

Article

Mitigating Overfitting and Physical Inconsistency in Flood Susceptibility Mapping: A Physics-Constrained Evolutionary Machine Learning Framework for Ungauged Alpine Basins

Chuanjie Yan ¹, Lingling Wu ², Peng Huang ³, Jiajia Yue ⁴, Haowen Li ¹, Chun Zhou ⁴, Congxiang Fan ^{5,6},
Yinan Guo ⁷ and Li Zhou ^{1,4,*}

¹ Institute for Disaster Management and Reconstruction, Sichuan University-Hong Kong Polytechnic University, Chengdu 610065, China; 2025226200009@stu.scu.edu.cn (C.Y.); 2024226200073@stu.scu.edu.cn (H.L.)

² Sichuan Hydrological and Water Resources Survey Center, Chengdu 610036, China; dji_wulingling@163.com

³ Sichuan Road and Waterway Construction Engineering Co., Ltd., Chengdu 610045, China; connection_512@163.com

⁴ State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource & Hydropower, Sichuan University, Chengdu 610065, China; yuejiajia@stu.scu.edu.cn (J.Y.); zhouchun982@gmail.com (C.Z.)

⁵ Sichuan Energy Internet Research Institute, Tsinghua University, Chengdu 610213, China; fancongyang@tsinghua-eiri.org

⁶ Tianfu Yongxing Laboratory, Chengdu 610213, China

⁷ Xizang Autonomous Region Meteorological Information and Network Centre, Lhasa 851000, China; m13989046694@163.com

* Correspondence: zhouli.scu@gmail.com

Abstract

Flood susceptibility mapping in high-altitude ungauged basins faces a structural dichotomy: physically based models often suffer from systematic biases due to uncertain satellite precipitation, whereas data-driven models are prone to overfitting and lack physical consistency in data-scarce regions. To resolve this, this study proposes a Physically constrained Particle Swarm Optimization–Random Forest (P-PDRF) framework, validated in the Lhasa River Basin. The core innovation lies in coupling a hydrological model with statistical learning by utilizing the maximum daily runoff depth as a “Relative Hydraulic Intensity Index.” This approach leverages the topological correctness of physical simulations to circumvent absolute forcing errors. Furthermore, a Physiographically Constrained Negative Sampling (PCNS) strategy and a PSO-optimized “Shallow Tree” configuration are introduced to enforce structural regularization against stochastic noise. Empirical results demonstrate that P-PDRF achieves superior generalization (AUC = 0.942), significantly outperforming standard Random Forest, Support Vector Machine, and Analytic Hierarchy Process models. Ablation studies confirm that the dynamic index outweighs the static Topographic Wetness Index in feature importance, effectively correcting topographic artifacts where static models misclassify arid depressions as high-risk zones. This study offers a scalable Physics-Informed Machine Learning solution for the global “Prediction in Ungauged Basins” initiative.

Keywords: physics-guided machine learning; flood susceptibility; ungauged basins; BTOP model; PCNS strategy; structural regularization



Academic Editor: Gwo-Fong Lin

Received: 5 February 2026

Revised: 10 March 2026

Accepted: 23 March 2026

Published: 7 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

Floods are among the most destructive natural disasters globally, and the intensified hydrological cycle driven by climate change has significantly amplified this threat [1,2]. These extreme events pose severe challenges to human safety, infrastructure integrity, and ecological stability, accounting for a high proportion of economic losses related to natural disasters [3]. Consequently, the development of high-precision Flood Susceptibility Mapping (FSM) has become a prerequisite for comprehensive flood risk management, territorial spatial planning, and the development of early warning systems [4]. Accurately identifying high-risk areas is crucial for optimizing the allocation of disaster mitigation resources, ensuring the precise implementation of defense measures [5], and thereby effectively reducing the loss of life and property caused by extreme hydrological disasters.

Despite the urgent need, reliable flood disaster assessment faces tremendous challenges in high-altitude mountainous regions, exemplified by the Qinghai–Tibet Plateau (QTP) [6]. Such environments possess extremely complex topographic features and highly heterogeneous precipitation patterns, while suffering from a chronic scarcity of ground hydro-meteorological observational data [7]. This “ungauged” dilemma is the core concern of the “Prediction in Ungauged Basins (PUB)” initiative launched by the International Association of Hydrological Sciences (IAHS) [8]. In data-scarce basins, traditional hydrological modeling is subject to significant uncertainty regarding parameter calibration and the quality of forcing data; meanwhile, complex nonlinear interactions between topography and saturation-excess runoff mechanisms further increase the difficulty of susceptibility delineation, rendering static assessment methods based on simple topographic factors often ineffective [9]. Traditional static indices, such as the Topographic Wetness Index (TWI), rely on steady-state assumptions and uniform precipitation, failing to capture the spatiotemporal heterogeneity of snowmelt and flow routing dynamics in alpine catchments. Consequently, static assessment methods often yield physically inconsistent results in complex terrain.

Historically, the methodology of flood susceptibility assessment has evolved along two distinct trajectories: process-based physical methods and data-driven machine learning (ML) techniques. Physically based models, such as distributed hydrological–hydrodynamic models, provide strong physical interpretability by simulating flow routing processes and mass conservation mechanisms [10,11]. However, their application in large-scale, ungauged basins is often constrained by high computational costs and systematic bias caused by satellite precipitation input errors [12]. Conversely, Multi-Criteria Decision Making (MCDM) methods like the Analytic Hierarchy Process (AHP), while relying on expert knowledge, are significantly affected by subjectivity and struggle to capture complex nonlinear interactions, often leading to misjudgments in low-lying arid areas [13].

In recent years, machine learning algorithms, represented by Support Vector Machines (SVM) and Random Forest (RF), have become the mainstream paradigm in this field due to their superior ability to model the nonlinear relationships between environmental factors and flood occurrence [14,15]. Although these data-driven models usually outperform traditional methods in prediction accuracy [16], they are often criticized as “black boxes” lacking physical constraints [17]. Particularly when training data is limited or samples are imbalanced, pure data-driven models are prone to overfitting, tending to memorize spatial coordinates or random noise rather than learning generalizable hydrological process laws [18].

To bridge this gap, the academic community has begun to explore the synergy between process understanding and data science. Current Physics-Informed Machine Learning (PIML) research typically follows two paradigms: embedding physical laws (e.g., partial differential equations) directly into the loss function of neural networks (Soft Constraints)

or integrating physical process outputs as informative priors into the feature space of statistical models (Physics-Guided Feature Engineering, PGFE). Given the discrete nature of tree-based algorithms (e.g., Random Forest) which precludes gradient-based loss function modification, this study adopts the PGFE paradigm. We construct a hybrid framework where the BTOP hydrological model provides the causal structure (flow routing topology), and the machine learning model handles the residual mapping, thereby achieving physical consistency without altering the solver's internal architecture [19]. For example, Ref. [20] demonstrated that integrating physical prior knowledge can significantly enhance the hydrological prediction capability of deep learning models; Ref. [4] attempted to introduce hydrological indices as covariates into statistical models to improve generalization. However, a critical logical trap remains in current PGFE applications for ungauged basins. Most hybrid studies treat physical model outputs (e.g., simulated inundation) as 'ground truth' features, ignoring a fundamental reality: in data-scarce regions, physical simulations typically suffer from significant amplitude biases due to the lack of gauge-corrected precipitation forcing. Feeding these biased physical values directly into ML models introduces 'systematic noise,' while pure ML models struggle to extract robust hydrological laws from sparse historical disaster records ($N < 200$). Consequently, a methodology that can extract topological reliability from imperfect physical simulations to constrain ML overfitting in data-vacuums is largely absent. Specifically, three critical gaps remain:

- (1) The 'Accuracy Paradox' in Ungauged Basins: In regions lacking ground monitoring, hydrological models driven by satellite precipitation (e.g., TRMM/CMFD) inevitably exhibit systematic amplitude errors (underestimation or overestimation). Existing frameworks lack a transformation strategy to convert these 'quantitatively inaccurate' absolute fluxes into 'qualitatively reliable' relative indicators (e.g., temporal phasing and flow routing paths), leading to a risk of physical bias propagation.
- (2) Trivial Learning in Sampling: Traditional random negative sampling strategies often draw non-flood points randomly from the entire domain (including vast high-altitude mountains). This practice easily leads to the "dominance of simple negative samples"—where the model only needs to learn the simple topographic rule that "high altitude equals safety" to achieve high accuracy, thus falling into the trap of trivial learning. The lack of targeted "Hard Negatives" (such as low-lying but safe terraces) makes it difficult for the model to finely distinguish between "flood-prone floodplains" and "safe zones" within topographically similar valley regions, leading to blurred decision boundaries [21].
- (3) Overfitting Dilemma in Small-Sample Scenarios: Alpine cryosphere basins typically suffer from a severe scarcity of historical flood records (Sample Size $\ll$ Feature Dimension). In such data-starved environments, unconstrained ML models (e.g., deep decision trees) tend to memorize high-frequency stochastic noise rather than learning generalizable laws. There is a lack of adaptive structural regularization strategies (specifically tailored for small datasets) to enforce the Principle of Parsimony, ensuring spatial continuity.

To address these limitations, this study proposes a novel Physically constrained Particle Swarm Optimization–Random Forest (P-PDRF) framework. This method deeply couples the semi-distributed BTOP with structurally optimized machine learning algorithms. The core strategies include introducing the "Relative Hydraulic Intensity Index" derived from BTOP to provide dynamic physical constraints, and proposing a Physiographically Constrained Negative Sampling (PCNS) strategy to purify training data. Crucially, the proposed P-PDRF framework is not a generic improvement but a targeted solution for the 'Ungauged' constraint: we utilize the relative Q_{depth} to bypass absolute forcing errors,

employ PCNS to resolve topographic ambiguity in complex alpine terrain, and strictly limit tree depth via PSO to match the low information content of sparse historical records.

The specific objectives of this study are as follows:

- (1) Construct a hybrid assessment framework that integrates the dynamic runoff generation and flow routing mechanisms of the BTOP model into machine learning classifiers, utilizing runoff depth (Q_{depth}) as a “Refinement Factor” to provide basin-scale hydrological connectivity constraints, thereby correcting topographic artifacts produced by static topographic factors.
- (2) Propose and validate the PCNS strategy to prevent the model from relying on spatial memorization by introducing “Hard Negatives,” thereby improving the specificity of the model in defining “safe zones” within complex valley environments.
- (3) Employ Particle Swarm Optimization (PSO) to identify the optimal “shallow tree” configuration, demonstrating how structural regularization prevents overfitting and ensures spatial continuity in data-scarce basins.
- (4) Analyze the model’s physical interpretability based on the SHAP (SHapley Additive exPlanations) method, verifying the consistency between the model’s decision rules and hydrological physical principles.

2. Study Area and Data

2.1. Study Area

The Lhasa River Basin (LRB) originates from the southern foothills of the Nyenchen Tanglha Mountains. It is the largest first-order tributary on the left bank of the middle reaches of the Yarlung Zangbo River and one of the highest basins within this river system. Geographically, the basin is situated between coordinates $90^{\circ}05'–93^{\circ}20'$ E and $29^{\circ}20'–31^{\circ}15'$ N. The main channel extends approximately 551 km, with a drainage area of 32,526 km² (Figure 1). As a typical alpine mountainous basin, the LRB exhibits extremely complex topographic features, with a vast elevation range rising from 3500 m at the outlet to 7100 m at the source. This results in significant vertical climatic zoning and geomorphological heterogeneity. The basin serves not only as a vital ecological barrier for the Qinghai–Tibet Plateau (QTP) but also as the socioeconomic core of the Tibet Autonomous Region. It encompasses the political and cultural center, Lhasa City, and its surrounding rapidly urbanizing zones, where the high concentration of population and infrastructure results in high exposure to flood disasters.

