DOI：10.3724/1000-6915.jrme.2025.0529

基于概率物理信息的机器学习区域滑坡易发性评估

崔红志[1, 2, 3]，裴 特[4]，姬 建[1, 5*]

（1. 绍兴文理学院 深部金属矿智能开采与装备全国重点实验室，浙江 绍兴 312000；2. 绍兴文理学院 浙江省岩石力学与地质灾害重点实验室，浙江 绍兴 312000；3. 扬州大学 土木与交通学院，江苏 扬州 225127；4. 纽约州立大学石溪分校，土木工程系，美国 纽约 11794；5. 河海大学 岩土工程科学研究所，江苏 南京 210024）

摘要：滑坡易发性评估是开展区域尺度潜在滑坡预测的重要手段。现有基于机器学习的数据驱动模型依赖于数据质量且缺乏物理可解释性，在复杂地质与水文条件下易出现泛化能力和物理一致性不足的问题；而基于力学机制的物理模型虽能够表征降雨入渗与边坡稳定性的关系，但在区域尺度应用中仍受参数空间非均质性与适用范围的限制。为解决上述挑战，提出一种概率物理信息机器学习（physics-informed machine learning，PIML）的方法：基于耦合一阶可靠度方法的概率物理模型（PRL-STIM），提取边坡安全系数与失效概率，并将其构建为物理一致性与风险一致性约束项，嵌入深度神经网络的损失函数中，从而实现岩土工程领域知识与数据驱动建模的有机融合。以 2013 年中国甘肃娘娘坝地区降雨诱发的浅层滑坡事件为工程背景，构建高质量样本集并采用空间交叉验证策略，系统评估所提模型在未知区域的泛化能力与预测不确定性。结果表明：PIML 模型在 5 个空间聚类区域中的平均 AUC 较纯机器学习模型提升了 13.6%，科学不一致性指标（scientific inconsistency index，*SI*）降低了 88.6%，显著增强了模型在跨区域滑坡预测中的稳健性、物理一致性与理论解释能力。为区域滑坡易发性建模提供了一种融合物理机制与数据驱动方法的有效路径，验证了在复杂空间环境下引入领域知识对提升模型科学性、泛化能力与工程适用性的可行性，凸显了构建具备岩土工程领域知识的数据驱动模型在滑坡地质灾害预测中的重要意义。

关键词：边坡工程；滑坡易发性评估；物理模型；机器学习；失效概率；一阶可靠度方法

中图分类号：P 642 **文献标识码：**A **文章编号：**1000－6915(2026)03－0790－18

第 一 作 者：崔红志（1993－），现任讲师，主要从事工程地质灾害风险和区域滑坡预测、岩土工程风险分析等方面的教学和研究工作。E-mail：hongzhi.cui@163.com

***通信作者：**姬 建（1983－），现任教授，主要从事工程地质灾害风险和区域滑坡预测、岩土工程风险分析等方面的教学与研究工作。E-mail：JI0003AN@e.ntu.edu.sg

A physics-informed machine learning framework for landslide susceptibility assessment based on probabilistic physically-based model

CUI Hongzhi[1, 2, 3], PEI Te[4], JI Jian[1, 5*]

(1. *State Laboratory of Intelligent Deep Metal Mining and Equipment*, *Shaoxing University*, *Shaoxing*, *Zhejiang* 312000, *China*; 2. *Zhejiang Key Laboratory of Rock Mechanics and Geohazards*, *Shaoxing University*, *Shaoxing*, *Zhejiang* 312000, *China*; 3. *School of Civil Engineering and Transportation*, *Yangzhou University*, *Yangzhou*, *Jiangsu* 225127; 4. *Department of Civil Engineering*, *The State University of New York at Stony Brook*, *New York* 11794, *America*; 5. *Geotechnical Research Institute*, *Hohai University*, *Nanjing*, *Jiangsu* 210024, *China*)

收稿日期：2025－07－21；**修回日期：**2025－10－28

基金项目：国家自然科学基金青年基金 C 类（42507242）；绍兴市基础公益专项项目（2025A13002）；国家自然科学基金联合基金重点项目（U22A20594）
Supported by the National Natural Science Foundation of China (Grant No. 42507242), Shaoxing Municipal Basic Public Welfare Special Project (Grant No. 2025A13002) and Joint Funds of National Natural Science Foundation of China (Grant No. U22A20594)

Abstract: Landslide susceptibility assessment (LSA) is critical for predicting potential landslides at regional scales. Existing machine learning based data driven models are highly dependent on data quality and generally lack physical interpretability, which often leads to limited generalization ability and physical inconsistency under complex geological and hydrological conditions. In contrast, physically-based models in mechanical principles can characterize the relationship between rainfall infiltration and slope stability, but their application at the regional scale remains constrained by spatial heterogeneity of parameters and limited model applicability. To address these challenges, this study proposes a probabilistic physics-informed machine learning (PIML) framework that explicitly incorporates geotechnical domain knowledge into the model training process. A simplified transient infiltration model and the first-order reliability method (PRL-STIM) are utilized to compute the factor of safety and failure probability, respectively. These physics-based outputs are integrated into a composite loss function designed to ensure both physical and risk consistency, thereby guiding the training of a neural network. The framework is validated using a rainfall-induced shallow landslide event that occurred in Niangniangba, Gansu Province, China, in 2013. A high-quality dataset is compiled, and a spatial cross-validation strategy is employed to assess the model′s generalization ability and predictive uncertainty in previously unseen areas. Results from five-fold spatial cross-validation indicate that the PIML model enhances the average AUC by 13.6% and reduces the scientific inconsistency index (*SI*) by 88.6% compared to the baseline model. These improvements demonstrate the proposed model′s enhanced robustness, physical consistency, and interpretability for regional LSA.

Key words: slope engineering; landslide susceptibility assessment; physically-based model; machine learning; probability of failure; first-order reliability method (FORM)

First author: CUI Hongzhi (1993–), assistant professor, is engaged in engineering geological disaster risk and regional landslide prediction, as well as geotechnical engineering risk analysis. E-mail: hongzhi.cui@163.com

***Corresponding author:** JI Jian (1983–), professor, is engaged in engineering geological disaster risk and regional landslide prediction, as well as geotechnical engineering risk analysis. E-mail: JI0003AN@e.ntu.edu.sg

1 引 言

滑坡作为一种严重的自然灾害，已在全球范围内造成了巨大的经济损失和人员伤亡[1]。近年来，极端事件的频发尤其是强降雨和地震活动加剧了滑坡的发生[2-3]。滑坡易发性评估(landslide susceptibility assessment，LSA)作为区域灾害风险管理的重要工具，旨在识别最可能发生滑坡的区域[4]。其基本原理是通过分析地形、地质、气候、人类活动等关键因素与历史滑坡事件之间的关系，量化特定区域的滑坡潜力[5]。当前 LSA 方法主要分为 2 类[6]：(1) 基于物理模型的方法；(2) 基于统计分析或机器学习(machine learning，ML)模型的数据驱动法。

物理模型方法以滑坡力学分析为基础，并结合极限平衡理论与水文地质模型[7]，模拟降雨、地震等触发因素对坡体稳定性的影响。此类方法的优势在于不依赖历史数据，具备良好的可解释性，尤其适用于滑坡记录稀缺的区域[8-9]。然而，由于在区域尺度下地质参数(如土壤厚度、强度指标等)信息不足，物理模型通常依赖经验公式或统计典型值估算输入参数，而忽视了参数内在不确定性，导致滑坡易发性常被高估[10-11]。近年来，姬 建等[12]在区域尺度下采用一阶可靠度方法(first order reliability method，FORM)以高效模拟参数不确定性下的降雨诱发浅层滑坡易发性，并开发了 PRL-STIM 模型。需要注意的是，物理模型的固有简化假设导致其应用仍存在一定的局限性。

数据驱动方法通过学习历史滑坡数据与环境因子之间的统计关系，实现滑坡区识别，尤其在滑坡清单数据丰富的地区表现出较高精度[13]。近年来，以随机森林(random forest，RF)、支持向量机(support vector machine，SVM)、极限梯度提升树(extreme gradient boosting，XGBoost)及深度神经网络(deep neural network，DNN)等为代表的机器学习模型已广泛应用于滑坡易发性评估。多项研究表明，集成学习模型(如 RF 与 XGBoost)在样本量有限或非线性特征强的情况下表现稳健，具有较高的可解释性；

而深度学习模型在多源高维特征融合与空间自适应性方面优势明显，尤其在结合地形、植被及降雨等多时空特征时能显著提高预测精度[14-15]。国际上，美国地质调查局(USGS)提出的基于地形形态特征的概率方法突破了传统清单依赖的局限，实现了无滑坡数据区的可靠易发性预测[16]；意大利萨莱诺大学与水文地质灾害研究所(CNR-IRPI)合作利用深度学习方法构建了降雨与滑坡发生的直接联系，推动了大范围、可操作化的降雨诱发滑坡预报[17]。国内方面，中国地质大学、成都理工大学、河海大学、南昌大学、重庆大学、同济大学及香港科技大学等单位在滑坡易发性机器学习建模、知识图谱构建与多源遥感数据融合等方面取得了重要进展[6, 12, 18-26]。总体而言，当前滑坡易发性研究重点逐步从小样本区域分析转向多源大数据集成与跨区域建模。随着深度学习、集成学习及基准数据集的不断发展，模型的计算精度与自动化程度显著提升。然而，大多数方法仍主要依赖统计关联学习，缺乏物理与工程机制约束，这在一定程度上限制了模型的可解释性与跨区域泛化能力。

此外，数据驱动方法高度依赖滑坡清单的质量，区域尺度上数据的稀疏性、空间不均性显著制约了其应用[27]。特别是，由于区域之间在地貌、地质与降雨等环境条件方面存在显著差异，训练域与测试域之间会出现明显的偏移(domain shift)。当训练数据不足或模型未考虑这种域偏移时，以深度神经网络模型为代表的 ML 模型容易产生与岩土工程机制不一致的预测结果，应用范围受限[28]。需要指出的是，传统 ML 模型的评价指标通常侧重于统计性能，如曲线下面积、准确率等，而忽略物理一致性。例如，T. Pei 和 T. Qiu 等[29]指出，部分表现优异的 ML 模型在物理解释上缺乏合理性，亦未能表征滑坡易发性与安全系数之间的单调负相关性。因此，提出能够兼顾物理一致性与统计性能的混合驱动方法，以提升滑坡易发性预测的科学性与泛化能力迫在眉睫。

近年来，融合领域知识与机器学习的物理信息机器学习方法逐渐成为滑坡预测的重要研究方向。通过将领域知识的物理约束嵌入 ML 模型中，以提升模型预测的泛化能力、可解释性及一致性。T. Pei 等[30]将基于安全系数(F_s)的单调性约束引入 ML 模型，有效提升了 LSA 在未见区域的预测能力；X. Wei 等[31]通过引入无限长边坡模型结果作为 ML 的输入，增强了混合模型在跨区域的泛化能力；H. Z. Cui 等[32]提出了将概率物理模型结果作为卷积神经网络的输入特征，用于增强 LSA 模型在参数不确定性下的稳健性。然而，当前针对基于物理模型与数据驱动方法相结合的研究仍存在三大挑战：(1) 现有物理模型虽然具有理论基础，但常因边界条件理想化、参数单值设定等因素，在区域尺度难以准确适应复杂地质环境；(2) 传统确定性建模方法难以精准量化岩土参数的不确定性，从而难以全面反映潜在滑坡风险；(3) 现有数据驱动模型普遍依赖经验统计关系，缺乏显式物理机制约束，尤其在训练样本不足或分布偏移区域，容易输出与力学原理相悖的预测结果，严重影响模型的泛化性能与工程适用性。

为解决上述问题，本文创新性地提出一种基于概率物理信息引导机器学习的方法(physics-informed machine learning，PIML)。该方法以简化瞬时入渗概率物理模型为基础，通过引入安全系数(F_s)与失效概率(P_f)两种稳定性评价指标，构建基于物理与风险双一致性的损失函数，结合区域尺度的空间交叉验证与不确定性分析，实现滑坡易发性的稳健预测与科学评估。为验证所提方法的实用性，以甘肃娘娘坝区域 2013 年强降雨诱发滑坡为工程背景，探讨 PIML 在区域尺度滑坡易发性评估中的应用潜力。

