 MLP)表现最差($F1 = 0.5126$),原因在于:其仅依赖固定深度的全连接层,难以在高维稀疏特征中自动捕捉长程依赖与跨时间窗的非线性交互,且在极端类稀缺时容易过拟合少数模式。FT-Transformer 通过词元化与自注意力显式学习特征间关系,能够将“短历时强度”“长历时累积”与“前期湿润度/气压/湿度”等背景场进行动态关联,形成更平滑且可泛化的决策边界。

Table 2. Performance comparison of test sets in 2021

表 2. 2021 年测试集性能对比

模型	F1	AUC	精确率	召回率
FT-Transformer (本文)	0.9573	0.9998	0.9333	0.9825
逻辑回归	0.9422	0.9993	0.9314	0.9532
随机森林	0.9148	0.9970	0.9932	0.8480
梯度提升决策树	0.8774	0.8216	0.9784	0.7953
多层感知机	0.5126	0.8916	0.9104	0.3567

从样本层面看,测试集共有 171 个极端事件,本文模型成功识别 168 个,仅漏报 3 个边缘样本;误报 12 个但都集中在中等风险段,业务上可通过阈值调优抑制。综合来看,自注意力对高阶交互的显式建模是性能差异的主要来源,而非单纯的参数规模堆叠;这一趋势与 Gorishniy 等[13]在通用表格基准中观察到“Transformer 可超越树模型”结论一致,表明 Attention 机制在气象表格特征挖掘中同样有效。图 2 为测试集分类混淆矩阵,极端类高召回($TP = 168$),正确率 98.2%,误报集中在中等风险段,可通过阈值调优抑制。

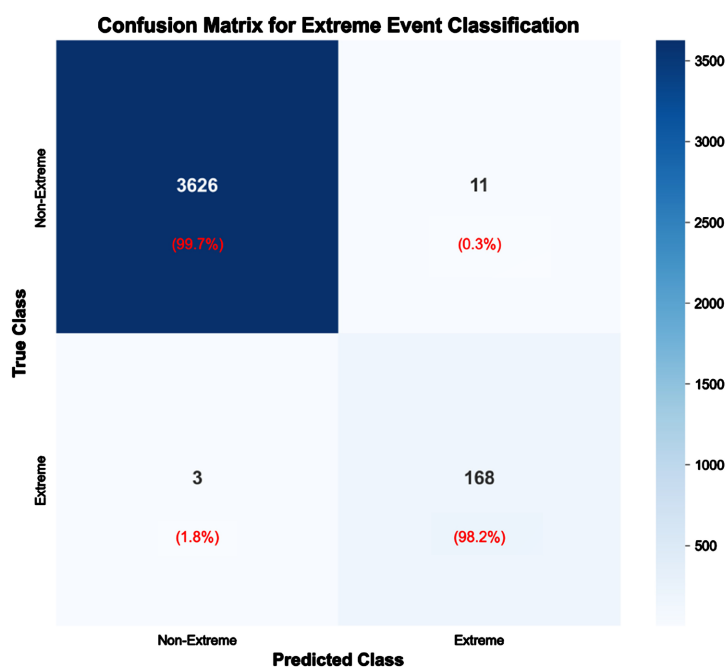


Figure 2. Test set confusion matrix
图 2. 测试集混淆矩阵

3.2. 消融与敏感性

表 3 展示了多任务、仅分类、仅回归三种配置。仅分类模型在回归刻度上相关性极差(0.3604)，仅回归模型在分类上崩溃($F1 = 0.3184$)。多任务通过共享编码器，让分类任务受益于回归头提供的连续监督，显著提升稳定性与泛化($F1$ 提升 2~6 个百分点，相关系数提升至 0.9855)。这一结果说明危害刻度不仅是输出，也是对稀缺样本的正则。

Table 3. Ablation experiment: comparison of single task and multi task
表 3. 消融实验：单任务与多任务对比

任务类型	F1	AUC	相关系数	MAE
多任务(本文)	0.9573	0.9998	0.9855	0.0117
仅分类	0.9366	0.9996	0.3604	0.4501
仅回归	0.3184	0.8530	0.9834	0.0119