The climate of the basin is primarily controlled by the plateau temperate semi-arid monsoon system, characterized by distinct wet and dry seasons. Mean annual precipitation is approximately 500 mm, with over 80% concentrated in the wet season from June to September. The spatial distribution of precipitation exhibits a clear gradient influenced by topographic uplift effects. The hydrological regime is driven by a mixture of atmospheric precipitation, seasonal snowmelt, and glacier melt, making runoff processes highly sensitive to cryosphere degradation caused by climate warming [22]. However, commensurate with its strategic importance, the hydro-meteorological monitoring network in the basin is remarkably sparse. As shown in Figure 1, only one national hydrological control station (Lhasa Station) and nine meteorological observation stations are distributed across this vast and rugged region. The average control area per station far exceeds the minimum density standards for mountainous areas (100–250 km²/station) recommended by the World Meteorological Organization (WMO). This typical data scarcity severely constrains the accuracy of traditional hydrological models, making this region an ideal testbed for verifying the applicability of hybrid physical-statistical models in ungauged basins. Furthermore, to construct a robust ground truth for flood susceptibility modeling, this study compiled an inventory of 119 historical flood and mountain torrent points spanning from 1840 to

2019. This extensive inventory was derived from the official dataset “*Compilation of cases of major mountain torrents along the Sichuan Tibet line and surrounding areas (1840–2019)*” provided by the National Tibetan Plateau Data Center (TPDC). This authoritative dataset systematically synthesized historical disaster yearbooks, national flood survey projects, and extensive field investigations (Figure 1). These disaster points are primarily distributed along the main river valley and alluvial fans of major tributaries, intuitively reflecting the high vulnerability of riparian settlements.

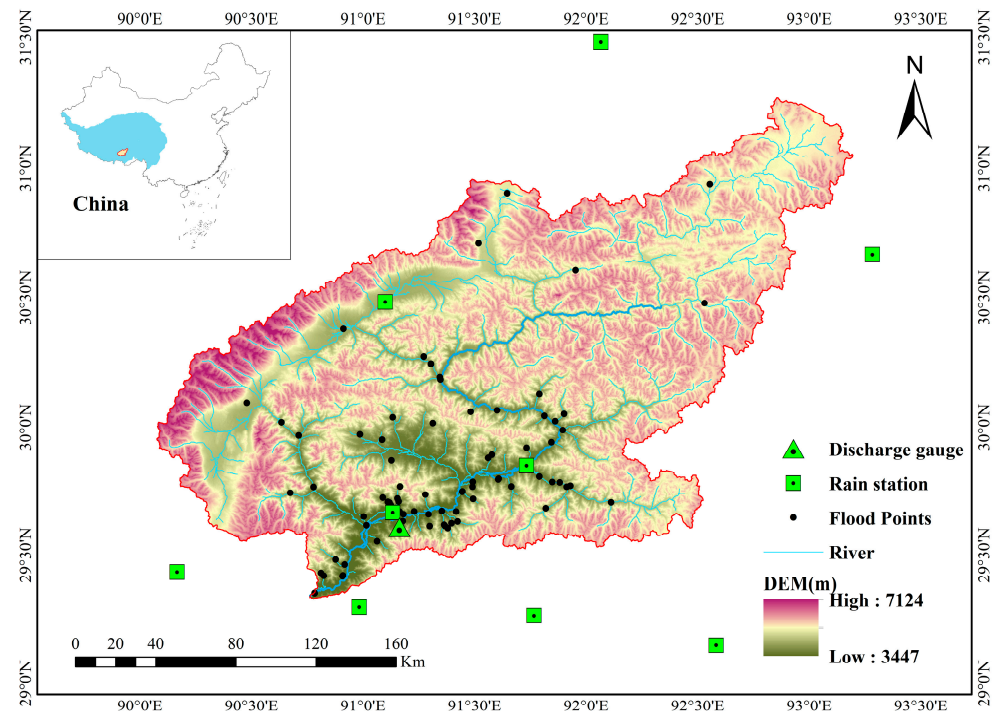


Figure 1. Geographical overview of the study area. Location of the Lhasa River Basin (LRB) on the Qinghai–Tibet Plateau; Distribution of historical flood points (black dots) overlaid on the Digital Elevation Model and river network; Spatial distribution of discharge gauge (green triangle) and rain stations (green squares); The dark blue bold river represents the main stream of the Lhasa River, and the others are tributaries; The inset map shows the location of the study area within China; The blue shaded area indicates the Qinghai–Tibet Plateau (QTP).

Methodologically, utilizing a long-term historical inventory (1840–2019) alongside a short-term hydro-meteorological baseline (2010–2015) is a deliberate design inherent to Flood Susceptibility Mapping (FSM). FSM focuses on identifying the time-independent spatial probability of hazard occurrence rather than temporal forecasting. Given that extreme flood events in alpine basins are infrequent, a long-term inventory is strictly necessary to capture a statistically sufficient number of positive samples representing the comprehensive spatial envelope of geomorphological vulnerability. Furthermore, because the macroscopic topography and river network topologies of the basin remain largely stationary over this timescale, the physical simulation (2010–2015) effectively serves to extract a normalized, climatological “relative accumulation potential” (Q_{depth}) that robustly matches the spatial locations of historical floods regardless of their specific occurrence dates.

2.2. Data Acquisition and Processing

A multi-source heterogeneous dataset was constructed in this study. The data are primarily categorized into two types: dynamic time-series hydro-meteorological data required to drive the physical BTOP, and static environmental conditioning factors required to construct the hybrid machine learning model. All spatial data were reprojected and

resampled to a 30 m spatial resolution and the WGS-1984 coordinate system prior to model input to ensure consistency in spatial analysis.

2.2.1. Hydrological and Meteorological Data (For BTOP)

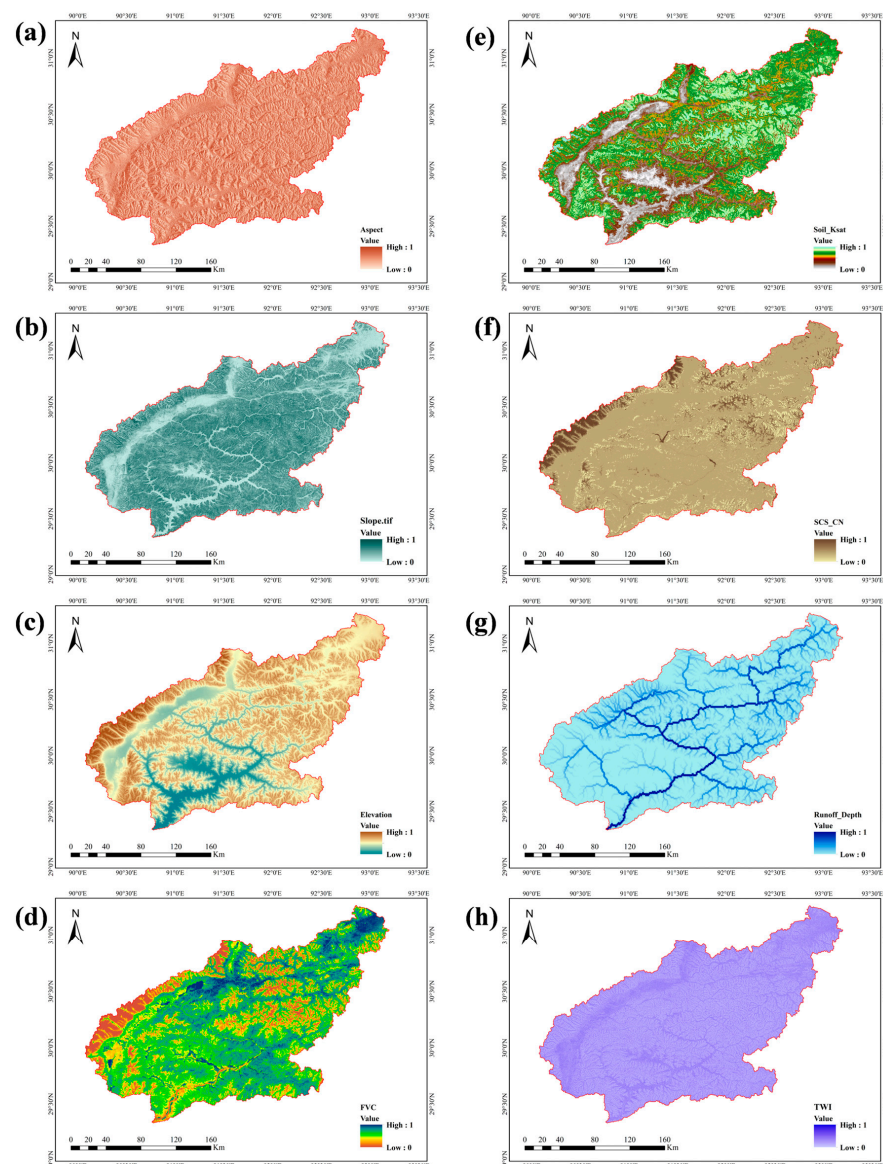
To ensure the reliability of physical process simulation, this study adopted a high-quality dataset configuration consistent with [9], which has been validated for runoff simulation in the Lhasa River Basin. First, benchmark data for model calibration and validation were obtained from daily streamflow observations at the Lhasa Hydrological Station (basin outlet) provided by the Hydrology and Water Resources Survey Bureau of the Tibet Autonomous Region. The period from 2010 to 2013 was designated as the calibration period, and 2014 to 2015 as the validation period. Second, regarding meteorological forcing data, given the sparse and uneven distribution of the rain gauge network within the basin, a multi-source fusion strategy was employed to reduce input uncertainty. Basic precipitation and temperature data were sourced from ground observation stations of the China Meteorological Administration (CMA). To fill spatial gaps in high-altitude ungauged areas, the China Meteorological Forcing Dataset (CMFD) was introduced. This dataset integrates CMA observational data, TRMM satellite precipitation estimates, and GLDAS reanalysis data, providing gridded meteorological fields with high spatiotemporal resolution (0.1° , 3 h) that effectively capture local precipitation events in alpine mountains [23]. Additionally, Potential Evapotranspiration (PET) data required for hydrological simulation were extracted from the ERA5-Land reanalysis dataset published by ECMWF, which demonstrates good applicability in the QTP regions and ensures the comprehensiveness of water balance calculations.