2　概率物理信息引导机器学习建模

2.1　概率物理模型

降雨诱发的区域滑坡主要表现为浅层滑坡[33]。为揭示此类灾害的发生机制，本文采用前期提出的简化瞬时入渗概率物理模型 PRL-STIM(同样指代下文所有提及的概率物理模型)，通过考虑降雨条件下的无限长边坡模型与一阶可靠度方法，实现对浅层滑坡概率的定量模拟[12]。PRL-STIM 稳定性分析模块聚焦于分析强降雨下非饱和土体中湿润锋的推进，并在无限边坡假设下构建安全系数表达式，以揭示降雨入渗过程对边坡稳定性的影响(见图 1)：

$$F_{\mathrm{s}}=\frac{c_{\mathrm{s}}'}{\gamma_{\mathrm{sat}} z_{\mathrm{w}} \sin\beta\cos\beta}+\frac{\tan\phi'}{\tan\beta}-\frac{\chi u_{\mathrm{w}}}{\gamma_{\mathrm{sat}} z_{\mathrm{w}}}\frac{\tan\phi'}{\sin\beta\cos\beta} \tag{1}$$

式中：γ_{sat} 为土壤饱和密度；β 为山坡角；c_{s}'，ϕ' 分别为土壤的黏聚力和有效摩擦角；z_{w} 为湿润锋深度；χ 为有效应力参数。

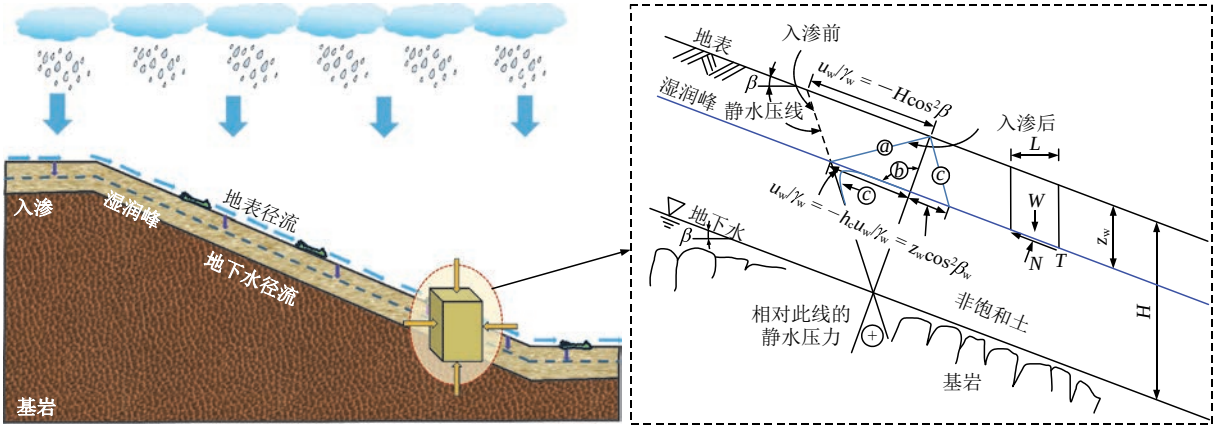


图 1　PRL-STIM 模型示意图(改编自姬　建等[12])

Fig.1　Schematic diagram of the PRL-STIM model(Adapted from J. Ji，et al[12])

该模型考虑了降雨作用下湿润锋深度的动态演化，进而影响孔隙水压力与有效抗剪强度。安全系数 F_s 定义为抗剪强度与剪应力之比，其表达式引入了湿润锋深度 z_w 与孔压参数 μ，并考虑了非饱和土体条件下有效应力的影响。模型中所用入渗量 I 并非等同于降雨强度，而是综合考虑了土壤初始湿度、毛细势梯度、入渗速率等因素的结果。土体厚度是评估边坡稳定性的关键参数之一[34]，本文引入经验型 Z 模型[12]，将土层厚度表示为海拔的函数，从而在区域尺度上高效估算浅层滑动面的深度范围。模型参数依据实测范围设置为 0.1～3.0 m，体现了不同地形单元下的土层空间差异。

除瞬时入渗模块(STIM)外，该模型采用的 GIS-FORM 能够在分析 F_s 基础上，评估土体参数(黏聚力、内摩擦角、坡度)和降雨强度不确定性对边坡稳定性的影响。其中，失效概率 P_f 的计算公式为

$$\left.\begin{aligned} P_f &= \int_{g(\boldsymbol{x})<0} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ g(\boldsymbol{x}) &= F_s(\boldsymbol{x}) - 1 \end{aligned}\right\} \tag{2}$$

式中：$\boldsymbol{x}=(c_s,\ \tan\phi,\ \beta,\ n,\ \gamma_{sat},\ \cdots\cdots)$ 为输入参数的随机变量集合，$f(x)$为随机变量 $\boldsymbol{x}$ 的联合概率分布函数，$F_s(\boldsymbol{x})$ 为安全系数。

由于直接求解该积分在区域尺度上计算量巨大，FORM 方法通过最可能失效点计算可靠度指标 β_f，进而求得近似失效概率：

$$P_f = \Phi(-\beta_f) \tag{3}$$

式中：$\Phi(\bullet)$ 为标准正态累积分布函数。

按照 B. K. Low 和 W. H. Tang[35]的简化岩土可靠度指标概念，β_f 由下式计算：

$$\beta_f = \sqrt{\boldsymbol{n}^{*\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{n}^{*}} = \sqrt{\left[\frac{x_i^{*}-u_i^{\mathrm{N}}}{\sigma_i^{\mathrm{N}}}\right]^{\mathrm{T}}\boldsymbol{R}^{-1}\left[\frac{x_i^{*}-u_i^{\mathrm{N}}}{\sigma_i^{\mathrm{N}}}\right]} \tag{4}$$

式中：x_i^{*} 为第 i 个随机变量的 MPP 值；u_i^{N}，σ_i^{N} 分别为该变量的等效正态平均值和标准偏差；$\boldsymbol{R}$ 为相关矩阵。为在物理空间中高效定位最可能失效点，本文采用 J. Ji 等[6]提出的 GIS-FORM 方法，该方法通过迭代更新变量向量 $\boldsymbol{x}_k$，逐步逼近最可能失效点，其迭代过程如下：

$$\left.\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{\mu}_k^{\mathrm{N}} + \frac{1}{\nabla g(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{T}_k\nabla g(\boldsymbol{x}_k)}[\nabla g(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{x}_k - \\ &\quad \boldsymbol{\mu}_k^{\mathrm{N}}) - g(\boldsymbol{x}_k)]\boldsymbol{T}_k\nabla g(\boldsymbol{x}_k) \\ \boldsymbol{T}_k &= [\sigma_k^{\mathrm{N}}]^{\mathrm{T}}\boldsymbol{R}[\sigma_k^{\mathrm{N}}] \end{aligned}\right\} \tag{5}$$

式中：$\boldsymbol{T}_k$ 为变换矩阵，$\boldsymbol{\mu}_k^{\mathrm{N}}$ 为将随机变量转换为正态分布的等效均值变量，$[\sigma_k^{\mathrm{N}}]$为对角矩阵。

需要指出的是，滑坡易发性[15, 19]表示某一地区在地形、地质和环境条件下发生滑坡的潜力，常用安全系数(F_s)进行确定性的定量描述，采用失效概率(P_f)进行不确定性的定量表征；而滑坡危险性(landslide hazards)[36]通常还需考虑规模、时间和影响范围等要素。

2.2　机器学习模型

相较于随机森林、梯度提升树等传统模型，神经网络具有良好的可拓展性和可微分性，便于通过自定义损失函数引入物理约束与岩土工程先验知识，从而提高模型的物理一致性与泛化能力，故本文采用由多层感知器组成的神经网络模型(neural networks，NN)作为滑坡易发性预测使用的机器学习模型[37]。其中，所建网络包含 3 个隐藏层，每层 36 个神经元，激活函数选用 ReLU，以增强非线性建模能力并加速收敛；输出层采用 sigmoid 函数，实现滑坡与非滑坡的概率二分类。该结构不仅具有良好的拟合能力，也为后续引入基于 F_s 或 P_f 的物理一致性损失提供了便利接口，是构建概率物理信息

神经网络(PIML)框架的重要基础[38]。

2.3 概率物理信息引导机器学习方法

2.3.1 基于岩土工程领域知识的滑坡易发性

在岩土工程领域，边坡与安全系数 F_s 之间存在公认的单调性：即 F_s 值越高，边坡越稳定，发生滑坡的可能性越低。然而，传统的机器学习模型往往难以捕捉这种物理单调性，预测结果缺乏科学解释性。因此，将物理规律纳入模型的训练过程，以考虑物理一致性预测至关重要；此外，滑坡易发性还较大程度上受到输入参数不确定性的影响，即在岩土工程领域认知中，边坡稳定性与失效概率 P_f 亦呈现单调性。通过将 F_s 和 P_f 整合到损失函数中，该模型能够学习基于边坡稳定性评估的确定性和基于风险的不确定性。

需要指出的是，在常规的区域滑坡预测中，很难确保 F_s 与 P_f 与滑坡易发性之间存在严格的一一对应关系，二者之间的映射也存在不确定性。为解决这一问题，本文在建模过程中设计了一种基于软约束(soft constraint)的单调性约束损失函数，通过将 F_s 与 P_f 的单调逻辑嵌入损失函数，对模型输出的滑坡易发性结果进行柔性约束，引导模型预测结果遵循岩土工程领域的基本物理规律。其中，滑坡易发性、F_s 和 P_f 之间的单调关系可表示为

$$\hat{y}_i - \hat{y}_j \leqslant 0 \text{ if } F_{si} \geqslant F_{sj},\ P_{fi} \leqslant P_{fj} \tag{6}$$

式中：$\hat{y}_i$，$\hat{y}_j$ 分别为样本 i 和 j 的预测滑坡易发性值。

为使模型学习边坡稳定性的物理规律，在训练过程中，将每一批训练样本按 F_s 值从高到低进行排序。相邻 2 个样本的模型预测值之差为

$$\Delta\hat{y}_i = \hat{y}_{i+1} - \hat{y}_i \tag{7}$$

由于负值违反了预期的单调趋势(即模型预测更稳定的边坡具有更高的易发性)，因此通过物理一致性损失对这种违反进行惩罚 L_{phy}：

$$L_{phy} = \frac{1}{n-1}\sum_{i=1}^{n-1}\max(-\Delta\hat{y},\ 0) \tag{8}$$

该损失项强化了滑坡易发性与 F_s 之间的预期递减关系，抑制模型将更稳定边坡判定为高滑坡易发性区域，提升了预测结果的物理一致性。

同理，为确定风险一致性，将每批训练样本按照滑坡失稳概率(P_f)从高到低排序，并使用式(7)以相同方式计算模型预测值之间的差异。若某一对样本中 P_f 值较高者对应的模型预测滑坡易发性值反而较低，则违反了 P_f 与易发性之间应有的单调递增关系。为此，本文引入基于风险的一致性损失项，对该类预测不一致性进行惩罚，从而引导模型在学习过程中遵循风险层面的物理逻辑：

$$L_{risk} = \frac{1}{n-1}\sum_{i=1}^{n-1}\max(\Delta\hat{y},\ 0) \tag{9}$$

通过最小化 L_{phy} 和 L_{risk}，该模型会同时保证基于确定性的物理约束和不确定性的风险约束，以确保预测结果符合岩土工程领域知识。

2.3.2 监督损失、物理一致性损失和风险一致性损失的整合

为实现滑坡易发性预测中同时兼顾精度与岩土工程领域知识的一致性，本文在损失函数中引入监督损失、物理一致性损失和风险一致性损失，从而构建出具备科学一致性(scientific consistency，SC)的训练框架。该设计不仅确保模型在数据驱动层面的预测准确性，同时也使预测结果遵循斜坡稳定性的基本物理与概率规律。其中，监督损失函数用于依据标注的滑坡样本优化模型预测结果，其定义如下：

$$L_{data} = \frac{1}{n}\sum_{i=1}^{n} y_i \lg\hat{y}_i + (1-y_i)\lg(1-\hat{y}_i) \tag{10}$$

式中：y_i 为实际观测的滑坡发生情况(标签)，$\hat{y}_i$ 为模型预测的滑坡概率。

为同时引入物理稳定性约束与风险一致性约束，本文提出的 PIML 总损失函数定义为

$$L_{PIML} = L_{data} + L_{SC} \tag{11}$$

式中：L_{SC} 为科学一致性损失，它结合了物理一致性损失和风险一致性损失，即

$$L_{SC} = \lambda_{phy}L_{phy} + \lambda_{risk}L_{risk} \tag{12}$$

正则化系数 λ_{phy} 和 λ_{risk} 分别控制物理约束与风险约束对模型训练的影响程度。λ_{phy} 强化对 F_S 单调性关系的学习(L_{phy})，确保模型遵循基本的边坡稳定性原理。λ_{risk} 强化对 P_f 的响应能力(控制着 L_{risk} 的权重)，使模型输出具有可靠的风险解释。