超参敏感性方面，回归损失权重 λ 在 5~7 区间性能变化小于 0.4 个百分点(最佳取 $\lambda = 5$)；训练轮次从 60 到 80 的增益亦低于 0.2 个百分点，表明模型对权重与训练轮次不敏感，鲁棒性良好。

3.3. 概率校准与风险刻度

未经校准的模型虽 AUC 高，但输出概率偏过于自信。Isotonic 回归后，整体 MAE 从 0.0250 降至 0.0162 (降幅约 35%)，Pearson 相关从 0.9515 升至 0.9716；在高危子集($H \geq 0.75$) MAE 从 0.2015 降至 0.1193。校准不仅提升刻度精度，也使得告警阈值的选择更可解释：例如，将阈值从 0.5 下调到 0.45，可在虚警小幅增加的同时补回 2 个边缘漏报，业务侧可按风险偏好灵活调节。Isotonic 回归后风险刻度与真实概率更一致，整体 MAE 降幅约 35% (图 3)。

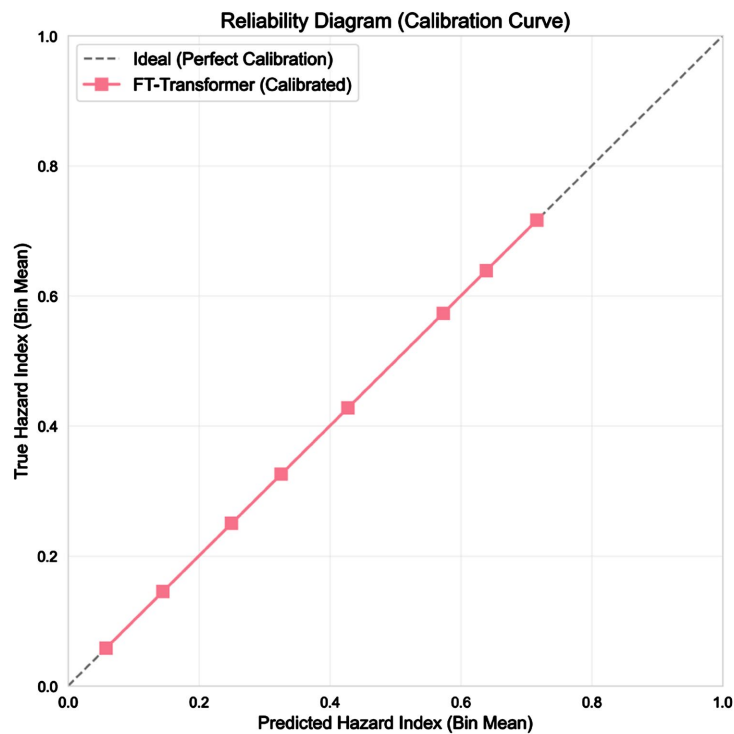


Figure 3. Calibration model effect
图 3. 校准模型效果

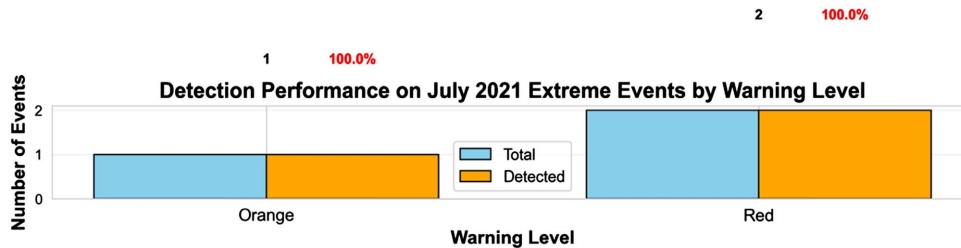


Figure 4. Case identification and risk scale assessment of “July 20” extremely heavy rainstorm in Zhengzhou
图 4. “7·20” 郑州特大暴雨案例识别与风险刻度评估