2.2.2. Static Conditioning Factors (For Machine Learning)

Based on the environmental characteristics contributing to disasters in the Lhasa River Basin, this study selected and constructed seven key flood conditioning factors (Table 1), covering topography, land cover, and hydraulic characteristics. Topographic data were derived from the ASTER GDEM V3 (Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model) published by NASA. Its high resolution of 30 m allows for the detailed delineation of valley micro-topography. Based on the DEM, three basic factors—elevation, slope, and aspect—were extracted. Elevation determines the distribution of gravitational potential energy, slope controls the velocity of surface flow routing, and aspect influences the snowmelt process by affecting solar radiation (Figure 2). Regarding land cover, Fractional Vegetation Cover (FVC) was calculated based on the Normalized Difference Vegetation Index (NDVI) from the NOAA Climate Data Record (CDR). The annual mean data for 2015 were used to characterize surface roughness and interception capacity during the validation period (Figure 2d); this dataset was provided by the National Tibetan Plateau Data Center (TPDC) [24]. Soil hydraulic properties are critical for runoff generation calculations. The soil saturated hydraulic conductivity (K_{sat}) was derived from the “China Soil Dataset for Land Surface Modeling (CSDLv2)” published by TPDC [25]. This dataset includes high-resolution (90 m) data on multi-layer soil texture (sand and clay content) and bulk density. This study utilized Pedotransfer Functions to convert these into a spatially continuous conductivity map (Figure 2e). Finally, to characterize the direct potential for runoff generation, a distribution map of the SCS Curve Number (CN) was synthesized (Figure 2f). This factor was calculated by overlaying the 30 m resolution China Land Cover Dataset (CLCD) provided by Professor Xin Huang’s team at Wuhan University [26] with Hydrologic Soil Groups, comprehensively reflecting rainfall-runoff response characteristics under different combinations of land use types and soil textures.

Table 1. List of Flood Conditioning Factors.

Factor Name	Resolution (m)	Source
Elevation Aspect Slope	30	ASTER GDEM (https://www.gscloud.cn/sources/accessdata/310?pid=302 , accessed on 11 December 2025)
FVC	250	The data set is provided by National Tibetan Plateau/Third Pole Environment Data Center (http://data.tpdc.ac.cn).
SCS_CN	30	Calculated from soil data and land use data. The sources of land use data are as follows: The 30 m annual land cover datasets and its dynamics in China from 1985 to 2022 (https://zenodo.org/record/8176941 , accessed on 15 December 2025)
Soil_Ksat	90	The data set is provided by National Tibetan Plateau/Third Pole Environment Data Center (http://data.tpdc.ac.cn).
Runoff_Depth	1000	/

**Figure 2.** Spatial distribution of the seven flood conditioning factors used in this study. (a–g) represent Aspect, Soil Saturated Hydraulic Conductivity (K_{sat}), Slope, SCS Curve Number (CN), Elevation,

Fractional Vegetation Cover (FVC), and the physics-derived Runoff Depth (Q_{depth}), respectively; (h) represent Topographic Wetness Index (TWI).

3. Methodology

Figure 3 presents the P-PDRF framework, which integrates the BTOP hydrological model to extract dynamic runoff depth (Q_{depth}) as a physical constraint. A hybrid training strategy is employed using Physiographically Constrained Negative Sampling (PCNS) and PSO-based structural regularization. The model is validated through statistical metrics, a physics necessity study, and SHAP interpretability analysis to generate the final flood susceptibility map.

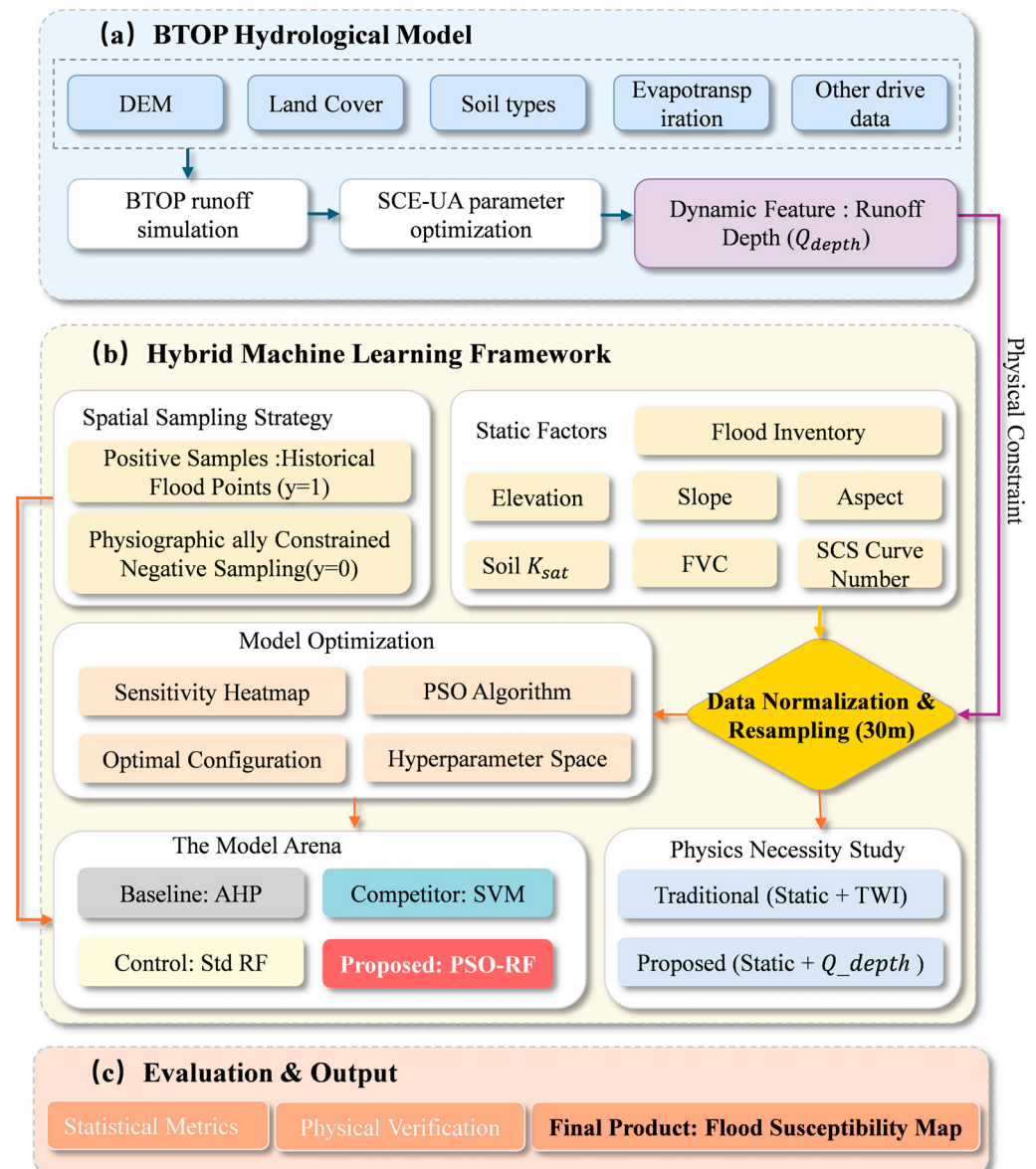


Figure 3. The methodological framework of the proposed P-PDRF approach. (a) Physical modeling: BTOP simulation to derive dynamic runoff depth (Q_{depth}) as a physical constraint; (b) Hybrid ML framework: Integration of PCNS, PSO-optimized Random Forest, and a physics necessity study (comparing Q_{depth} with TWI) within a competitive “Model Arena.” (c) Evaluation and output: Statistical validation, physical consistency check (SHAP), and final flood susceptibility mapping.

3.1. Physical Process Modeling via BTOP

3.1.1. Model Principles and Runoff Generation Mechanism

To overcome the inherent limitations of static topographic descriptors (e.g., slope, aspect) in characterizing complex hydrological connectivity and flow accumulation processes in mountainous terrain, this study adopted the BTOP. BTOP is a semi-distributed hydrological model designed specifically for large-scale, ungauged basins [27]. Unlike lumped models that homogenize basin characteristics, BTOP discretizes the watershed into grid cells, realizing a spatially distributed representation of the runoff generation process based on the “variable source area” concept originally proposed in TOPMODEL [11].

The core physical mechanism of the model relies on the saturation-excess runoff principle, where surface runoff is generated when the soil moisture deficit is filled. The local subsurface flow q_i at grid cell i is mathematically governed by the joint control of the topographic index and soil hydraulic conductivity. The governing equation is modified as follows:

$$q_i = T_0 \cdot \tan(\beta_i) \cdot \exp\left(-\frac{S_{i,t}}{m}\right) \quad (1)$$

where T_0 represents the lateral saturated transmissivity (m^2/h), $\tan(\beta_i)$ denotes the local slope gradient, $S_{i,t}$ is the local soil moisture deficit at time t (m), and m is the decay parameter controlling the rate at which transmissivity declines with soil depth. This formula explicitly establishes the physical relationship between runoff generation and topography, ensuring that low-lying, convergent areas reach saturation and generate runoff more rapidly than steep hillslopes.

The generated runoff is subsequently routed through the river network using the Muskingum–Cunge method [28]. This routing scheme accounts for the propagation and attenuation effects of physical waves in the channel, capturing the dynamic hydrological connectivity and spatiotemporal lag that purely static factors cannot express.

3.1.2. Parameter Optimization and Calibration Configuration

Hydrological models are often subject to parameter uncertainty, particularly in basins with complex topography. To identify the global optimal parameter set, this study employed the Shuffled Complex Evolution (SCE-UA) algorithm [29]. This algorithm combines the advantages of random search and competitive evolution to effectively handle nonlinear parameter spaces; the objective function was defined to maximize the Nash–Sutcliffe Efficiency (NSE) between simulated and observed streamflow.

Data Configuration: Adhering to the standard protocol established for recent research in the Lhasa River Basin (LRB) [9], this study utilized daily observed streamflow data from the Lhasa Hydrological Station (basin outlet) for model calibration. Meteorological forcing data were derived from a fusion product of ground observations from the China Meteorological Administration (CMA) and the China Meteorological Forcing Dataset (CMFD) to ensure input quality under sparse station conditions. The simulation timeline was strictly divided into two phases to verify model robustness:

Calibration Period (2010–2013): Used for SCE-UA parameter optimization to capture the baseline hydrological response characteristics of the basin.

Validation Period (2014–2015): Used to evaluate the model’s transferability and prediction reliability under unseen meteorological conditions.

3.1.3. Construction of the Relative Hydraulic Intensity Index

A critical innovation of this framework lies in the strategic translation of physical model outputs into machine learning constraints. Given that satellite precipitation forcing in data-scarce regions inevitably introduces amplitude errors into absolute discharge sim-

ulations (as analyzed in Section 4.1), we extracted the spatially distributed runoff depth to serve as a structurally reliable “Relative Hydraulic Intensity Index.” To ensure methodological transparency and reproducibility, the mathematical definition, derivation, and implementation mechanism of this index (denoted as Q_{depth}) are detailed below.

The derivation of the hydrological feature sequentially processes the raw physical outputs through temporal aggregation and logarithmic transformation. The core physical index, prior to standardization, is mathematically defined as follows:

$$Q_{log} = \log_{10} \left(\max_{t \in T} (q_t) + 1 \right) \quad (2)$$

Here, q_t represents the raw simulated runoff depth (in mm) at a daily resolution for time step t within the entire continuous simulation period T (2010–2015).

- (1) Temporal Aggregation: The inner term $\max_{t \in T} (q_t)$ extracts the maximum daily runoff depth generated during historical flood events, capturing the extreme hydrological state most representative of disaster-causing conditions.
- (2) Logarithmic Smoothing: Because extreme runoff values exhibit a heavily right-skewed spatial distribution (i.e., massive volumes in main channels, near-zero on hillslopes), a base-10 logarithmic transformation is applied (adding 1 to prevent undefined operations for zero-runoff pixels). This critical step compresses extreme outliers and enhances the continuous spatial texture of the flow network.
- (3) Resampling and Normalization: To align with the 30 m static environmental variables, the processed 1 km Q_{log} raster is spatially downscaled using bilinear interpolation. Finally, to eliminate magnitude bias in subsequent machine learning algorithms, this interpolated raster is normalized to a dimensionless $[0, 1]$ interval using the Min-Max scaling method (as defined in Equation (5)), yielding the final input feature Q_{depth} .