在此框架下，本文所提出的“引导机制”实质上是将 PRL-STIM 输出的 F_s 与 P_f 嵌入损失函数，通过构造 L_{phy} (约束 F_s —易发性单调递减)与 L_{risk} (约束 P_f —易发性单调递增)两种一致性损失(见式(6)～(9))，共同构成科学一致性损失 L_{SC}。该损失作为对监督损失 L_{data} 的附加软约束在训练过程中引导模型输出遵循岩土工程的物理规律与风险响应关系。该多源一致性损失设计能够实现数据驱动学习与科学规律的有机融合，确保模型在保持预测性能的同时，

具备良好的可解释性和工程适用性。

2.4　PIML 建模流程与科学一致性约束设计

本文提出的概率物理信息机器学习框架如图 2 所示，主要包括 4 个关键步骤：数据准备、模型训练与评估、集成模型构建及滑坡易发性图生成。

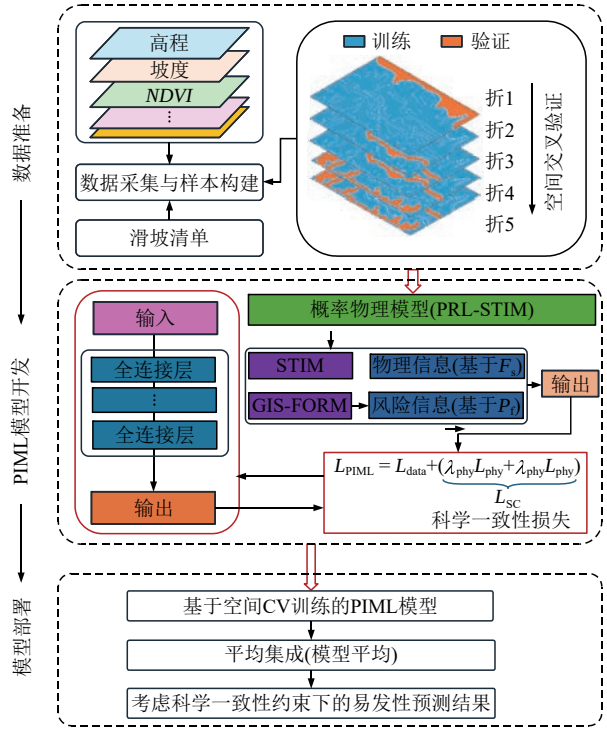


图 2　融合概率物理模型与科学一致性约束的 PIML 滑坡易发性预测流程

Fig.2　Prediction framework of landslide susceptibility based on PIML integrating probabilistic physical model and scientific consistency constraints

2.4.1　数据集准备和空间交叉验证

本文所提出的 PIML 建模框架所需数据包括 2 类：(1) 输入特征(input features)，用于描述边坡稳定的环境背景，主要包括地形与环境因子，如坡度、坡向、地形湿度指数(topographic wetness index，*TWI*)、土壤类型、地质条件、归一化植被指数(normalized difference vegetation index，*NDVI*)等；(2) 监督信号(supervising signals)，包括滑坡/非滑坡标签以及物理与概率指标，即基于 PRL-STIM 计算得到的安全系数(F_s)和失效概率(P_f)，分别代表边坡在给定条件下的稳定性和潜在失稳风险。模型训练过程依据节 2.3.2 所述的损失函数设计，并结合本节描述的空间交叉验证策略实现。滑坡事件、滑坡条件因子及其物理模型参数将在后文进行详细阐述。

滑坡易发性建模本质上是一个二元分类问题，其中已发生的滑坡位置作为正样本。由于滑坡清单通常仅包含正样本记录，本研究通过在远离已知滑坡区域随机采样生成负样本，并结合地理坐标提取环境因子，构建完整的训练数据集。需要特别注意的是，地理空间数据具有显著的空间自相关性(spatial autocorrelation)。传统的 *k* 折交叉验证方法由于假设样本独立同分布，会导致训练集与测试集之间存在空间依赖性，从而高估模型的泛化性能，因此传统的验证方法难以真实反映模型在新区域的预测能力[39]。基于此，本文提出了一种基于空间交叉验证(spatial cross-validation，CV)的改进方法，以解决传统验证方式在滑坡易发性评估中的局限性。

具体而言，首先利用 *k*-means 聚类算法根据样本地理坐标将其划分为 5 个空间簇(见图 3)，在交叉验证过程中确保训练集和验证集来自不同的空间区域，从而有效降低空间依赖性导致的过拟合风险。为验证该方法的有效性，进一步采用 t-SNE 非线性降维技术对特征空间进行可视化分析[40]。与传统随机划分相比，空间交叉验证能更好地保持数据的空间异质性，避免因空间邻近样本同时出现在训练集和验证集而造成的“信息泄露”问题，从而更准确地评估模型在未知区域的泛化性能。

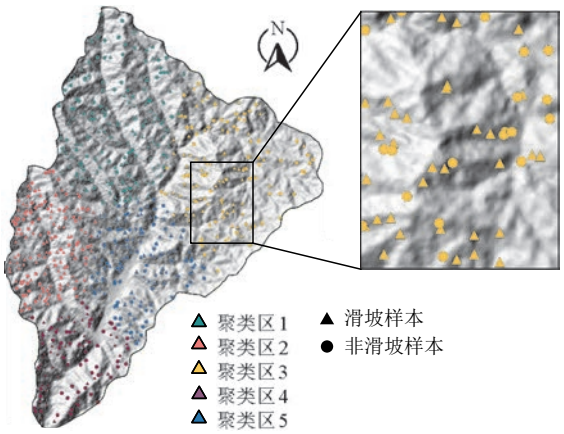


图 3　5 个聚类中滑坡和非滑坡样本的空间分布

Fig.3　Spatial distribution of landslide and non-landslide samples across five clusters

2.4.2 ML 训练、集合建模和滑坡易发性图的生成

为实现滑坡易发性的高效预测，本文基于构建的样本数据训练神经网络模型。在训练过程中，若损失函数中同时引入物理一致性项与风险一致性项，模型即作为物理信息神经网络(PIML)运行；若仅使用监督损失，则作为基准 NN 模型。2 种模型均采用相同的超参数配置：优化器为 Adam，学习率 0.001，训练轮次为 200，以确保对比的公平性。为评估模型在不同地理区域的泛化能力与鲁棒性，本文采用五折空间交叉验证。具体地，每轮交叉验证中，模型在 4 个簇上训练，在剩余一个簇上验证，

从而测试其对未见区域的预测能力。考虑到模型训练的随机性与数据的不确定性，本研究在性能评价阶段对空间交叉验证过程进行了 5 次随机重复，即在相同的五折空间划分下使用不同随机种子训练模型。由此，每一折(fold)均可获得 5 次试验的性能指标，从而更加稳健地评估模型的泛化能力与稳定性。在完成模型评价后，为生成全研究区的滑坡易发性分布图，本研究从 5 次重复试验中选取其中一次完整的 CV 模型集，并对该重复中的 5 个子模型进行集成平均(Ensemble averaging)，得到全区滑坡易发性预测结果。其预测不确定性通过该子模型集之间的预测标准差进行估计。这种方法不仅提升了预测的稳定性，也有助于识别潜在高风险且置信度较低的区域，为风险防控提供更可靠依据。

2.4.3 性能评估

本文采用接收者操作特征曲线(receiver operating characteristic，ROC)及其曲线下面积(area under the curve，*AUC*)评估模型的分类性能。ROC 曲线通过展示不同阈值下的伪阳性率(false positive rate，F_{PR})与真阳性率(true positive rate，T_{PR})关系，反映模型的判别能力；*AUC* 值则作为单一指标衡量整体分类表现，*AUC* = 0.5 表示模型无判别能力，*AUC* = 1.0 表示完美分类。*AUC* 越高，模型区分滑坡与非滑坡样本的能力越强。为实现更全面评估，本文还引入 5 个关键分类指标：准确率(A_c)、精确率(P_{re})、召回率(R_{ec})、F_1 – 分数($F_{1\text{-}Score}$)及科学不一致性指标(scientific inconsistency index，*SI*)。前 4 项为常用统计指标，基于以下 4 类预测结果计算：

$$\left.\begin{aligned} A_c &= \frac{T_P+T_N}{T_P+F_N+T_N+F_P} \\ F_{1\text{-}Score} &= \frac{2P_{re}R_{ec}}{P_{re}+R_{ec}} \\ T_{PR} &= \frac{T_P}{T_P+F_N} \\ F_{PR} &= \frac{F_P}{F_P+T_N} \end{aligned}\right\} \tag{13}$$

式中：T_P 为真阳性，指被正确分类为滑坡的滑坡样本数量；T_N 为真阴性，指被正确分类为非滑坡的非滑坡样本数量；F_P 为假阳性，指被错误分类为滑坡的非滑坡样本数量；F_N 为假阳性，指被错误分类为非滑坡的滑坡样本数量；A_c 为精度，表示被正确分类的样本数占样本总数的比例。

在上述常用机器学习统计指标之外，本文定义科学不一致性指标 *SI* 用于衡量模型预测是否遵循边坡稳定性的岩土工程领域认知，由 L_{phy} 和 L_{risk} 组成：

$$SI = L_{phy} + L_{risk} \tag{14}$$

SI 的降低，表明模型在物理力学规律与风险趋势上的一致性增强。*SI* 与 *AUC* 等传统指标结合使用，可在保证分类精度的同时，系统评估预测结果是否符合地质力学原理，从而兼顾“判别性能”与“科学合理性”，提升其实际应用中的可靠性。

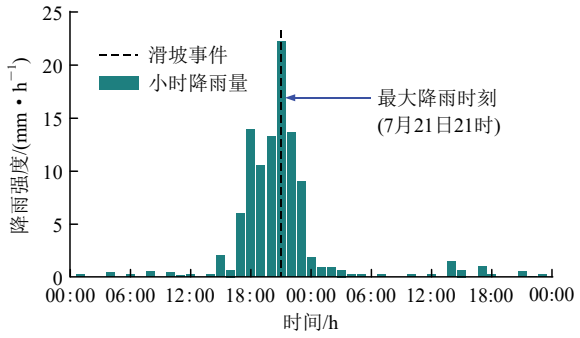
为避免对 *SI* 理解上的歧义，本文进一步对其定义及作用进行澄清与说明。虽然 *SI* 由物理一致性损失(L_{phy})与风险一致性损失(L_{risk})叠加构成，其本质并非模型训练阶段需最小化的目标函数项，而是作为一种独立的后验评价指标，用于衡量模型预测结果是否符合滑坡发生的物理过程与风险。在训练阶段，L_{phy} 和 L_{risk} 作为正则化项引导模型在参数更新过程中遵循特定的单调性约束，从而提升其科学一致性。而 *SI* 的计算则在模型训练完成后进行，通过统计验证集上违反滑坡物理规律(如安全系数 F_s 与失效概率 P_f 之间的单调关系)的预测样本比例，量化模型的解释性与一致性。*SI* 不参与模型训练过程，但依赖训练后模型的预测结果，且与传统统计指标(如 *AUC*，F_1 分数)不同，*SI* 侧重衡量模型在科学合理性方面的表现。因此，较低的 *SI* 值并不直接反映模型总体误差最小化，而是表示模型的科学一致性越高。

3　甘肃娘娘坝降雨诱发山体滑坡案例

3.1　研究区概况及滑坡数据数据

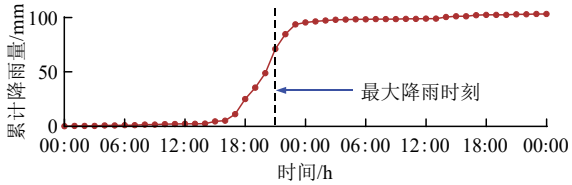
为评估所提 PIML 框架在滑坡易发性评估中的适用性，本文选取中国甘肃省娘娘坝镇东北部小流域作为研究区域。该地位于秦岭山脉西段南麓，地形复杂、生态和地貌类型多样，具有良好的代表性。据报道[41]，该小流域面积约 53.81 km^2，海拔 1 483～2 109 m，高差达 626 m。地质方面，区域地表主要由滑坡堆积物($Q4^{del}$)和冲积堆积物($Q4^{pal}$)构成，冲积物多分布在山谷残积阶地，地层松散、密度较低。土壤层较浅，下伏为更新世马兰黄土、新元古代泥岩与砂岩、鞍山期花岗岩及上泥盆统板岩和石英砂岩。气候上，该区域位于半湿润与半干旱气候交汇带，属温带大陆性气候。1981～2010 年资料显示，年均降雨量约 500.7 mm，相对湿度平均为 66%。其中 85%的降雨集中在 4～10 月，尤其以 7～9 月为主，占全年总降雨量约 68%。强降雨是主要的滑坡诱发因素。