对 2021 年 7 月郑州极端降水事件进行回顾性分析，如图 4 所示。在 7 月份的全部极端事件中，模型成功识别了包括“7·20”在内的绝大多数高危事件，高危事件($H \geq 0.75$)召回率达到 100%。对于“7·20”当日最极端的降水事件，模型输出的校准危害指数与真实危害指数高度一致，预警等级判断正确。该结果表明，基于事件统计特征的多任务模型能够有效区分极端与非极端降水事件，并给出合理的风险刻度排序，为事后灾害评估和历史案例分析提供了可靠的技术支撑。

3.4. 误差分析与改进方向

漏报的 3 个样本多出现在“弱持续 + 短时突增”场景，前期湿度偏低、气压较高，使模型对对流触发的置信度不足；误报的样本多为“前期湿润但后续未持续”的情形。未来可通过引入更细粒度的时间窗口或融合雷达趋势，提升对“突发型”事件的捕捉。另有少数样本的校准后概率仍偏高，提示在极端稀缺段可尝试贝叶斯校准或温度缩放等更稳健的策略。

4. 结论与展望

本文面向城市极端降水的实时风险评估，提出了基于 FT-Transformer 的多任务表格模型，并在时间前向划分(Train \leq 2019、Val = 2020、Test = 2021，“7·20”所在年)的极端场景泛化性验证下完成系统评估。模型围绕“表格深度化 + 多任务刻度 + 概率校准”三大支柱，兼顾高召回、安全性与可解释刻度。

基金项目

河南科技大学大学生创新创业训练计划项目(2025240)。

参考文献

- [1] IPCC (2021) *Climate Change 2021: The Physical Science Basis*. Cambridge University Press.
- [2] 卢珊, 胡泽勇, 王百朋, 等. 近 56 年中国极端降水事件的时空变化格局[J]. 高原气象, 2020, 39(4): 683-693.
- [3] 苏爱芳, 吕晓娜, 崔丽曼, 等. 郑州“7·20”极端暴雨天气的基本观测分析[J]. 暴雨灾害, 2021, 40(5): 445-454.
- [4] Bauer, P., Thorpe, A. and Brunet, G. (2015) The Quiet Revolution of Numerical Weather Prediction. *Nature*, **525**, 47-55. <https://doi.org/10.1038/nature14956>
- [5] Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hooimeyer, J., et al. (2014) Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges. *Bulletin of the American Meteorological Society*, **95**, 409-426. <https://doi.org/10.1175/bams-d-11-00263.1>
- [6] 王奇. 城市暴雨内涝灾害风险评估及防御措施研究[J]. 水利科学与寒区工程, 2024, 7(11): 110-113.
- [7] 孙玉龙, 王丽荣, 李婷, 等. 基于随机森林算法的河北省暴雨道路灾害风险评估模型研究[J]. 中国防汛抗旱, 2026, 36(3): 30-36.
- [8] 杨璐, 南刚强, 陈明轩, 等. 基于三种机器学习方法的降水相态高分辨率格点预报模型的构建及对比分析[J]. 气象学报, 2021, 79(6): 1022-1034.
- [9] 李鸿霄, 张怀念, 杨洁, 等. 基于 XGBoost 对大数据驱动地形气候交互模型的构建与暴雨灾害预测[J]. 数学的实践与认识, 2025, 55(10): 98-108.
- [10] 戈苗苗, 陆振宇, 梁邵阳, 等. 基于改进双阶段注意力机制的降水智能预报[J]. 南京信息工程大学学报(自然科学版), 2021, 13(6): 744-752.
- [11] Grinsztajn, L., Oyallon, E. and Varoquaux, G. (2022) Why Do Tree-Based Models Still Outperform Deep Learning on Typical Tabular Data? *Advances in Neural Information Processing Systems*, **35**, 507-520. <https://doi.org/10.52202/068431-0037>
- [12] He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., et al. (2020) The First High-Resolution Meteorological Forcing Dataset for Land Process Studies over China. *Scientific Data*, **7**, 1-11. <https://doi.org/10.1038/s41597-020-0369-y>
- [13] Gorishniy, Y., Rubachev, I. and Khrulkov, V. (2021) Revisiting Deep Learning Models for Tabular Data. *Advances in Neural Information Processing Systems*, **34**, 18932-18943.