Due to the discrete, non-differentiable mathematical architecture of tree-based algorithms (e.g., Random Forest), embedding physical differential equations directly into a gradient-based loss function is unfeasible. Therefore, this study operationalizes the “hydrological connectivity constraint” via the Physics-Guided Feature Engineering (PGFE) paradigm.

Operationally, Q_{depth} is implemented as an additional dynamic feature concatenated with the six static conditioning factors, constructing a 7-dimensional input feature space vector for the classifier: $X_{input} = [Elevation, Slope, Aspect, FVC, Soil\_K_{sat}, SCS\_CN, Q_{depth}]$.

Rather than functioning as an external post-processing rule, Q_{depth} acts as a physical constraint structurally embedded within the recursive node-splitting process of the decision trees. When coupled with the PCNS strategy (Section 3.2.2), which intentionally introduces “hard negative” samples from low-lying but historically safe terraces, the algorithm is forced to rely on Q_{depth} to break topographic ambiguity. Consequently, the model internalizes Q_{depth} as a hierarchical “conditional trigger”: a location is classified as high-risk if and only if both the static topographic preconditions (e.g., flat terrain) and the dynamic hydrological connectivity (Q_{depth} exceeding a learned critical threshold) are satisfied. This mechanism explicitly corrects the terrain-related errors inherent in pure data-driven models, which frequently misclassify hydrologically isolated depressions (dry lowlands) as flood-prone based solely on local elevation.

Mathematically, unlike linear models that compute a weighted sum, the decision trees in the PSO-RF algorithm partition the multi-dimensional feature space into distinct regions R_m . The constraint effect of Q_{depth} acts as an orthogonal splitting boundary. The logical

decision path leading to a high-risk flood prediction ($Class = 1$) can be conceptualized as an intersection of multiple threshold conditions:

$$P(Flood | X) = \begin{cases} 1, & \text{if } (Elevation \leq \tau_{elev}) \cap (Slope \leq \tau_{slope}) \cap (Q_{depth} > \tau_q) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where τ_{elev} , τ_{slope} , and τ_q represent the specific physical splitting thresholds optimized by the learning algorithm. Through this non-linear mathematical formulation, Q_{depth} acts as an explicit multiplier (or “veto” operator): even if a pixel is located in a low-elevation flatland ($Elevation \leq \tau_{elev}$), the flood probability will be suppressed to 0 if its hydrological connectivity is insufficient ($Q_{depth} \leq \tau_q$). This formalizes the structural mechanism of our physical constraint within the ML architecture.

3.2. Dataset Construction and Sampling Strategy

3.2.1. Selection and Quantification of Flood Conditioning Factors

The scientific selection of conditioning factors constitutes the cornerstone of susceptibility modeling, determining the environmental constraints on flood occurrence. Based on the alpine hydrological mechanisms of the Lhasa River Basin (LRB) and data availability, this study constructed a total of seven key conditioning factors, including one dynamic physical factor (runoff depth, see Section 3.1) and six static environmental descriptors [15,30].

In addition to conventional topographic factors (elevation, slope, aspect) and Fractional Vegetation Cover (FVC), this study focused on the rigorous derivation of two key hydraulic parameters controlling infiltration and runoff generation processes: K_{sat} quantifies the water transmission capacity of soil in a saturated state and is a critical parameter determining the proportion of rainfall converted into surface runoff. Distinct from simplified look-up table methods, this study employed the Pedotransfer Functions (PTFs) proposed by Saxton and Rawls to calculate the spatially continuous distribution of K_{sat} based on multi-layer soil texture data (sand, clay, organic matter) and bulk density (BD) [31]. The calculation process first estimated the field capacity (θ_{33}) and saturated water content (θ_s), subsequently deriving the governing equation for K_{sat} (mm/h):

$$K_{sat} = 1930 \cdot (\theta_{s_{adj}} - \theta_{33})^{3-\lambda} \quad (4)$$

$$\lambda = (\ln(\theta_{33}) - \ln(\theta_{1500})) / (\ln(1500) - \ln(33)) \quad (5)$$

where $\theta_{s_{adj}}$ is the porosity adjusted by bulk density, and θ_{1500} is the wilting point water content. This physically based derivation ensures that the model accurately captures the spatial heterogeneity of soil permeability [32].

The SCS-CN method is a standard empirical model for estimating direct runoff potential in ungauged basins [33]. We constructed a high-resolution CN distribution map by spatially superimposing the 30 m resolution CLCD land use data with Hydrologic Soil Groups (HSG) derived from soil texture:

$$CN_i = f(LULC_i, HSG_i) \quad (6)$$

High CN values (e.g., impervious surfaces, $CN = 98$) indicate high potential for runoff generation, while low values (e.g., dense forests, $CN = 30$) indicate strong retention capacity [34].

To eliminate heterogeneity in dimensions among different factors (e.g., meters for elevation versus mm/h for K_{sat}), particularly for distance-sensitive algorithms such as

Support Vector Machines (SVM), all continuous variables were normalized to the [0, 1] interval using Min-Max Scaling [35]:

$$X_{norm} = (X_i - X_{min}) / (X_{max} - X_{min}) \quad (7)$$

where X_i is the original raster value, and X_{min} and X_{max} are the global minimum and maximum values of the factor, respectively.

3.2.2. Physiographically Constrained Negative Sampling (PCNS) Strategy

In the binary classification problem of flood susceptibility assessment, the quality of non-flood samples (negative class, $y = 0$) is of equal importance to that of flood samples (positive class, $y = 1$) [36]. Based on historical disaster survey data, this study extracted 119 verified historical flood points within the Lhasa River Basin (LRB). Spatial analysis reveals that these points exhibit distinct linear aggregation characteristics along river valleys (Figure 1), primarily distributed in the low-lying alluvial fans of the main Lhasa River and its downstream tributaries. This reflects the physical reality that flood risk in the basin is highly concentrated around the river network.

In flood susceptibility assessment, the selection of non-flood samples directly determines the decision boundary of the classifier. As a typical alpine mountainous region, the LRB possesses a vast elevation range (3500 m–7100 m). If traditional global random sampling is employed, negative samples will likely fall within extremely high-altitude or steep ridge areas. These samples are physically improbable to flood and are considered “easy negatives.” If a model is trained solely on such samples, it is highly prone to falling into the “trivial learning” trap—that is, the model only needs to learn the simple topographic rule that “high altitude equals safety” to achieve high accuracy, thereby masking the true contribution of hydrodynamic processes. To overcome this deficiency and force the model to learn differences in hydrological connectivity within similar topographic backgrounds, this study proposes a Physiographically Constrained Negative Sampling (PCNS) strategy (also referred to as “hard negative mining”). The core logic of this strategy is to construct a highly challenging negative sample candidate pool (S_{neg}) within the valid basin space by setting multidimensional physical thresholds. Its mathematical definition is as follows:

$$S_{neg} = \{x_i \in \Omega \mid L(x_i) = 0 \cap D_{norm} > \delta_d \cap H_{norm} < \delta_h \cap S_{norm} < \delta_s\} \quad (8)$$

where the following definitions apply:

Water avoidance constraint ($D_{norm} > 0.01$): A minute distance threshold (normalized distance of 0.01) was set solely to exclude pixels directly falling within river water bodies to prevent sampling errors, rather than employing a broad buffer exclusion.

Elevation upper limit constraint ($H_{norm} < 0.6$): High-mountain areas with a normalized elevation greater than 0.6 (corresponding to the top 40% of the basin’s elevation range) were excluded.

Slope upper limit constraint ($S_{norm} < 0.5$): Steep areas with a normalized slope greater than 0.5 were excluded.

Through these constraints, this study generated hard negatives for model training. These samples are concentrated in “low-altitude, gentle-slope, non-channel” areas, with an average normalized elevation of 0.3815 (e.g., river terraces and intermontane plains). This strategy artificially weakens the discriminative power of topographic factors (elevation and slope) between positive and negative samples, thereby forcing the machine learning model (particularly PSO-RF) to rely on dynamic hydraulic factors such as runoff depth (Q_{depth}) to distinguish between “flood-prone floodplains” and “safe terraces.” This significantly enhances the model’s sensitivity to hydro-physical mechanisms [37].

3.2.3. Hybrid Dataset Allocation and Validation Scheme

While the PCNS strategy provides a high-intensity training environment, an evaluation process relying solely on hard negatives would construct an “artificial dataset” detached from real-world geographical distributions, leading to biased assessment results. To achieve rigorous training and unbiased validation, this study adopted a hybrid dataset allocation strategy to decouple the training logic from global generalization testing.

First, during the training phase (guided by structural regularization), negative samples in the training set (70% of total samples) were strictly selected from the hard negative pool generated by PCNS. This “adversarial” training design aims to maximize the weight of the physical feature Q_{depth} in model decision-making by increasing classification difficulty, thereby preventing the model from developing spurious spatial memory based solely on topographic coordinates. Second, during the testing phase (global generalization check), negative samples in the test set (30% of total samples) were generated using global random sampling, with their spatial distribution covering the entire non-flood domain from low-altitude valleys to high-altitude mountains (as shown in Figure 4). This “check set” design is of critical academic significance: it not only verifies the model’s fine-grained discriminative power in complex valley environments but also tests the model’s rejection ability (i.e., reducing the false alarm rate) across the entire basin by including a large number of mountainous background samples. This ensures that the finally reported AUC and Kappa metrics truly reflect the universality and generalization performance of the model across diverse real-world topographic backgrounds [38,39].