通过比对极端降雨前后的高分三号卫星影像以及谷歌影像发现，本文所关注的研究区滑坡以浅层小型滑坡为主[41]。根据统计分析表明，实际滑坡主要分布在海拔 1 506～2 109 m、坡度 22°～33°的区域，滑动体通常由厚度不足 3 m 的风化岩、泥岩层或含砾土构成，具有明显的空间聚集性。特别是在 2013 年 7 月，一场极端暴雨引发了大规模滑坡事件。此前一个月，娘娘坝镇累计降雨远超多年同期的 3 倍；其中，7 月 21～23 日降雨量达 104.3 mm，最大降雨强度出现在 7 月 21 日 21 时，这也是滑坡爆发的直接诱因(见图 4)。本文使用的物理模型参数详见前期研究结果[12]。



(a) 小时降雨强度与滑坡发生时间对比

(a) Hourly rainfall intensity and comparison with landslide occurrence time



(b) 累计降雨量随时间变化

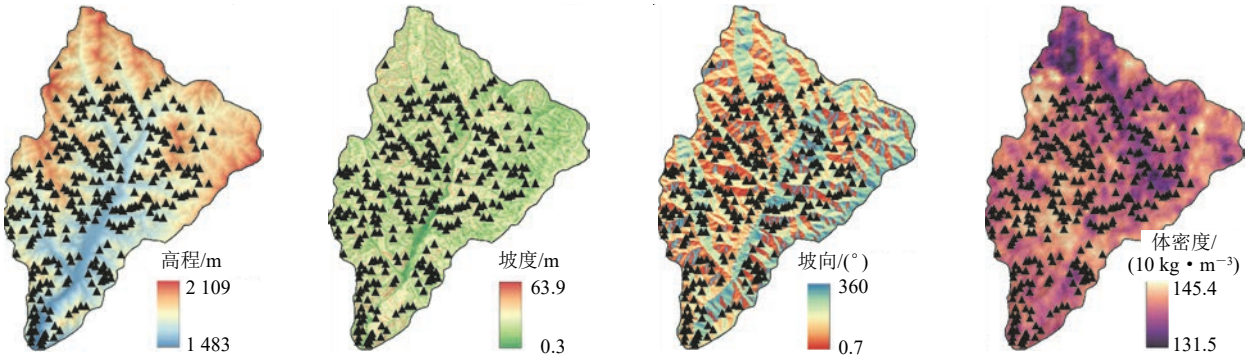
(b) Temporal variation of cumulative rainfall

图 4　2013 年 7 月滑坡事件期间研究区域的小时降雨分布

Fig.4　Hourly rainfall distribution in the study area during the July 2013 landslide event

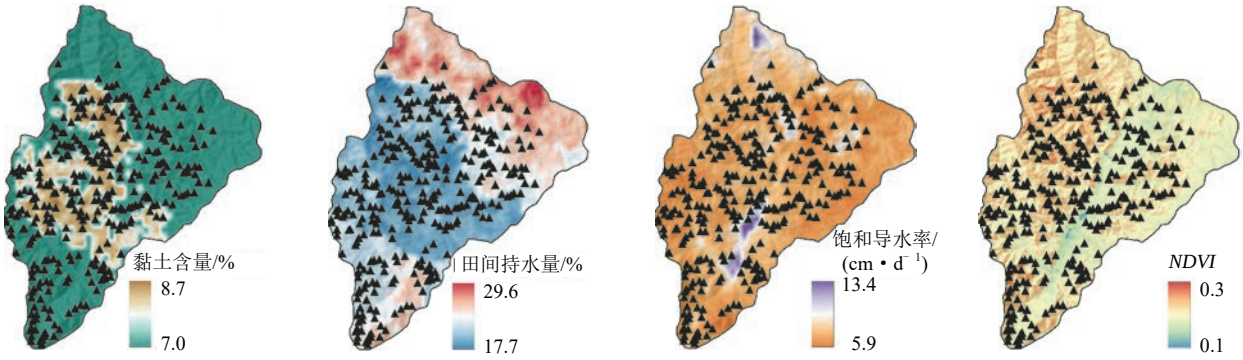
3.2　滑坡条件因子

为准确表征地形、水文与土壤特性对滑坡发生的影响，合理确定条件因子至关重要。如图 5 所示，本文结合研究区的数据可获取性及滑坡活动特征，遴选了 12 个具有代表性的因子，涵盖了 3 类主要控制机制：地形因子(海拔、坡度、坡向平面曲率与剖面曲率)、水文因子(地形湿度指数，*TWI*)、植被因子(*NDVI*)及土壤因子(土壤颗粒组成、容重、饱和导水率 k_s、田间持水量)。其中高程数据来源于 ALOS PALSAR 提供的 12.5 m 空间分辨率数字高程模型(digital elevation model，DEM)；坡度与坡向由 DEM 计算，分别反映坡面的陡峭程度与方位属性；地形



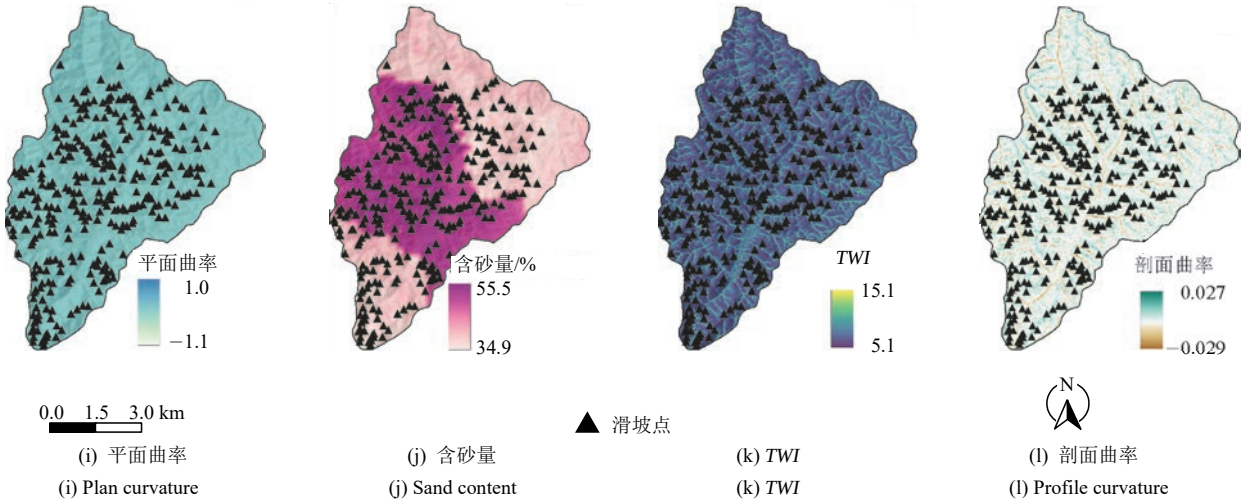
(a) 高程　(b) 坡度　(c) 坡向　(d) 体密度

(a) Elevation　(b) Slope　(c) Aspect　(d) Bulk density



(e) 黏土含量　(f) 田间持水量　(g) 饱和导水率　(h) *NDVI*

(e) Clay content　(f) Field capacity　(g) Saturated hydraulic conductivity　(h) *NDVI*



(i) 平面曲率　(j) 含砂量　(k) *TWI*　(l) 剖面曲率

(i) Plan curvature　(j) Sand content　(k) *TWI*　(l) Profile curvature

图 5　研究区域的滑坡影响因素

Fig.5　Landslide contributing factors for the study area

湿润指数 *TWI* 用于衡量水分在坡面上的积聚潜力，其计算公式如下：

$$TWI = \ln \frac{A_s}{\tan \beta} \tag{15}$$

式中：A_s 为上坡汇水面积，β 为坡度。*TWI* 综合考虑地形汇水能力与坡度，是反映地形对水分积累与渗流路径影响的重要指标。

归一化植被指数 *NDVI* 用于表征地表植被覆盖状况。土壤特征数据(包括砂含量、黏土含量、容重与田间持水量)来自 SoilGrids 全球土壤数据库[42]，该数据库提供 0～200 cm 深度范围内多个标准剖面的高分辨率(250 m)土壤属性估计值。饱和导水率(k_s)数据则引自其衍生产品 HiHydroSoil v2.0。在滑坡易发性分析中，各类土壤因子发挥着关键作用：砂与黏土含量影响渗透性、剪切强度与排水性能；容重反映土体的压实度与孔隙结构，从而影响水分入渗能力与植物根系发育；田间持水量衡量土壤保水能力，以表征长期湿润条件下的水分状态；k_s 代表饱和条件下水分的迁移速率，从而表征地下水运动。

3.3　影响因子相关性分析

在数据驱动的滑坡易发性建模中，输入因子之间的冗余性和潜在共线性可能影响模型的稳定性与预测性能。为此，本文对研究区所选滑坡条件因子进行了皮尔逊相关性分析(见图 6)，其中颜色柱表示各因子之间的相关系数 r 大小，取值范围为 −1～1，为量纲一的量。结果表明，大多数因子相关系数绝对值较低($|r|<0.4$)，整体独立性较强，冗余性对建模影响有限。仅少数因子之间存在显著的相关性：例如，含砂量与黏土含量呈强负相关($r=-0.77$)，反映了土壤质地组成的互补关系；含砂量与田间持水量呈强负相关($r=-0.82$)，体现了砂质土壤持水能力较差的水文学特征；黏土含量与体密度呈中等正相关($r=0.36$)，与土体压实度随黏土含量增加而升高的物理规律相一致。综上，研究区环境因子在整体上未呈现出多重共线性现象，部分因子间较强的相关性亦符合土力学与水文机制。

4　模型性能对比与科学一致性分析

4.1　概率物理模型的性能

图 7 为基于概率物理模型得到的滑坡易发性空间分布图，其中，图 7(a)显示 F_s 较低的区域对应高滑坡易发性，图 7(b)则显示 P_f 较高的区域表征潜在高风险区。结果表明，F_s 低值区与已发生滑坡位置高度吻合，P_f 高值区亦呈现出与实际滑坡分布相符的聚集特征。通过引入合理的物理模型与岩土参数，岩土工程的理论(如无限长边坡稳定性分析与土体力学参数估计)可有效服务于区域滑坡易发性建模。特别是，PRL-STIM 模型所计算的 F_s 与 P_f 值并非最终滑坡易发性预测结果，而是代表区域边坡稳定性(F_s)与概率风险水平(P_f)的两类中间推理变量。

为建立物理力学参数与滑坡空间分布特征之间的定量关联，需构建考虑多源环境因子(详见节 3.2)的滑坡易发性模型。其中，F_s 与 P_f 可作为区域滑坡易发性的物理与风险约束因子，其基于岩土力学可靠性分析的理论基础为机器学习模型提供了先验知

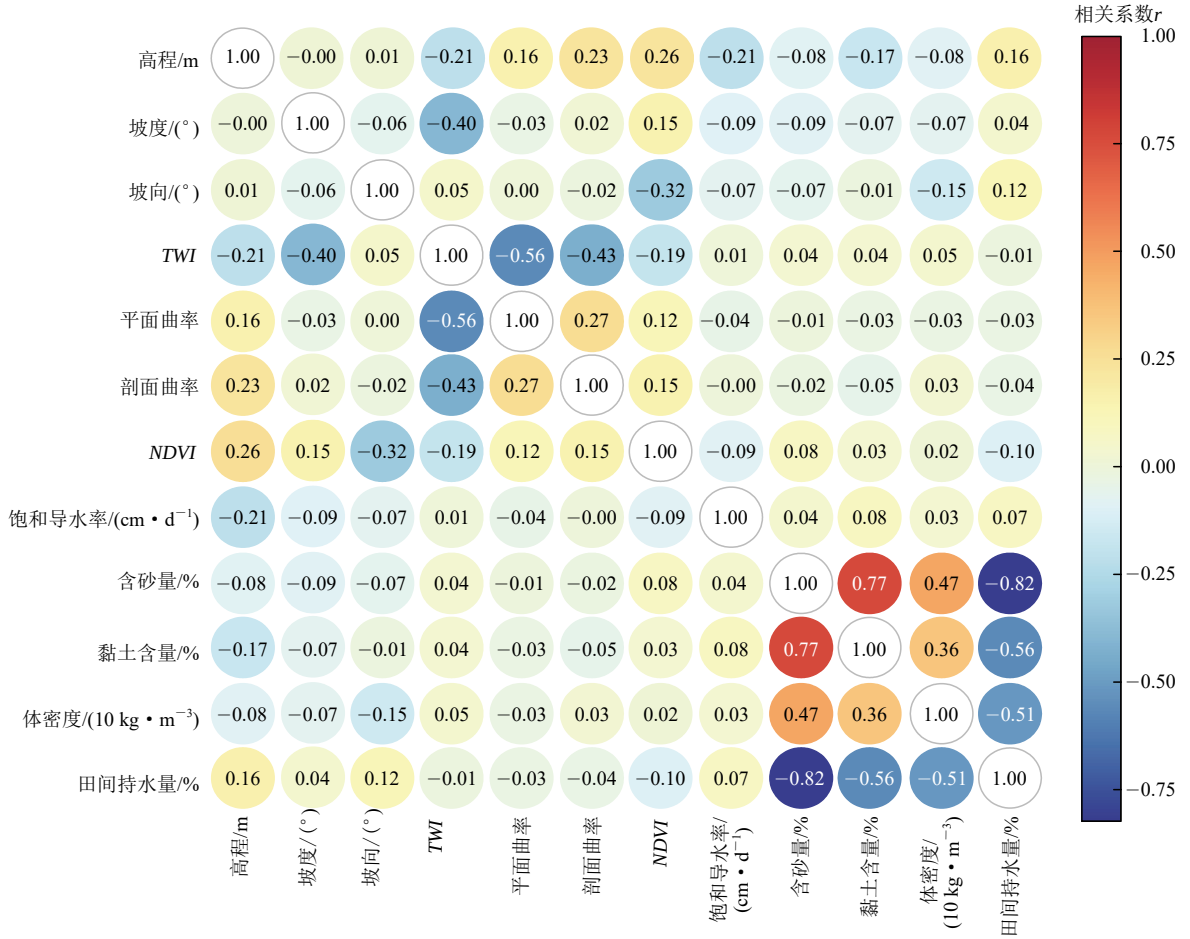
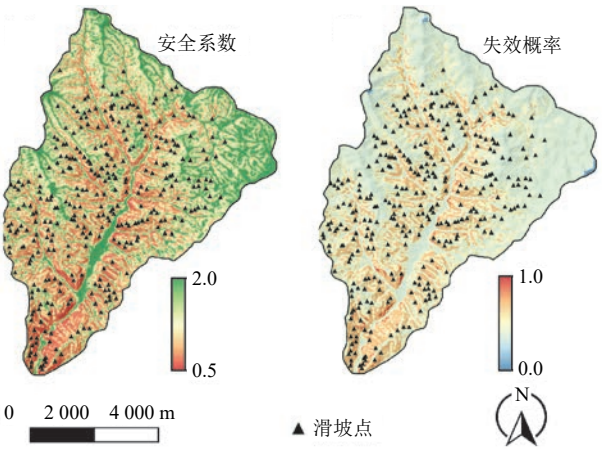


图 6　滑坡条件因子的皮尔逊相关性分析

Fig.6　Pearson correlation analysis of landslide conditioning factors



(a) 基于 F_s 的结果　(b) 基于 P_f 的结果

(a) Derived from the F_s　(b) Derived from the P_f

图 7　基于 PRL-STIM 的研究区域滑坡易发性图

Fig.7　Landslide susceptibility maps for the study area based on the PRL-STIM

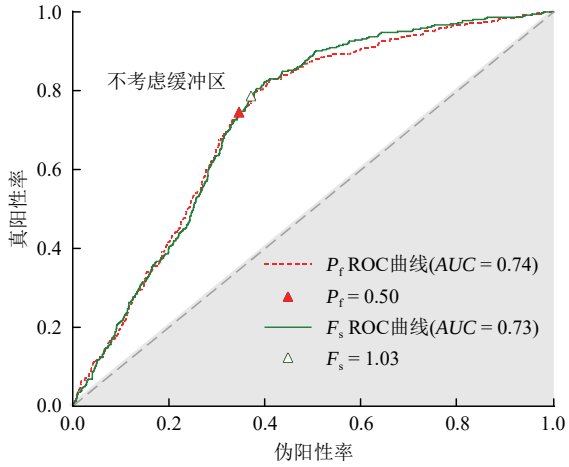


图 8　基于整个数据集的 PRL-STIM 模型性能评估

Fig.8　Performance evaluation of the PRL-STIM model on the entire dataset

识。为进一步量化 PRL-STIM 模型的预测性能，图 8 展示了基于安全系数(F_s)与失效概率(P_f)所得的 ROC 曲线及其对应的最优分类阈值。在全样本上，基于安全系数的 AUC = 0.73，而基于失效概率的 AUC = 0.74。可以看到，ROC 曲线基本一致，但 F_s 与 P_f 分别来源于 PRL-STIM 的极限平衡与 GIS-FORM 计算。尽管在物理意义上高度相关，但本文在建模时并未将二者同时作为独立特征输入 NN，而是分别构建为物理一致性损失(L_{phy})与风险一致性损失(L_{risk})的正则约束，用于校正模型预测结果的合理性，从而避免了直接输入可能导致的多重共线性与过拟合问题。

此外可以看到，模型确定的最优分类阈值分别为 F_s = 1.03 与 P_f = 0.50。从理论上看，若概率物理模型校准合理，则 $F_s \approx 1.0$ 表示边坡处于极限平衡状态，$P_f \approx 0.5$ 也是采用概率方法进行易发性评价常用的分类阈值[43]，即稳定与失稳的分界。J. E. Bowles[44]基于大量边坡失稳案例总结指出，滑坡通常发生于 F_s <1.07，有时发生于 1.07< F_s <1.25，极少发生于 F_s >1.25 的范围内。由此可见，本文通过 ROC 曲线分析得到的阈值 F_s = 1.03 与 P_f = 0.50 均处于该理论区间内，表明概率物理模型在区域尺度上校准合理，能够从数据驱动的角度恢复出与理论期望一致的判别边界，具有良好的物理一致性。这也进一步说明 PRL-STIM 预测结果可为机器学习模型训练过程提供先验信息，从而支撑物理－数据融合建模的实现。

4.2　数据驱动机器学习模型性能分析

为评估纯机器学习模型(采用 NN 模型)在不同聚类区域的泛化能力，本文基于空间 CV 策略开展模型验证，并通过 5 次重复试验以增强评估结果的稳定性与可靠性。图 9 展示了模型在各空间聚类中的预测性能，其中实线表示 5 次重复试验平均值，阴影区域对应标准差：相较于 PRL-STIM(见图 8)，NN 模型整体性能略低，五折空间 CV 的平均 *AUC* = 0.66，且不同折之间性能差异较大，这说明地理异质性影响显著。需要指出的是，空间 CV 的核心目标在于评估模型对未见区域的泛化能力，其分组策略有效规避了空间自相关对评估结果的干扰。*AUC* 得分偏低凸显了 NN 模型对训练数据分布的高度依赖，缺乏对新区域的适应性。这主要源于 NN 模型

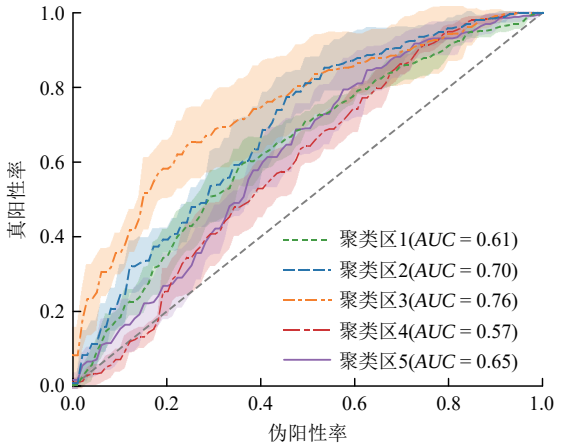


图 9　NN 模型在不同空间聚类区域下基于空间 CV 的性能对比

Fig.9　Performance of NN models based on spatial CV in different cluster areas

缺乏物理机制约束，完全依赖历史样本中所学习的统计特征，一旦该特征模式迁移至新的空间环境，性能便会显著下降。这可从第 4 组看出，相应 *AUC* = 0.57，接近随机分类器水平(*AUC* = 0.5)，几乎无法实现有效判别。综上所述，纯数据驱动模型在应对强空间异质性区域时存在明显的不稳定性与泛化能力不足；相比之下，引入物理约束的模型在跨区域迁移中表现出更强的鲁棒性。

不同于 *AUC* 等传统统计指标，*SI* 指标用于评估模型预测是否违背 2 个关键的物理与概率单调性约束：其中，F_s 源自 PRL-STIM 的极限平衡稳定性结果；P_f 源自 PRL-STIM 的 GIS-FORM 结果，反映边坡失稳的概率。*SI* 值越低，表明模型预测更符合物理力学与风险理论的预期规律；反之，*SI* 值升高则意味着模型在 F_s 较高(更稳定)或 P_f 较低(失稳潜力较低)时仍输出高滑坡易发性，违反基本物理逻辑。这种空间层面的科学不一致性突显出传统机器学习模型在滑坡易发性评估中的局限：尽管其具备较强的统计学习能力，能够从历史数据中提取复杂模式，但由于缺乏显式的领域知识嵌入，难以确保预测结果在物理与概率层面的合理性，这也在地质构造复杂或参数空间异质性显著的区域尤为突出。鉴于此，单纯依赖数据驱动模型可能产生与稳定性理论和失效概率逻辑相悖的预测结果，进而削弱模型的可解释性与滑坡易发性评估的科学性，降低模型在灾害防控决策中的可信度。

在完成神经网络模型的训练与评估后，本文从 5 次空间 CV 重复试验中随机选取其中一次，并将该次试验中 5 个折训练得到的子模型进行平均集成，构建最终的 NN 集合模型。图 10 为此方法生成的滑坡易发性空间分布图，其结果为 0～1 的连续值，分别对应最低与最高滑坡易发性。可以看到，尽管高易发性区域与滑坡清单在空间上表现出较强一致性，但这种表象匹配可能掩盖过拟合问题。一次 CV 验证结果表明，NN 集合模型在独立测试区域的平均 *AUC* = 0.66，明显低于训练集的结果，反映出 NN 模型在不同地理空间上的泛化能力有限。此外，高 *SI* 值进一步表明该模型在科学一致性方面的不足：部分区域预测结果违背了 F_s 所表征的稳定性与 P_f 所反映的风险水平，说明该模型未能遵循物理与概率的单调性原则。

总之，尽管 NN 模型能够从训练样本中学习并生成表面合理的空间预测图，但其在空间泛化能力

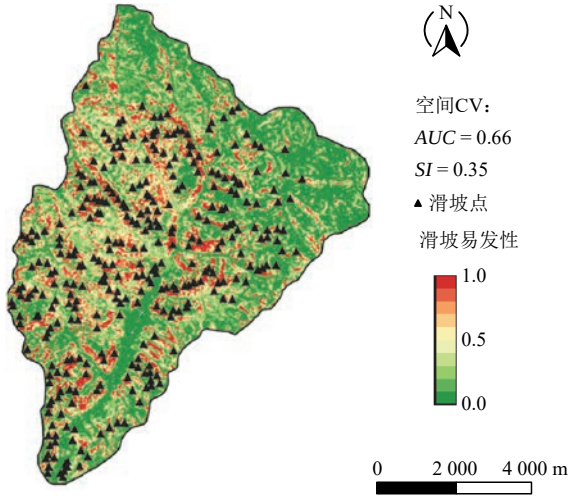


图 10　基于一次空间 CV 重复的 NN 组合模型绘制研究区域滑坡易发性图

Fig.10　Landslide susceptibility map for the study area based on the NN ensemble model from one spatial CV repetition

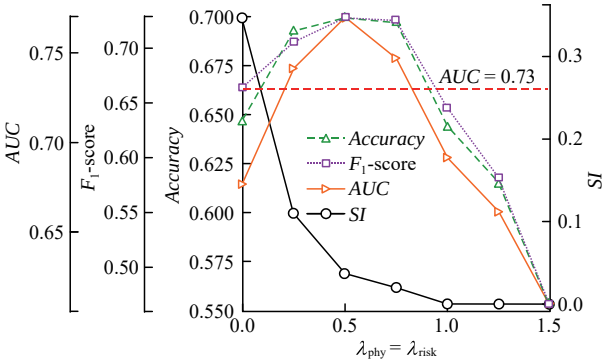


图 11　λ 联合变化下($\lambda_{phy}=\lambda_{risk}$)对 PIML 模型性能的影响

Fig.11　Effect of PIML model performance with varying λ ($\lambda_{phy}=\lambda_{risk}$)

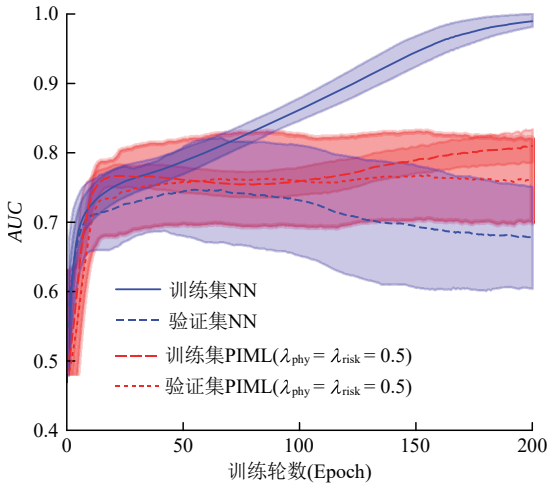
与可解释性方面仍存在明显的局限性。基于纯数据驱动的滑坡易发性评估方法在复杂地质背景下的不足，凸显了引入领域知识以增强模型的稳健性、可解释性和必要性。