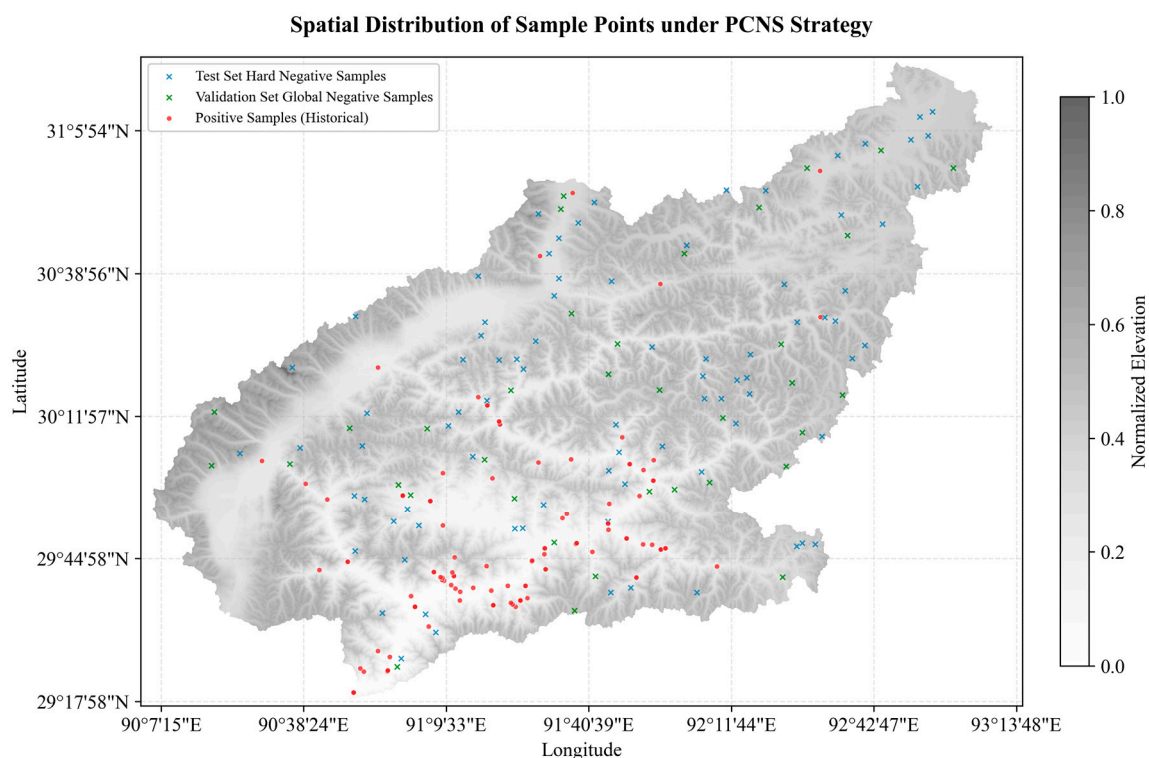


Figure 4. Spatial distribution of training and testing samples under the Physiographically Constrained Negative Sampling (PCNS) strategy. The map illustrates the hybrid sampling design used to mitigate trivial learning and evaluation bias. Red dots represent the 119 verified historical flood events (positive samples), which exhibit a distinct linear aggregation along the river valleys. Blue crosses denote the “Hard Negative Samples” generated via the PCNS strategy ($H_{norm} < 0.6$ and $S_{norm} < 0.5$); these points are concentrated in low-lying, flat riparian terraces to force the model (PSO-RF) to learn complex hydrological drivers (Q_{depth}) rather than simple elevation rules during the training phase.

Green crosses represent the “Global Negative Samples” drawn randomly from the entire basin domain; these are utilized exclusively in the testing set to provide an unbiased evaluation of the model’s generalization capability across the vast alpine background. The background grayscale indicates the normalized elevation (ASTER GDEM), ranging from 0 (valley floors) to 1 (mountain peaks).

3.2.4. Ablation Study Design

An ablation study is designed to rigorously validate the necessity and contribution of the complex BTOP physical simulation. The primary objective is to isolate and quantify the predictive gain provided by the dynamic hydrological connectivity constraint (Q_{depth}) relative to traditional static terrain indices. Two competing model configurations are established for a rigorous comparative analysis. The baseline model is trained on six static conditioning factors (Elevation, Slope, Aspect, FVC, Soil_Ksat, and SCS_CN) alongside the Topographic Wetness Index (TWI) (Figure 2h). TWI serves as a standard static descriptor of soil saturation potential based on steady-state topography assumptions. This configuration represents a traditional data-driven approach that relies solely on static geographic features without awareness of dynamic meteorological forcing. Conversely, the proposed model replaces TWI with the physics-derived dynamic factor, Q_{depth} , to construct the proposed P-PDRF feature space. To eliminate confounding variables, the remaining six static factors, the PCNS strategy, and all hyperparameter configurations of the PSO-RF are kept strictly identical to those of the baseline model. The performance difference between the models is evaluated in Section 4.4 across two dimensions: the shift in feature importance ranking (Mean Decrease Impurity) to determine whether the learning algorithm prioritizes dynamic routing over static wetness, and the quantitative improvement in global classification capability as measured by ROC curves and AUC. The underlying hypothesis of this design is that Q_{depth} provides critical time-lagged routing and heterogeneous precipitation information that TWI intrinsically cannot capture, thereby effectively mitigating terrain-related misclassifications within ungauged basins [39].

3.3. Hybrid Model Construction: Optimization and Comparison

To rigorously evaluate the contribution of the proposed Physically constrained Particle Swarm Optimization–Random Forest (P-PDRF) framework, this study established a competitive evaluation framework termed “The Model Arena.” This arena comprises three benchmark models representing varying degrees of data dependency and complexity, alongside the proposed physically constrained hybrid model. It aims to verify the critical roles of physical mechanisms and evolutionary optimization in enhancing disaster identification accuracy through multi-dimensional performance comparison. Specific model configurations and optimization parameters are detailed in Tables 2 and 3.

Table 2. Hyperparameter configuration and optimization space for machine learning models.

Model	Hyperparameter	Description	Search Space/Method	Optimal Value
PSO-RF	$n_{estimators}$	Number of trees	PSO Search: [20, 200]	60
	max_{depth}	Maximum depth of trees	PSO Search: [2, 20]	4
	Swarm Size	Particle population	Fixed	20
	Iterations	Optimization rounds	Fixed	50
Standard RF	$n_{estimators}$	Number of trees	Fixed (Default)	1000
	max_{depth}	Maximum depth of trees	Fixed (Unconstrained)	None
	Criterion	Splitting rule	Default	Gini Impurity
SVM	Kernel	Kernel function type	Fixed (Non-linear)	Radial Basis Function
	C	Regularization parameter	Grid Search: [0.1, 1, 10, 50, 100, 200]	1
	γ	Kernel coefficient	Grid Search: [1, 0.1, 0.01, 0.001, scale]	‘scale’

Table 3. Pairwise comparison matrix and weight allocation for the AHP model.

Factors	Weights	CI	RI	CR
Elevation	0.2413	0.0339	1.32	0.0257
Slope	0.1604			
Aspect	0.0333			
FVC	0.0705			
Soil_Ksat	0.1058			
SCS_CN	0.0333			
Runoff_Depth	0.3555			

3.3.1. Benchmark Models

Analytic Hierarchy Process (AHP): AHP is a classic Multi-Criteria Decision Making (MCDM) method that decomposes complex unstructured problems into a hierarchical structure, utilizing expert experience to quantify the relative importance of factors [40]. It does not rely on training data but reflects prior cognitive understanding of disaster mechanisms. In implementation, a pairwise comparison matrix was constructed based on the characteristics of the Lhasa River Basin (LRB). As shown in Table 3, runoff depth (Q_{depth}) was assigned the highest weight (0.3555) due to its direct physical significance in causing disasters. The weight allocation passed the Consistency Ratio test ($CR = 0.0257 < 0.1$) to ensure logical transitivity. The core function of AHP is to provide a reference frame based on human prior knowledge, highlighting the advantages of data-driven methods in mining implicit nonlinear patterns [13,41].

Support Vector Machine (SVM): SVM is a supervised learning algorithm based on the Structural Risk Minimization (SRM) principle [42]. Distinct from traditional empirical risk minimization, SVM aims to minimize the upper bound of the generalization error by seeking an optimal hyperplane in a high-dimensional feature space to maximize the geometric margin between flood and non-flood samples [43]. Given the complex heterogeneity of flood triggers, this study selected the Radial Basis Function (RBF) kernel to handle highly nonlinear decision boundaries, as this kernel has proven superior to traditional linear or polynomial kernels in geohazard assessment [44,45]. To ensure fair comparison and eliminate parameter randomness, a Grid Search strategy under 5-fold cross-validation was employed to fine-tune the penalty factor C (search range [0.1, 200]) and kernel coefficient γ (search range [0.001, 1]), finally locking the optimal configuration as $C = 1, \gamma = \text{'scale'}$ (Table 2). SVM assumes the critical task of verifying performance differences between hyperplane-based classifiers and rule-based ensemble algorithms in handling hydrological threshold effects [5,46].

Standard Random Forest (Std RF): Random Forest is an ensemble learning method that constructs a multitude of decorrelated decision trees via Bootstrap Aggregation (Bagging), utilizing Gini Impurity as the splitting criterion to reduce model variance [14,15]. It possesses strong anti-noise and nonlinear fitting capabilities. To simulate the overfitting risk often encountered in routine applications, we configured an “unconstrained” standard RF model, setting the number of trees to 1000 without limiting the maximum depth ($\text{max\_depth} = \text{None}$). Under this configuration, decision trees are allowed to grow fully until leaf nodes are pure (i.e., each leaf contains samples of only one class). This model serves as an experimental control group, aiming to reveal that fully grown trees in small-sample datasets tend to maximize variance and memorize random noise (salt-and-pepper noise), rather than learning generalizable physical laws [38]. This provides a direct basis for comparison to introduce PSO for structural regularization.

3.3.2. Proposed Model: Physics-Constrained PSO-RF

To address the overfitting problem of standard RF and break the “black box” limitation, this study integrated the Particle Swarm Optimization (PSO) algorithm with Random Forest. PSO is a swarm intelligence optimization technique inspired by bird flocking behavior. Through social information sharing among particles and individual historical experience learning, it can rapidly converge to the global optimal solution in complex non-differentiable search spaces [47,48].

We transformed the hyperparameter optimization problem of Random Forest into a global optimization problem within a two-dimensional search space. Each particle represents a potential hyperparameter combination vector $x = [n_{estimators}, max_{depth}]$. The search space was set as: number of trees $n \in [20, 200]$, and maximum depth $d \in [2, 20]$ (Table 2). The population size was set to 20, with a maximum of 50 iterations. Particle velocity (v) and position (x) updates follow these dynamic equations:

$$v_i^{t+1} = \omega \cdot v_i^t + c_1 r_1 (p_{best,i}^t - x_i^t) + c_2 r_2 (g_{best}^t - x_i^t) \quad (9)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (10)$$

where ω is the inertia weight, balancing global exploration and local exploitation; c_1, c_2 are acceleration coefficients representing cognitive and social learning rates, respectively. The fitness function was defined to maximize the Area Under the Curve (AUC) on the validation set to ensure the model’s generalization performance on unseen data.

The core innovation of this study lies in utilizing PSO to actively explore the bias-variance tradeoff [49]. As shown in the subsequent sensitivity heatmap (Section 4.2), the optimization process did not pursue an extremely complex model but converged to a “shallow tree” configuration ($n = 60, max_{depth} = 4$). This explicit constraint on tree depth acts as a powerful means of structural regularization. It forces the model to make decisions based only on the most dominant physical features (such as elevation and Q_{depth}) near the root nodes, thereby effectively filtering out high-frequency noise and topographic artifacts that would otherwise be captured by deep branches, achieving a dual improvement in physical consistency and statistical accuracy [50,51].

3.4. Evaluation and Interpretation

To comprehensively validate the robustness and reliability of the proposed Physically constrained Particle Swarm Optimization–Random Forest (P-PDRF) framework in data-scarce alpine environments, this study established a dual evaluation system. This system quantifies the generalization accuracy on an independent test set (30%) using statistical performance metrics and introduces physical interpretability analysis to determine whether the internal decision logic of the machine learning “black box” adheres to fundamental hydrological physical laws.

3.4.1. Statistical Performance Metrics

The predictive efficacy of the model was quantified multidimensionally based on the confusion matrix, where TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively.