4.3　PIML 模型性能分析

4.3.1　科学损失函数的影响

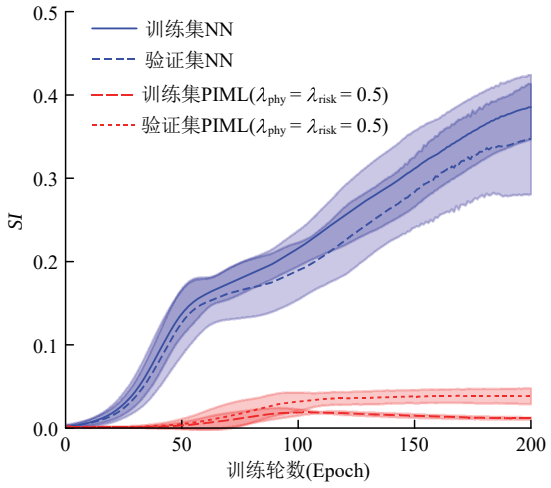
在本文中，正则化参数 λ_{phy} 和 λ_{risk} (式(12))是控制 PIML 模型中物理一致性损失与风险一致性损失权重的关键超参数。图 11 展示了在 $\lambda_{phy}=\lambda_{risk}$ 不同情况下，模型在空间 CV 中的科学不一致性指标(*SI*)与分类性能指标(*Accuracy*，F_1-score，*AUC*)的变化趋势，揭示了领域知识约束对模型科学一致性与分类性能的双重影响：当 $\lambda_{phy}=\lambda_{risk}=0$ (未考虑科学一致性约束)时，PIML 模型退化为传统 NN 模型；随着正则化强度的提升，物理与风险约束的引导作用增强，模型输出逐渐趋于更符合岩土工程原理的预测结果。图 11 显示，当 λ 值增大时，*SI* 指标显著下降，表明模型预测结果更加遵循由 F_s 和 P_f 所定义的单调性关系(即 F_s 越高或 P_f 越低，滑坡易发性越小)；与此同时，分类性能亦同步提升，并在 $\lambda_{phy}=\lambda_{risk}\approx0.5$ 附近达到最优。然而，进一步增加正则化强度将导致模型性能下降，甚至低于 PRL-STIM 与 NN 基准模型。这表明，过强的物理约束可能与训练样本中的标签存在冲突。例如，部分被标注为“非滑坡”的样本，在 PRL-STIM 分析中可能对应较低的 F_s 或较高的 P_f，进而违背单调性约束，导致分类误差。因此，在确保物理一致性的同时，必须在 λ_{phy} 与 λ_{risk} 的设置中合理权衡预测精度与约束强度。结果表明，适度正则化强度下的 PIML 模型在 *SI* 指标上显著优于 NN，同时在分类性能方面也优于 PRL-STIM 与 NN，展现出良好的泛化能力与预测稳定性。

如图 12 所示，在分类性能方面(见图 12(a))，PIML 模型在整个训练过程中验证集 *AUC* 始终优于 NN 模型，平均提升幅度约为 13.6%。虽然 NN 模型在训练集上的 *AUC* 随迭代次数增加不断上升，但在超过 100 轮训练后，其验证集 *AUC* 出现明显下降，表现出典型的过拟合特征；相比之下，PIML 模型在保持较高精度的同时维持了稳定的泛化性能，表明引入物理与风险一致性约束不仅提升了整体预测性能，也有效解决了过拟合问题。在科学一致性方面(见图 12(b))，PIML 模型的 SI 指标在整个训练过程中始终维持在 0.05 以下，而 NN 模型则随训练轮数持续攀升，在 200 轮后接近 0.4，二者差距高达 88.6%。这说明 PIML 的预测结果能够持续遵



(a) 验证集 *AUC* 随训练轮数的变化

(a) Variation of validation *AUC* with training epochs



(b) 科学不一致性指标(*SI*)随训练轮数的变化

(b) Variation of Scientific Inconsistency(*SI*) index with training epochs

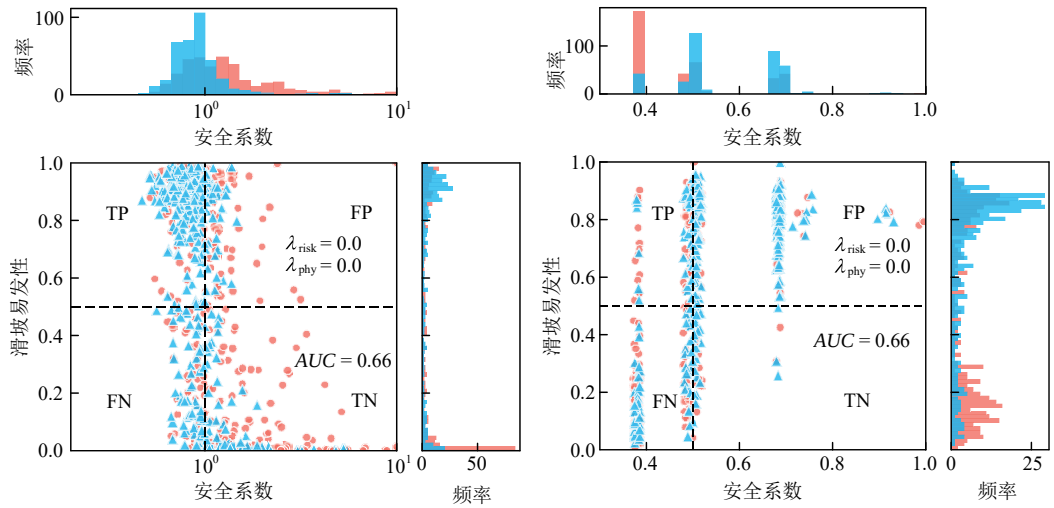
图 12　基于空间 CV 的 NN 与 PIML 模型性能指标(*AUC* 和 *SI*)随训练轮数的变化曲线

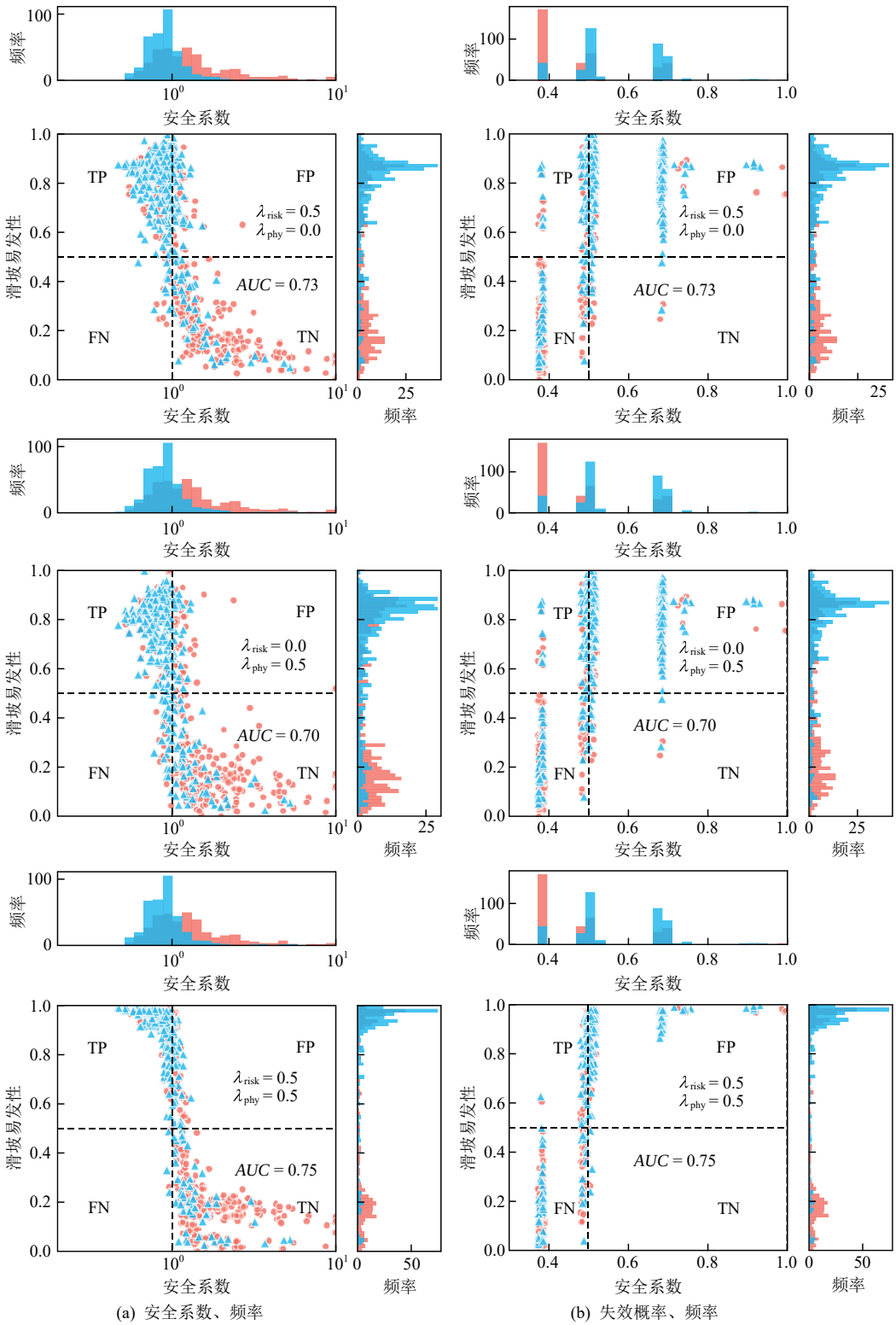
Fig.12　Variation of NN and PIML model performance metrics (*AUC* and *SI*) with training epochs under spatial CV

循“F_s-易发性单调递减与 P_f-易发性单调递增”的物理规律，而 NN 模型随着训练迭代逐渐违背科学一致性，表现出与岩土工程物理逻辑显著背离的趋势。

图 13 展示了边际直方图散点图，用以揭示 PIML 模型预测的滑坡易发性与概率物理模型(PRL-STIM)计算的安全系数(F_s)及失效概率(P_f)之间的关系。为便于分析，每个子图均采用默认阈值划分为 4 个象限($F_s=1.0$，$P_f=0.5$)，分别对应：真阳性(TP：易发性≥0.5 且 $F_s<1.0$，$P_f\geqslant0.5$)、真阴性(TN：易发性<0.5 且 $F_s\geqslant1.0$，$P_f<0.5$)、假阳性(FP：易发性≥0.5 且 $F_s\geqslant1.0$，$P_f<0.5$)与假阴性(FN：易发性<0.5 且 $F_s<1.0$，$P_f\geqslant0.5$)。如图 13 所示，在未引入任何领域知识约束($\lambda_{phy}=\lambda_{risk}=0.0$)时，PIML 模型等价于 NN，虽具备一定区分能力，但其输出与 F_s 缺乏明确的单调关系，表现出明显的科学不一致性。例如，部分 F_s 较低的样本被预测为低易发性(FN)，而 F_s 较高的样本反而被预测为高易发性(FP)；同样，P_f 与预测结果之间亦未体现出预期的风险趋势。对比图 13 各分图可见，单独纳入物理一致性(λ_{phy})或风险一致性(λ_{risk})约束虽能一定程度提升一致性表现，但提升幅度有限，当同时引入两类约束时，模型的科学一致性与预测性能提升更为显著。

此外，图 13 显示大量非滑坡点处于易发性≥0.5、$F_s<1$、高失效概率区域。这是由于非滑坡点采用随机采样策略生成，且排除了陡崖、河道等极端地形位置，并参考基准数据避免落在不可判定区。尽管在 $F_s<1$ 或 $P_f>0.5$ 区间仍存在部分非滑坡点，但这类点代表“潜在不稳定但未发生”的真实边坡状态，符合实际地质环境。此类样本有助于提高模型在复杂环境下的鲁棒性与泛化能力。对比图 11 与图 13 可见，当考虑物理一致性与风险一致性约束后，PIML 模型在空间交叉验证中 *SI* 指标显著下降，表明预测结果更符合岩土工程物理规律，预测性能亦同步提升。这说明本文方法有效改善了 PRL-STIM 在应对空间异质性与样本稀缺条件下泛化能力不足的问题。因此，PIML 模型在融合概率物理模型与数据驱动学习的基础上，兼顾了预测性能与科学一致性，展现出良好的泛化能力与解释能力。同时，该结果也引发了一个思考：纯数据驱动模型的预测可能违背已有领域知识。将物理一致性与风险一致性作为训练目标进行动态评估，是构建稳健、可信的滑坡易发性建模体系的关键步骤。





(a) 安全系数、频率　　(b) 失效概率、频率

(a) Factor of safety，frequency　　(b) Probability of failure，frequency

图 13　科学一致性引导的损失函数 L_{SC} 对模型预测的影响

Fig.13　Influence of the scientific-consistency-guided loss function (L_{SC}) on model prediction

4.3.2 PIML 优化模型的性能

图 14 展示了在正则化参数设定为 $\lambda_{phy}=\lambda_{risk}=0.5$ 条件下，PIML 模型在空间 CV 中的 ROC 结果，其中实线表示 5 次重复试验平均值，阴影区域对应标准差。结果表明，PIML 模型在各空间聚类区域均具备较强的分类能力，平均 $AUC=0.75$，优于概率物理模型(见图 8)与机器学习模型(见图 9)。此外，不同空间交叉折以及重复试验下模型性能波动

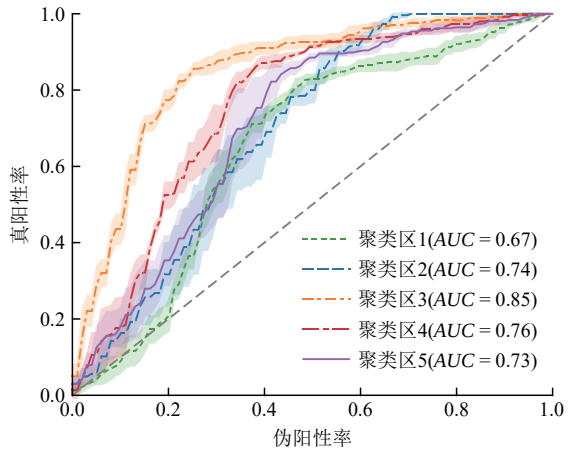


图 14　不同聚类区域基于空间交叉验证的 PIML 模型平均性能($\lambda_{phy}=\lambda_{risk}=0.5$)

Fig.14　Average performance of the PIML model ($\lambda_{phy}=\lambda_{risk}=0.5$) based on spatial CV in different cluster areas

较小，体现出良好的稳定性与区域泛化能力。

表 1 进一步对比了 PIML 与传统 NN 模型在不同聚类区域上的分类性能与科学一致性表现。在 $\lambda_{phy}=\lambda_{risk}=0.5$ 设定下，PIML 模型在绝大多数聚类区域中均表现出更优的综合分类能力。从分类准确率(*Ac*)来看，PIML 在 5 个聚类区域中有 4 个实现了明显提升，整体平均准确率由 NN 的 0.63 提升至 0.70，平均相对提升幅度达 12.35%。其中，聚类区 4 的提升最为突出，准确率由 0.57 显著提高至 0.75，提升幅度高达 31.58%。在 F_1 分数方面，PIML 的优势更加明显：其平均 F_1 值由 0.63 提升至 0.73，平均增幅达到 17.18%。即便在样本分布相对均衡、传统模型已具备一定判别能力的聚类区 2，PIML 仍实现了 5.71%的提升。对于曲线下面积 *AUC*，总体上亦保持显著的提升趋势，相比于 NN 模型，PIML 模型平均 *AUC* 提高了 14.61%，其中聚类区 4 的提升幅度均超过 30%，这也验证了其在复杂地质条件下的适应性与稳健性。在科学一致性方面，PIML 模型在 *SI* 指标上同样优于 NN 模型，平均降低幅度达 88.19%，这意味着模型预测结果更加符合 F_s 与 P_f 所体现的单调性，避免了常见的 NN 模型在“高 F_s 区域却预测高易发性或高 P_f 区域却预测低易发性”的不合理情况。

图 15 为基于集成模型的研究区滑坡易发性分布图。其中图 15(a)为 PIML 模型在 $\lambda_{phy}=\lambda_{risk}=0.0$ 时的均值和标准差，此时 PIML 模型等价于 NN 模型。在未引入任何约束下，模型尽管在训练集上拟合良好，但在空间推广时存在明显过拟合现象、泛化能力有限，尤其是在滑坡清单稀疏或地形条件复杂区域，其预测结果也展示出较高的不确定性。相较之下，图 15(b)，(c)展示了 PIML 模型在引入物理一致性约束($\lambda_{phy}>0$)或风险一致性约束($\lambda_{risk}>0$)后的预测结果，模型不仅在空间分布上更加清晰有序，同时预测的不确定性显著减小，泛化性能显著提升。在 $\lambda_{phy}=\lambda_{risk}=0.5$ 的综合约束下，PIML 集合模型(见图 15(d))表现出最优性能，其对高滑坡易发性区域的预测更加清晰，且不确定性显著收敛，有效克服了传统 NN 模型易产生的过拟合问题。这主要在于本研究采用了空间 CV 策略，训练与测试样本在地理空间上严格分离，尤其在训练阶段未涉及测试区域的地理空间。相反，NN 模型较低的分类性能反映了训练区域的过拟合特征，而本文提出的 PIML 在未见区域上表现更稳健，更能捕捉与滑坡易发性相关的泛化特征，而非局部特征模式。

4.3.3　不同空间区域模型性能差异分析

空间 CV 结果表明，PIML 模型在不同聚类区域中的预测性能存在一定差异(见图 14 与表 1)，其中区域 3 和区域 4 明显优于区域 1 和区域 5。为探讨差异成因，本文从地形地貌特征、滑坡分布及环境因子差异进行分析：区域 3 和区域 4 主要位于研

表 1　数据驱动机器学习模型(NN)与物理信息机器学习(PIML)模型泛化性能比较($\lambda_{phy}=\lambda_{risk}=0.5$)

Table 1　Comparison of model generalization performance between NN and PIML of ML ($\lambda_{phy}=\lambda_{risk}=0.5$)

聚类 (clusters)	准确率 (A_c)			F_1 - 分数 (F_1 - score)			曲线下面积(area under the curve) *AUC*			科学不一致性指标(*SI*) (scientific inconsistency index (*SI*))		
	NN	PIML	*RE*/%	NN	PIML	*RE*/%	NN	PIML	*RE*/%	NN	PIML	*RE*/%
1	0.60	0.66	10.00	0.61	0.67	9.84	0.61	0.67	9.84	0.31	0.05	−83.87
2	0.68	0.66	−2.94	0.70	0.74	5.71	0.70	0.74	5.71	0.39	0.03	−92.31
3	0.67	0.76	13.43	0.59	0.73	23.73	0.76	0.85	11.84	0.31	0.05	−83.87
4	0.57	0.75	31.58	0.58	0.78	34.48	0.57	0.76	33.33	0.43	0.03	−93.02
5	0.62	0.68	9.68	0.66	0.74	12.12	0.65	0.73	12.31	0.33	0.04	−87.88
均值	0.63	0.70	12.35	0.63	0.73	17.18	0.66	0.75	14.61	0.35	0.04	−88.19

注：表中结果来自五折空间交叉验证计算；*RE* 代表“相对误差(relative error)”的英文缩写。

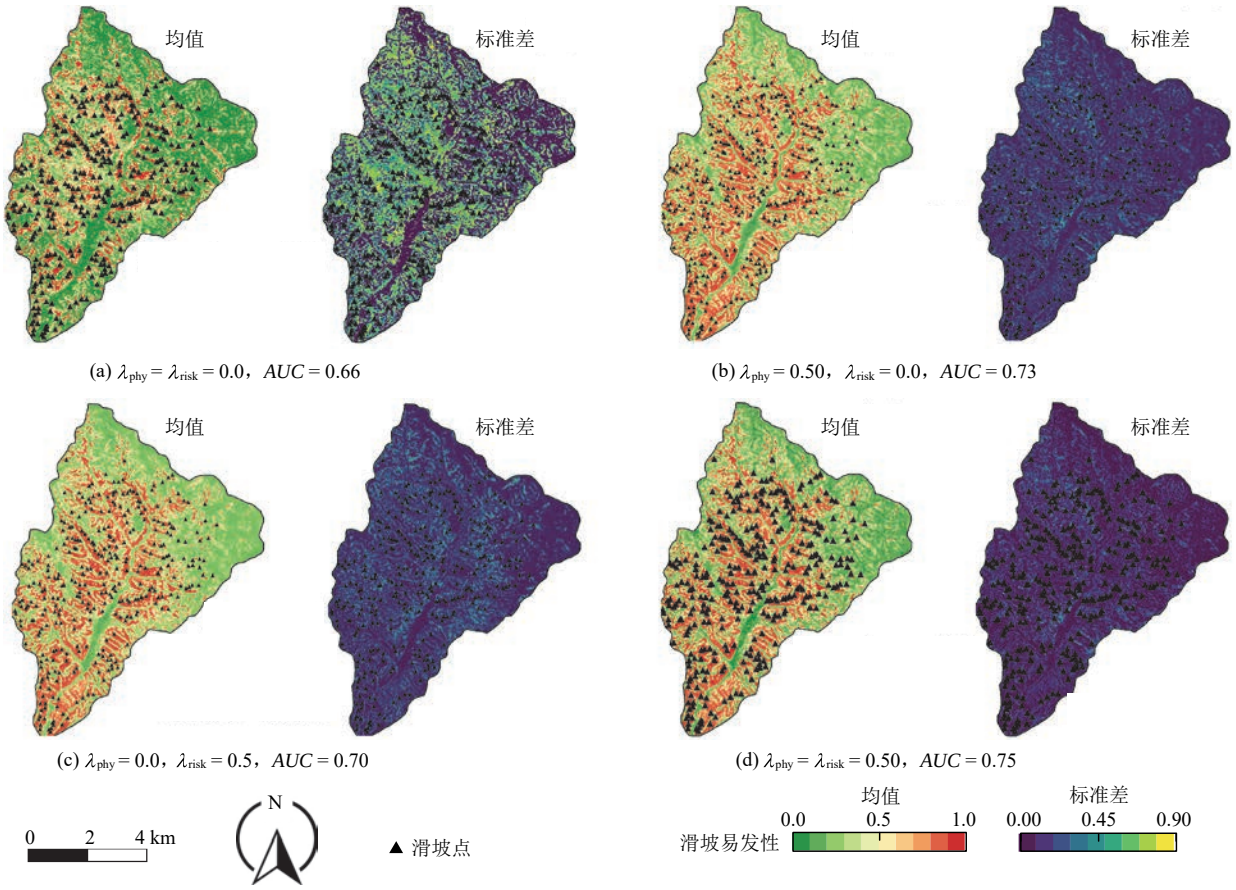


图 15　NN 以及不同配置的 PIML 集成模型滑坡易发性对比

Fig.15　Comparison of landslide susceptibility maps by NN and PIML ensemble models with different configurations

究区中部与南部的丘陵－低山地带地形起伏较大、坡度和曲率变化显著，以红层和页岩为主，土层浅薄、岩土体易风化崩解，滑坡类型相对单一且分布密集，呈空间聚集型，多为浅层降雨型滑坡。此类区域滑坡诱发机制清晰、环境因子与滑坡响应相关性较强，PIML 模型通过 F_s 与 P_f 等约束信息更易识别地形－水文－地质耦合特征，因此表现出更高的精度与一致性。相比之下，区域 1 主要分布于研究区北部的台地或宽谷地带，整体地势平缓，滑坡样本稀少且分布离散；同时，受耕地开发与建设等人类活动影响，滑坡诱发机制复杂多样，环境因子与滑坡响应关系模糊，滑坡与非滑坡样本在特征空间中高度重叠，监督学习受标签不确定性困扰。尽管 PIML 在该类区域性能较 NN 模型有所提升，但受制于环境异制性，物理约束机制的边际作用递减。

综合来看，PIML 模型在不同聚类区域中的表现差异，既反映了其对空间特征的适应能力，也揭示了滑坡发育过程中的空间异质性。结果表明，科学一致性约束在成因清晰、地质条件相对稳定的区域效果更为显著，而在地貌复杂或标签不确定性较高的区域，模型仍存在泛化能力受限的问题。鉴于此，未来研究可在现有 PIML 基础上发展分区建模策略，结合区域地质与地貌类型，构建地质－地貌－机制耦合的多层次模型体系，以进一步提升模型在复杂环境下的稳健性、可解释性与预测精度。

5　结　论

针对区域滑坡易发性评估问题，本文提出了一种融合岩土工程领域知识的概率物理信息机器学习(PIML)方法。与纯数据驱动模型相比，该方法通过引入概率物理模型 PRL-STIM，将边坡安全系数(F_s)与失效概率(P_f)转化为物理与风险一致性约束，并嵌入神经网络损失函数，实现了物理机制与数据驱动方法的有机结合。以 2013 年我国甘肃娘娘坝降雨诱发浅层滑坡事件为案例，主要结论如下：

(1) PIML 在损失函数中引入 F_s 和 P_f 约束，有效避免了纯数据驱动模型中常见的物理不一致现象(如 F_s 升高却预测高易发性、P_f 增大却预测低易发性)，显著增强了模型的物理合理性和可解释性。

(2) 在5个空间聚类区域的交叉验证中，PIML平均AUC达到0.75，较NN模型提升13.6%，准确率和F_1-score分别提高11.1%和15.9%，表明所提方法不仅提高了预测精度，还在地质条件和空间异质性显著区域呈现出稳健性和跨区域泛化能力。

(3) 引入物理与风险一致性约束后，相较于NN模型，科学不一致性(*SI*)降低了88.6%，表明模型预测结果能够更严格地遵循F_s－易发性单调递减与P_f－易发性单调递增的物理规律，进一步凸显了科学一致性提升模型物理合理性与可解释性的作用，确保了滑坡预测的稳定性，降低了跨区域的不确定性。

综上，本文提出的PIML方法在复杂地质条件和样本有限的区域环境下，展现出优异的预测性能、稳健性和物理可解释性。研究结果为解决区域滑坡易发性建模中“物理机制表达不足、关键参数难以获取、空间尺度适应性弱”等关键难题提供了一条融合物理机制与数据驱动方法的实践路径，并在一定程度上为构建具备科学一致性与工程可行性的滑坡预测与区划技术体系奠定了基础。

参考文献(References)：

[1] 宋　健，陆朱汐，谢华威，等. 基于极限平衡原理的地震边坡浅层和深层耦合滑移分析[J]. 岩土工程学报，2023，45(6)：1 141－1 150.(SONG Jian，LU Zhuxi，XIE Huawei，et al. Analysis of coupled shallow and deep sliding of slopes induced by earthquake based on limit equilibrium method[J]. Chinese Journal of Geotechnical Engineering，2023，45(6)：1 141－1 150.(in Chinese))

[2] EMBERSON R，KIRSCHBAUM D，STANLEY T. Global connections between El Nino and landslide impacts[J]. Nature Communications，2021，12(1)：2262.