Receiver Operating Characteristic (ROC) and AUC: The ROC curve characterizes the discriminative ability of the model by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) under different decision thresholds. The Area Under the Curve (AUC) serves as a core metric for measuring the global performance of the classifier. Its value ranges from 0.5 to 1, where values closer to 1 indicate a higher degree of distinction

between flood and non-flood pixels. This metric exhibits low sensitivity to class distribution and possesses good robustness [52]:

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (11)$$

Cohen's Kappa (κ): Considering that the spatial scarcity of flood samples may lead to chance agreement in prediction results, this study employed Cohen's Kappa coefficient to correct for random bias [53]. The κ value reflects not only the observed accuracy but also the reliability of the model beyond random guessing. According to the grading standards of [54], a $\kappa > 0.8$ is considered to indicate a "substantial to almost perfect" agreement between prediction and observation:

$$\kappa = (P_o - P_e)/(1 - P_e) \quad (12)$$

where $P_o = (TP + TN)/N$ is the observed accuracy, and P_e is the probability of chance agreement calculated based on marginal probabilities.

Sensitivity and Specificity: Given that the Physiographically Constrained Negative Sampling (PCNS) strategy implemented in Section 3.2.2 intentionally retained a substantial number of "hard negative samples" (i.e., safe points situated in low-lying, flat areas that remained uninundated) within the training set, the significant improvement in specificity directly confirms that the model avoided the trivial learning trap of equating "low elevation" with "flood occurrence." Conversely, it demonstrates that the model successfully leveraged physical features (such as runoff depth) to precisely identify "absolute safe zones" (e.g., river terraces) within topographically similar valley environments, thereby effectively mitigating spatial false positives [55].

$$Sensitivity = TP/(TP + FN) \quad (13)$$

$$Specificity = TN/(TN + FP) \quad (14)$$

F1-Score and Threshold Optimization: The F1-Score is the harmonic mean of Precision and Sensitivity, offering a more objective evaluation than simple accuracy when dealing with imbalanced hazard data. Additionally, this study utilized the Youden Index (J) to determine the optimal threshold for converting continuous probabilities into binary flood susceptibility classes, aiming to identify the optimal cut-off point for model discrimination efficiency by maximizing $Sensitivity + Specificity - 1$ [56].

3.4.2. Physical Interpretability

Machine learning models are often questioned in Earth science applications due to a lack of physical transparency. To ensure that the P-PDRF framework is not only statistically accurate but also mechanistically sound, this study adopted a "global-local" synergistic interpretability strategy.

(1) Mean Decrease Impurity (MDI)

Utilizing the Gini Impurity reduction mechanism internal to Random Forest, the MDI metric quantifies the cumulative weight of each environmental factor's contribution to model prediction during node splitting [14]. This study used MDI to verify whether the physical driving factor (runoff depth Q_{depth}) was identified by the model as a dominant constraint, thereby assessing the intensity of physical mechanism intervention.

(2) SHAP (SHapley Additive exPlanations)

To further analyze the directional contribution and marginal effects of each factor on flood susceptibility, this study introduced the SHAP method based on cooperative game theory [57]. The core of SHAP lies in assigning a physically meaningful contribution value (ϕ) to feature j by calculating its average marginal contribution across all possible feature combinations:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_x(S \cup \{j\}) - f_x(S)] \quad (15)$$

where F is the full feature set, and $f_x(S)$ is the model prediction for feature subset S . Unlike traditional importance ranking, SHAP can identify positive and negative correlations between features and target values.

This study utilized SHAP Beeswarm plots to deeply explore the nonlinear response patterns of features [58]. For instance, according to hydrological physical principles, high Q_{depth} and high rainfall should exhibit positive driving effects (positive SHAP values), while elevation gain should exhibit distinct inhibitory effects (negative SHAP values). This verification mechanism identifies whether the model captures real physical laws of flood evolution and reveals whether the model is anomalously learning specific topographic artifacts, thereby ensuring the scientific validity of the assessment results [59].

4. Results and Discussion

4.1. Hydrological Simulation and Physical Constraints

Figure 5 illustrates the daily streamflow simulation performance of the BTOP at the Lhasa Hydrological Station. To quantitatively evaluate the reliability of physical process simulation, this study employed a multi-dimensional evaluation metric system comprising the Nash-Sutcliffe Efficiency (NSE), coefficient of determination (R^2), and relative bias (R_{bias}).

During the calibration period (2010–2013), the model exhibited exceptional robustness, with both NSE and R^2 reaching 0.85 (Figure 5a). The simulated hydrograph (red dashed line) matches the observed discharge (black solid line) closely in morphology, accurately capturing not only the baseflow recession processes during the dry season but also responding sensitively to the rapid rising limbs of monsoon floods. According to the hydrological model evaluation standards established by [60], this NSE level is classified as “Very Good,” providing strong evidence that the parameter set optimized via SCE-UA successfully characterizes the physical mechanisms dominated by saturation-excess runoff in the Lhasa River Basin (LRB).

However, a noticeable performance decline was observed during the validation period (2014–2015), with the NSE dropping to 0.58 and R_{bias} reaching -29.5% (Figure 5b). While the model correctly synchronized the temporal occurrence of all major flood events, it exhibited systematic underestimation in peak magnitudes, particularly during the extreme streamflow events of 2014. This discrepancy is primarily attributed to the inherent limitations of satellite-retrieved precipitation products (e.g., CMFD) in data-scarce alpine terrain. Satellite sensors suffer from severe spatial smoothing effects and signal attenuation, frequently failing to capture localized, high-intensity convective storms that drive extreme summer floods. Consequently, the BTOP model, constrained by underestimated forcing inputs, inevitably produced dampened discharge peaks [6,61].

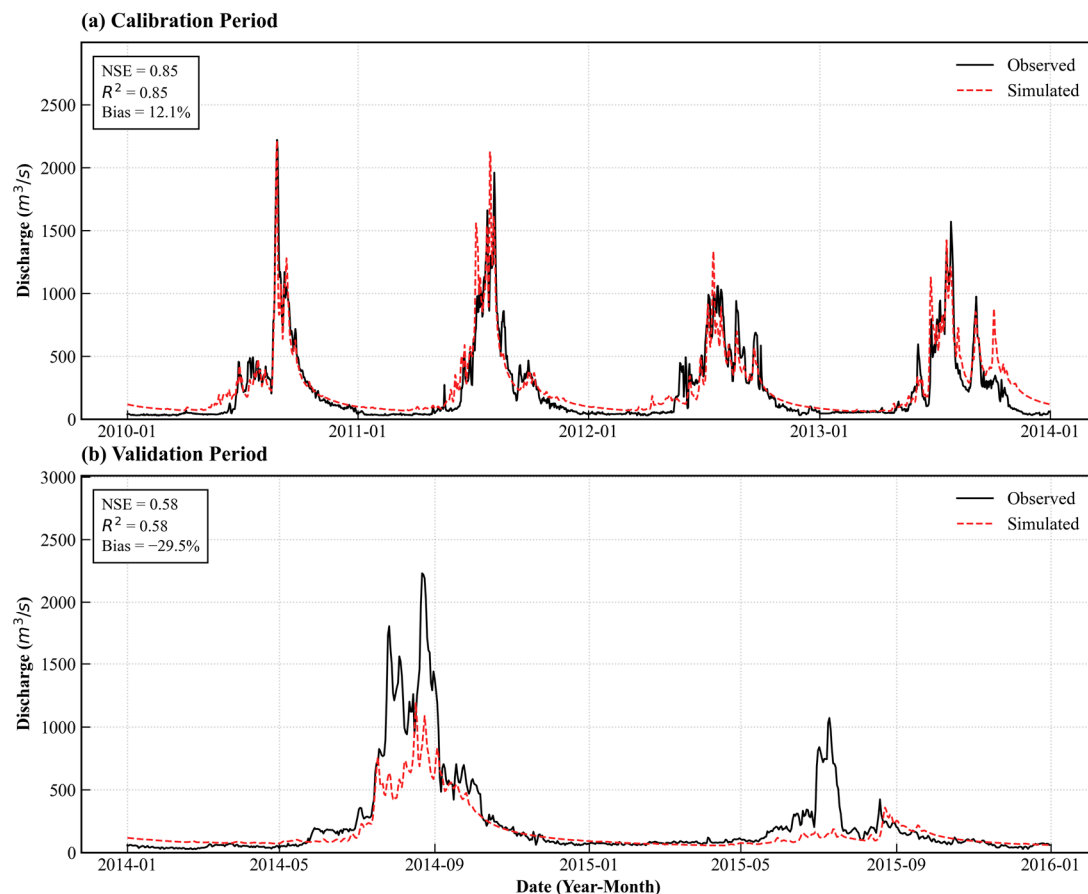


Figure 5. Observed and simulated daily discharge hydrographs of the BTOP model at Lhasa station. (a) Calibration period (2010–2013); (b) Validation period (2014–2015).

Despite this limited performance in absolute magnitude, it is critical to clarify why this discrepancy does not undermine the validity and robustness of the proposed P-PDRF framework. The core philosophy of this hybrid modeling paradigm relies on topological correctness rather than absolute quantitative precision.

As mathematically formalized in Section 3.1.3, the raw runoff outputs are not directly utilized. Instead, they undergo logarithmic transformation and Min-Max normalization to generate the dimensionless $[0, 1]$ index (Q_{depth}). Consequently, the machine learning classifier learns the relative spatial distribution and flow accumulation pathways, entirely bypassing the absolute discharge volume. Even when driven by systematically biased precipitation forcing, the BTOP model robustly preserves the fundamental physical reality (governed by gravity and mass conservation): water converges into structurally connected valleys and bypasses isolated hillslopes.

Therefore, the normalized Q_{depth} successfully provides a highly reliable dynamic topological constraint for the ML model. This design highlights the exact necessity of the hybrid framework in ungauged basins: we extract the reliable causal structure (hydrological connectivity) from an imperfect physical simulation to prevent ML spatial overfitting, while relying on the ML decision tree architecture to effectively insulate the final susceptibility map from the physical model's amplitude errors [62].

4.2. Optimization Landscape and Structural Regularization

To investigate the impact of model structure on prediction performance, we plotted a hyperparameter sensitivity heatmap based on the AUC improvement (ΔAUC) on the validation set (Figure 6). Hyperparameter sensitivity analysis visualized by the heatmap

of AUC improvement), revealing the complex landscape during the PSO process. The color gradient maps the performance gain of optimized configurations relative to the unconstrained Standard Random Forest (Std RF, $\text{max\_depth} = \text{None}$).

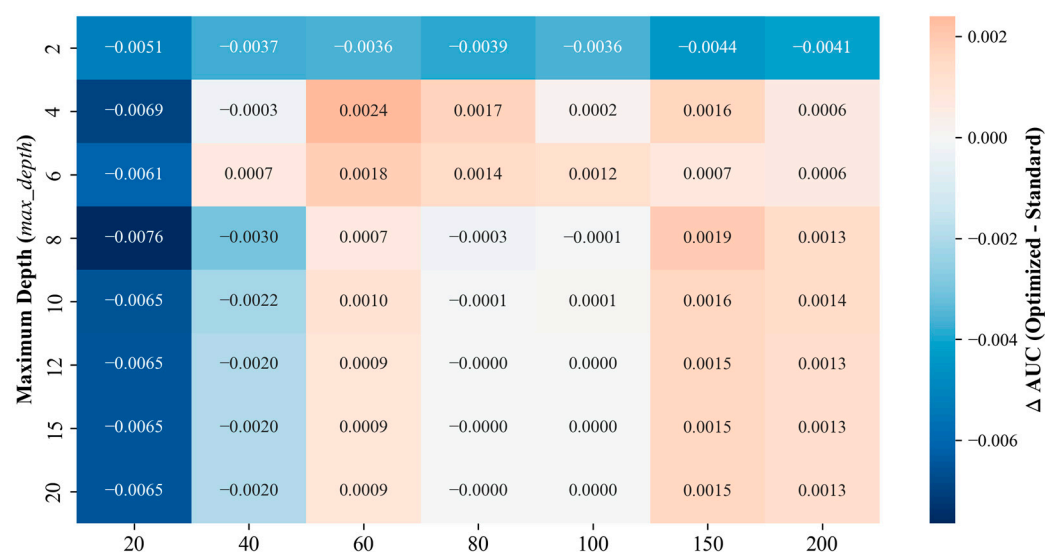


Figure 6. Hyperparameter sensitivity analysis visualized by the heatmap of AUC improvement (ΔAUC). The color gradient represents the performance gain of the optimized RF over the standard RF ($\text{max\_depth} = \text{None}$), highlighting the effectiveness of the “Shallow Tree” strategy (orange/red zones).

The heatmap reveals a distinct “Sweet Spot” (warm orange/red region), highly concentrated in the range of shallower tree depths ($\text{max\_depth} \in [4, 6]$), while being relatively insensitive to the number of trees ($n\_estimators$). Within this interval, the optimized model achieved a stable AUC improvement of approximately 0.002 to 0.007 relative to the benchmark model. Conversely, as tree depth increased beyond 10 (moving toward the bottom of the heatmap), the performance gain showed a diminishing trend or even turned negative (blue region). This result indicates that in this specific geographical environment, deeper and more complex model structures did not translate into higher prediction accuracy.

This phenomenon provides empirical evidence for the bias-variance tradeoff in hydrological modeling of ungauged basins [17,49]. Overfitting via Complexity in Deep Trees: The standard RF model adopts an unconstrained growth strategy, allowing trees to split until leaf nodes are pure. In datasets where samples are scarce and potential noise exists (such as the 119 historical flood points in this study), overly deep tree structures tend to undergo “spatial overfitting”—that is, the model begins to memorize random noise or specific spatial coordinates of training samples rather than learning generalizable physical laws [18]. Generalization via Regularization in Shallow Trees: The PSO algorithm successfully identified a restricted depth ($d \approx 4$) as the global optimal solution. This explicit constraint on tree depth acts as a powerful means of “structural regularization” [63]. By limiting the complexity of decision boundaries, the “shallow tree” strategy forces the model to rely only on the highest-weighted physically consistent features (such as Q_{depth} and elevation) for decision-making near the root nodes, thereby effectively filtering out high-frequency random noise. This finding aligns with the view of Nearing et al. [20] that in systems dominated by physical laws, simpler model architectures often demonstrate superior transferability and robustness in ungauged regions.

4.3. Comparative Model Performance

4.3.1. Statistical Evaluation and Structural Regularization Effects

Quantitative evaluation results based on the independent test set (30%) (see Table 4 and Figure 7) indicate that the PSO-RF model exhibited optimal generalization capabilities across all metrics (AUC = 0.942, Kappa = 0.75). This performance advantage is not coincidental; rather, the comparison with other models reveals the adaptability differences in various algorithmic architectures in addressing the complex hydrological-geomorphological relationships in alpine ungauged basins.

Table 4. Performance metrics of the four models on the independent testing set.

Model	AUC	Accuracy	Recall	Specificity	Kappa
PSO-RF	0.942	0.875	0.889	0.861	0.750
Standard RF	0.919	0.861	0.889	0.833	0.722
SVM	0.913	0.861	0.889	0.833	0.722
AHP	0.853	0.639	0.389	0.889	0.278

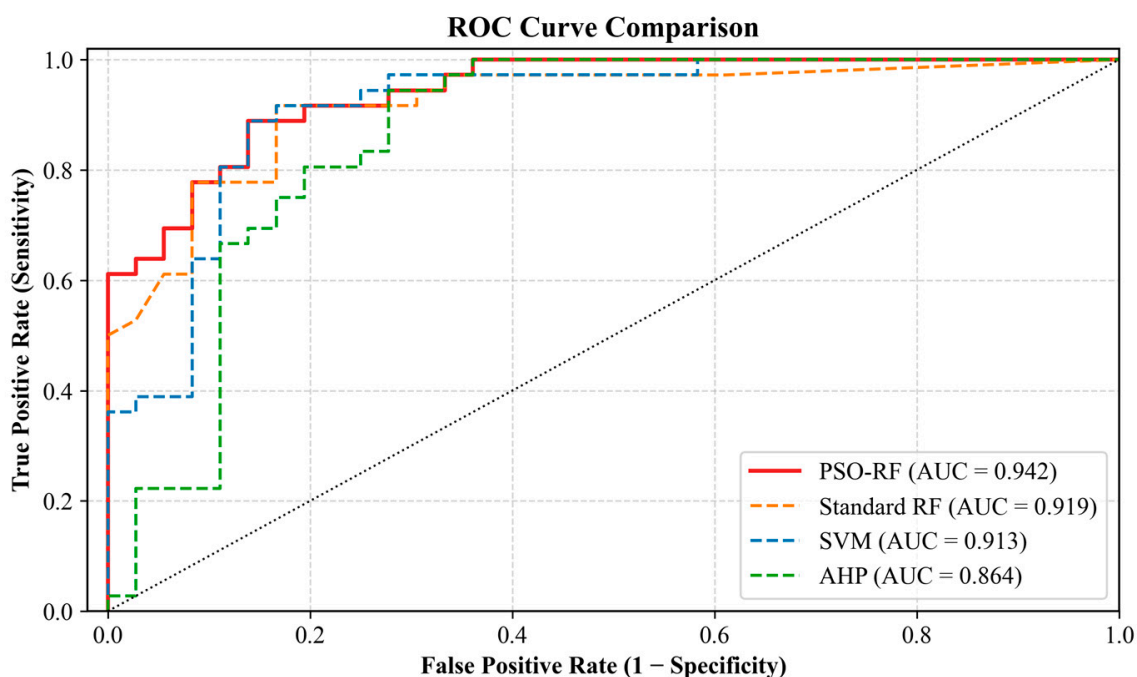


Figure 7. Receiver Operating Characteristic (ROC) curve comparison of the four models (PSO-RF, Standard RF, SVM, and AHP) on the independent testing set. The black dotted line represents the performance of a random classifier (AUC = 0.50).

Comparison with Standard Random Forest (Std RF, AUC = 0.919): Although Standard RF performed robustly due to the advantages of ensemble learning, its slightly inferior AUC exposed the potential risks of unconstrained tree growth ($\text{max\_depth} = \text{None}$). In geological hazard datasets with limited sample sizes, overly deep tree structures tend to capture stochastic noise within training samples, leading to increased model variance. In contrast, the PSO-optimized “shallow tree” configuration ($d \approx 4$) effectively implemented a powerful form of structural regularization. According to the bias-variance tradeoff theory, this explicit constraint on model complexity effectively filtered out high-frequency noise fluctuations, forcing the model to focus on more robust dominant physical laws, thereby constructing more transferable decision boundaries.

Comparison with Support Vector Machine (SVM, AUC = 0.913): Although SVM effectively handled nonlinear features using the Radial Basis Function (RBF) kernel, its performance remained inferior to tree-based ensemble models. From the perspective of algorithmic mechanisms, SVM constructs continuous smooth hyperplanes, whereas decision trees construct hierarchical orthogonal rectangular boundaries. In the characterization of hydrological processes, rule-based “threshold interactions”—for example, “runoff is triggered if and only if runoff depth (Q_{depth}) exceeds threshold α and slope is below threshold β ”—are often more prevalent than continuous mappings. Consequently, the tree structure of PSO-RF is naturally more aligned with such step-wise hydrological response mechanisms.

4.3.2. Sampling Strategy Verification

Analysis of the confusion matrix reveals a key diagnostically significant feature: the PSO-RF model achieved a high specificity (True Negative Rate) of 0.861, implying exceptional performance in suppressing false alarms. This result provides strong empirical support for the Physiographically Constrained Negative Sampling (PCNS) strategy implemented in Section 3.2.2. Traditional global random sampling often selects high-altitude mountainous areas as negative samples, prone to inducing the model to learn the simplistic topographic rule that “high altitude equals safety.” Conversely, the PCNS strategy intentionally retained a substantial number of “Hard Negatives” within the training set—namely, safe zones situated in low-lying, flat terrain that historically remained uninundated (e.g., riparian terraces or arid basins). By maintaining high specificity under the pressure of such challenging classification boundaries, it is confirmed that the PSO-RF model did not fall into the trap of “spatial memorization.” Instead, it successfully utilized dynamic physical constraints, such as runoff depth (Q_{depth}), to precisely identify areas that are topographically “susceptible” yet lack sufficient hydrological dynamics, thereby achieving a profound understanding of the intrinsic physical differences between flood-prone zones and hydrologically safe zones [5,64].

4.3.3. Limitations of the AHP Model

The performance of the Analytic Hierarchy Process (AHP) (AUC = 0.864) provides a methodologically significant benchmark. Based on the pairwise comparison matrix, expert prior knowledge correctly assigned the highest weight to the physically most significant runoff depth (0.356), followed by elevation (0.241) and slope (0.160); however, its prediction accuracy remained significantly lower than that of machine learning models.

This performance gap highlights the intrinsic limitation of the linear superposition principle in complex hazard assessment. The physical mechanism of flood occurrence is highly nonlinear, involving multi-factor “conditional probabilities”—that is, high runoff potential (Q_{depth}) translates into actual flood risk often contingent upon the satisfaction of preconditions (e.g., flat local terrain), rather than a simple weighted sum [13]. The weighted sum algorithm of AHP fails to capture these complex nonlinear coupling relationships, whereas the PSO-RF model automatically detects and internalizes these physical thresholds through its hierarchical tree structure, thereby achieving a performance leap of approximately 8% relative to the linear benchmark [65].

4.4. Physical Mechanism Verification

To rigorously validate the necessity of introducing the complex BTOP simulation, we conducted an ablation study comparing the proposed dynamic physical factor (Q_{depth}) against the traditional static Topographic Wetness Index (TWI). TWI is a standard descriptor of soil saturation potential based on steady-state topography assumptions.

4.4.1. Superiority of Dynamic Constraints over Static Indices

Figure 8 presents the comparative results between the Baseline Model (Static Factors + TWI) and the Proposed Model (Static Factors + Q_{depth}).