[3] LÜ Q，WU J，LIU Z，et al. The Fuyang shallow landslides triggered by an extreme rainstorm on 22 July 2023 in Zhejiang，China[J]. Landslides，2024，21(11)：2 725－2 740.

[4] 杨仲康，张世殊，邓建辉，等. 斜坡单元尺寸和滑坡取样方式对区域滑坡易发性评价的影响[J]. 岩石力学与工程学报，2025，44(3)：602－617.(YANG Zhongkang，ZHANG Shishu，DENG Jianhui，et al. The impact of slope unit scale and landslide sampling methods on regional landslide susceptibility assessment[J]. Chinese Journal of Rock Mechanics and Engineering，2025，44(3)：602－617.(in Chinese))

[5] 曾韬睿，邬礼扬，金必晶，等. 基于stacking集成策略和SBAS-InSAR的滑坡动态易发性制图[J]. 岩石力学与工程学报，2023，42(9)：2 266－2 282.(ZENG Taorui，WU Liyang，JIN Bijing，et al. Landslide dynamic susceptibility mapping based on stacking ensemble strategy and SBAS-InSAR[J]. Chinese Journal of Rock Mechanics and Engineering，2023，42(9)：2 266－2 282.(in Chinese))

[6] JI J，CUI H Z，ZHANG T，et al. A GIS-based tool for probabilistic physical modelling and prediction of landslides：GIS-FORM landslide susceptibility analysis in seismic areas[J]. Landslides，2022，19(9)：2 213－2 231.

[7] BAUM R L，SAVAGE W Z，GODT J W. TRIGRS：a Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis，version 2.0[R]. [S. l.]：[s. n.]，2008.

[8] PICIULLO L，CALVELLO M，CEPEDA J M. Territorial early warning systems for rainfall-induced landslides[J]. Earth-Science Reviews，2018，179：228－247.

[9] WHITELEY J S，CHAMBERS J E，UHLEMANN S，et al. Geophysical monitoring of moisture-induced landslides：A review[J]. Reviews of Geophysics，2019，57(1)：106－145.

[10] ZHU D S，XIA L，GRIFFITHS D V，et al. Reliability analysis of infinite slopes with linearly increasing mean undrained strength[J]. Computers and Geotechnics，2021，140：104442.

[11] 郭万里，蔡正银，简富献. 滑坡堰塞体材料空间变异性研究的试验方法[J]. 岩土工程学报，2025，47(10)：2 076－2 085.(GUO Wanli，，CAI Zhengyin，JIAN Fuxian. Test method for studying spatial variability of landslide dam materials[J]. Chinese Journal of Geotechnical Engineering，2025，47(10)：2 076－2 085.(in Chinese))

[12] 姬　建，崔红志，佟　斌，等. 基于物理过程不确定性的降雨诱发浅层滑坡易发性快速区划：GIS-FORM技术开发与应用[J]. 岩石力学与工程学报，2024，43(4)：838－850.(JI Jian，CUI Hongzhi，TONG Bin，et al. Fast zoning of rainfall-induced shallow landslide susceptibility based on physical process uncertainty：development and application of GIS-FORM[J]. Chinese Journal of Rock Mechanics and Engineering，2024，43(4)：838－850.(in Chinese))

[13] CHANG Z L，JIANG S H，HUANG F M，et al. Prediction of overburden layer thickness based on spatial heterogeneity analysis and machine learning models in hillslope regions[J]. Geoscience Frontiers，2025，16(5)：102109.

[14] MERGHADI A，YUNUS A P，DOU J，et al. Machine learning methods for landslide susceptibility studies：A comparative overview of algorithm performance[J]. Earth-Science Reviews，2020，207：103225.

[15] REICHENBACH P，ROSSI M，MALAMUD B D，et al. A review of statistically-based landslide susceptibility models[J]. Earth-Science Reviews，2018，180：60－91.

[16] WOODARD J B，MIRUS B B. Overcoming the data limitations in landslide susceptibility modeling[J]. Science Advances，2025，11(8)：eadt1541.

[17] MONDINI A C，GUZZETTI F，MELILLO M. Deep learning forecast of rainfall-induced shallow landslides[J]. Nature Communications，2023，14(1)：2466.

[18] 窦　杰，向子林，许　强，等. 机器学习在滑坡智能防灾减灾中的应用与发展趋势[J]. 地球科学，2023，48(5)：1 657－1 675.(DOU Jie，XIANG Zilin，XU Qiang，et al. Application and development trend of machine learning in landslide intelligent disaster prevention

and mitigation[J]. Earth Science，2023，48(5)：1 657－1 674.(in Chinese))

[19] MIAO F，ZHAO F，WU Y，et al. Landslide susceptibility mapping in three gorges reservoir area based on GIS and boosting decision tree model[J]. Stochastic Environmental Research and Risk Assessment，2023，37(6)：2 283－2 303.

[20] ZHANG J，ZHU C，XIE L，et al. Comparative study on landslide susceptibility assessment of different models：A case study of alpine mountainous region in Xinjiang[J]. Bulletin of Engineering Geology and the Environment，2025，84(2)：95.

[21] 郭 飞，赖 鹏，黄发明，等. 基于知识图谱的滑坡易发性评价文献综述及研究进展[J]. 地球科学，2024，49(5)：1 584－1 606.(GUO Fei，LAI Peng，HUANG Faming，et al. Literature review and research progress of landslide susceptibility mapping based on knowledge graph[J]. Earth Science，2024，49(5)：1 584－1 606.(in Chinese))

[22] 黄发明，吴敦筱，常志璐，等. 滑坡样本缺失下易发性规律和潜在滑坡识别的易发性-InSAR 多源信息法[J]. 岩石力学与工程学报，2025，44(3)：584－601.(HUANG Faming，WU Dunxiao，CHANG Zhilu，et al. Landslide susceptibility law under the absence of landslide sample and the susceptibility-InSAR multi-source information method considering the potential landslide identification[J]. Chinese Journal of Rock Mechanics and Engineering，2025，44(3)：584－601.(in Chinese))

[23] 仉文岗，何昱苇，王鲁琦，等. 基于水系分区的滑坡易发性机器学习分析方法：以重庆市奉节县为例[J]. 地球科学，2023，48(5)：2 024－2 038.(ZHANG Wengang，HE Yuwei，WANG Luqi，et al. Machine learning solution for landslide susceptibility based on hydrographic division：Case study of Fengjie County in Chongqing[J]. Earth Science，2023，48(5)：2 024－2 038.(in Chinese))

[24] LIU S L，WANG L Q，ZHANG W G，et al. A physics-informed data-driven model for landslide susceptibility assessment in the Three Gorges Reservoir area[J]. Geoscience Frontiers，2023，14(5)：101621.

[25] PENG M，WANG Y，MA C，et al. Landslide susceptibility assessment integrating deep transfer learning and physical models in the Baihetan reservoir area，China[J]. Journal of Rock Mechanics and Geotechnical Engineering，2025，DOI：10.1016/j.jrmge.2025.07.007.

[26] WANG H，WANG L，ZHANG L. Transfer learning improves landslide susceptibility assessment[J]. Gondwana Research，2023，123：238－254.

[27] MEHRABI N，MORSTATTER F，SAXENA N，et al. A survey on bias and fairness in machine learning[J]. ACM Computing Surveys，2022，54(6)：1－35.

[28] PEI T，QIU T. Landslide susceptibility mapping using physics-guided machine learning：A case study of a debris flow event in Colorado Front Range[J]. Acta Geotechnica，2024，19：6 617－6 641.

[29] PEI T，QIU T. Machine learning with monotonic constraint for geotechnical engineering applications：An example of slope stability prediction[J]. Acta Geotechnica，2024，19(6)：3 863－3 882.

[30] PEI T，QIU T，SHEN C. Applying knowledge-guided machine learning to slope stability prediction[J]. Journal of Geotechnical and Geoenvironmental Engineering，2023，149(10)：4023089.

[31] WEI X，ZHANG L L，GARDONI P，et al. Comparison of hybrid data-driven and physical models for landslide susceptibility mapping at regional scales[J]. Acta Geotechnica，2023，18(8)：4 453－4 476.

[32] CUI H Z，TONG B，WANG T，et al. A hybrid data-driven approach for rainfall-induced landslide susceptibility mapping：Physically-based probabilistic model with convolutional neural network[J]. Journal of Rock Mechanics and Geotechnical Engineering，2025，17(8)：4 933－4 951.

[33] 蒋水华，刘 贤，黄发明，等. 融合观测信息的降雨诱发斜坡失稳机理及可靠度分析[J]. 岩土工程学报，2022，44(8)：1 367－1 375.(JIANG Shuihua，LIU Xian，HUANG Faming，et al. Rainfall-induced slope failure mechanism and reliability analyses based on observation information[J]. Chinese Journal of Geotechnical Engineering，2022，44(8)：1 367－1 375.(in Chinese))

[34] SEGONI S，ROSSI G，CATANI F. Improving basin scale shallow landslide modelling using reliable soil thickness maps[J]. Natural Hazards，2012，61(1)：85－101.

[35] LOW B K，TANG W H. Efficient spreadsheet algorithm for first-order reliability method[J]. Journal of Engineering Mechanics，2007，133(12)：1 378－1 387.

[36] ALEOTTI P，CHOWDHURY R. Landslide hazard assessment：summary review and new perspectives[J]. Bulletin of Engineering Geology and the Environment，1999，58(1)：21－44.

[37] TSAI W P，FENG D，PAN M，et al. From calibration to parameter learning：harnessing the scaling effects of big data in geoscientific modeling[J]. Nature Communications，2021，12(1)：5988.

[38] SHEN C，APPLING A P，GENTINE P，et al. Differentiable modelling to unify machine learning and physical models for geosciences[J]. Nature Reviews Earth and Environment，2023，4(8)：552－567.

[39] ROBERTS D R，BAHN V，CIUTI S，et al. Cross-validation strategies for data with temporal，spatial，hierarchical，or phylogenetic structure[J]. Echography，2017，40(8)：913－929.

[40] BELKINA A C，CICCOLELLA C O，ANNO R，et al. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets[J]. Nature Communications，2019，10(1)：5415.

[41] 何简吟. 降雨型浅层滑坡稳定性研究——基于 TRIGRS 模型和 Scoops3D 模型[硕士学位论文][D]. 西安：西北大学，2021.(HE Jianyin. Stability analysis of rainfall-induced shallow landslide——Based on TRIGRS model and Scoops3D model[M. S. Thesis][D]. Xi′an：Northwest University，2021.(in Chinese))

[42] HENGL T，MENDES DE JESUS J，HEUVELINK G. B. M，et al. SoilGrids250m：Global gridded soil information based on machine learning[J]. Plos One，2017，12(2)：e0169748.

[43] MEDINA V，HÜRLIMANN M，GUO Z，et al. Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale[J]. Catena，2021，201：105213.

[44] BOWLES J E. Physical and geotechnical properties of soils[M]. New York，USA：McGraw-Hill，1981：448－449.