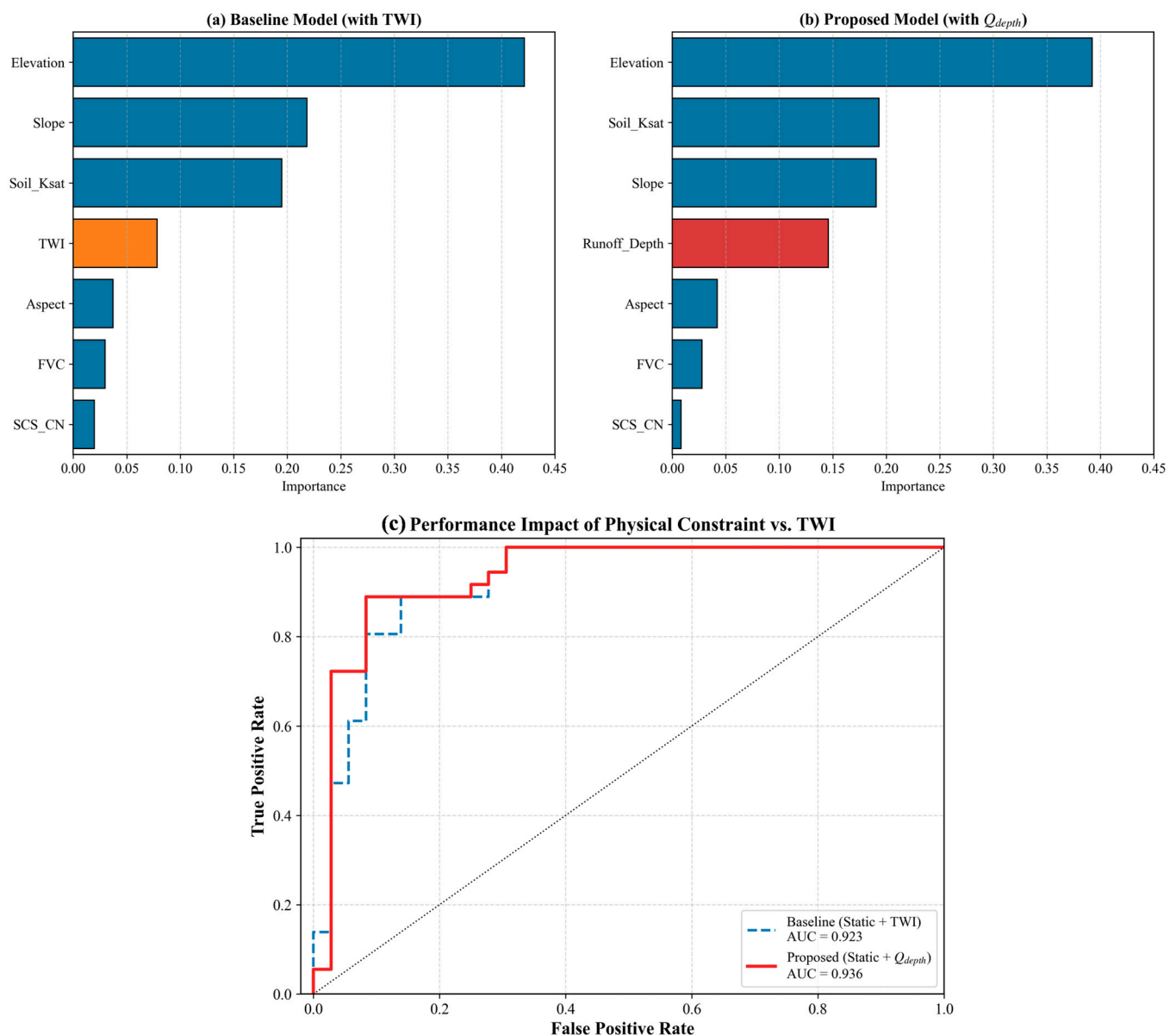


Figure 8. Ablation study comparing the proposed physics-constrained model (with Q_{depth}) against the traditional topographic baseline (with TWI). **(a,b)** Feature importance ranking: Q_{depth} (0.15) significantly outweighs TWI (0.08); **(c)** ROC curves: The proposed model achieves higher AUC (0.936 vs. 0.923) and better performance in the low false-positive rate region.

Feature Importance Contrast (Figure 8a,b): In the Baseline model, TWI ranks relatively low with an importance score of approximately 0.08. In contrast, when replaced by Q_{depth} in the Proposed model, the physical factor's importance rises to 0.15, nearly doubling that of TWI and surpassing static factors like Aspect and FVC.

Performance Gain (Figure 8c): The ROC curves indicate a consistent advantage for the Proposed Model, particularly in the low False Positive Rate region (0.0–0.2). The introduction of Q_{depth} yielded a net performance gain, increasing the AUC from 0.923 to 0.936. A DeLong's test confirmed that this improvement is statistically significant ($p < 0.05$) [66]. It is observed that the AUC of 0.936 obtained in this ablation run exhibits a minor variation

compared to the 0.942 reported in the global model comparison (Section 4.3). This slight discrepancy is attributed to the inherent stochasticity of the ensemble learning process and the resampling variability during data partitioning. Such marginal fluctuations in performance metrics are expected in complex spatial modeling and do not alter the fundamental finding: the physics-derived Q_{depth} provides superior information gain and stronger dynamic constraints than the static TWI, ensuring more reliable susceptibility mapping across different experimental iterations.

The quantitative superiority of Q_{depth} over TWI stems from “Dynamic Forcing Awareness.” TWI assumes spatially uniform precipitation and steady-state saturation, which oversimplifies the complex reality of alpine catchments [67]. In the Lhasa River Basin, precipitation and snowmelt are highly heterogeneous in space and time. The BTOP model integrates this spatiotemporal heterogeneity (derived from CMFD forcing data) and captures the time-lagged flow routing. Consequently, Q_{depth} correctly identifies “dry lowlands”—areas where local topography is low but actual accumulated flow is minimal due to uneven upstream rainfall distribution—whereas TWI erroneously flags these areas as saturated solely based on local slope and contributing area [62].

Despite the high importance of Q_{depth} , the global ranking in Figure 8a,b shows that Elevation and Slope consistently remain the top two predictors, cumulatively accounting for over 50% of the total importance weight in both models. Q_{depth} ranks third, acting as a supplementary factor rather than the primary driver. This ranking structure defines a hierarchical “Global Static + Local Dynamic” control mechanism.

Global Context: Gravitational potential (represented by static Elevation and Slope) defines the fundamental spatial domain of potential hazards, acting as a coarse filter to exclude vast highland areas.

Critical Refinement: Q_{depth} functions as a “Refinement Factor” rather than a standalone dominant driver. Although derived from a coarser resolution (1 km) simulation compared to the 30 m static factors, Q_{depth} provides the necessary basin-scale hydrological connectivity information. It refines the static prediction by vetting the “low-lying candidates” proposed by the DEM and rejecting those that are hydrologically disconnected. This confirms that the model successfully integrates macro-scale physical laws with micro-scale topographic details.

4.4.2. Directional Consistency via SHAP

The SHAP beeswarm plots (Figure 9) further validates this refinement mechanism. For samples with low elevation (typically high risk), if the Q_{depth} value is low, the SHAP interaction value suppresses the final risk probability. This confirms that the model has successfully learned to prioritize hydrological connectivity over simple topographic proximity.

Elevation and Soil Saturated Hydraulic Conductivity (K_{sat}) exhibit significant monotonic negative response characteristics in the plot. High-value points (red dots) for the elevation feature are densely distributed at the far left of the SHAP axis; this strong negative contribution confirms that the model has internalized “gravitational potential energy” as the primary physical constraint on flood occurrence. Similarly, high values of K_{sat} correspond to distinct negative SHAP values, indicating that the model correctly identified the “hydraulic braking” effect of high infiltration capacity on surface runoff generation. This acute capture of physical inhibitory factors ensures the model’s low false alarm rate in high-altitude and high-permeability zones.

In sharp contrast, the physically derived runoff depth (Q_{depth}) displays a significant long-tail positive distribution. As the value of Q_{depth} increases (color changes from blue to red), its SHAP value rises exponentially. This morphology reveals that the model captured the “threshold effect” of flood occurrence—that is, disaster risk rises sharply only when

the accumulated flow exceeds a specific critical point. This proves that Q_{depth} is not merely a redundant feature but acts as a core dynamic trigger dominating the determination of high-risk areas.

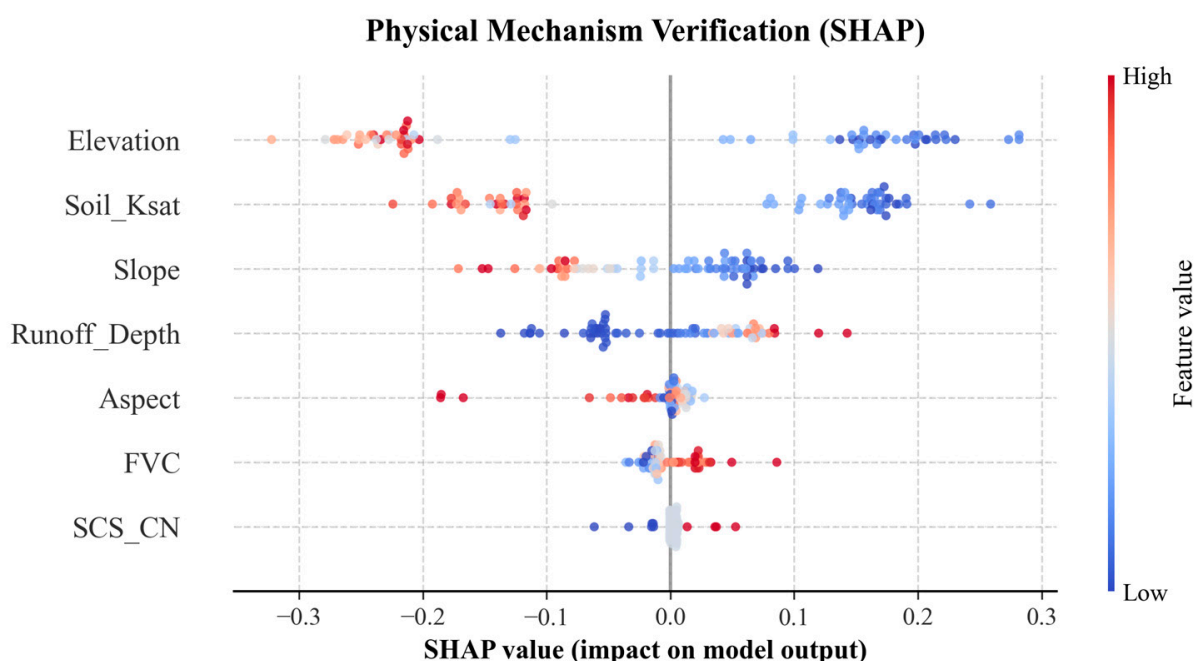


Figure 9. SHAP beeswarm summary plot illustrating the impact of top features on model output. Each dot represents a sample, with color indicating feature value (Red = High, Blue = Low). The horizontal axis shows the SHAP value, where positive values indicate a higher contribution to flood probability.

The most critical finding lies in the deep integration of the interaction between “static topography” and “dynamic hydrology” by the model. In traditional static assessments, low-altitude areas are typically defaulted to high risk. However, SHAP analysis reveals an implicit “veto mechanism”: for samples located at low altitudes (supposedly high risk) but with low Q_{depth} , their final SHAP contribution values are significantly suppressed or even turn negative. This implies that the P-PDRF framework successfully utilized dynamic hydraulic constraints to “veto” false high-risk signals caused solely by topographic factors. This precise identification capability for “arid low-lying areas” is the fundamental reason why this study effectively corrects topographic artifacts in traditional models and achieves physically consistent mapping [68].

4.5. Spatial Susceptibility Mapping and Artifact Correction

To verify the physical plausibility of the model from a spatial dimension, this study generated flood susceptibility maps for the Lhasa River Basin and conducted an in-depth analysis of the model output from three aspects: general spatial patterns, correction of topographic artifacts, and suppression of noise.

4.5.1. General Spatial Pattern and Risk Zone Statistical Characteristics

The flood susceptibility maps generated by the four models (Figure 10) exhibit distinctly different spatial distribution characteristics, with the Analytic Hierarchy Process (AHP) and the machine learning model group showing significant binary divergence. The risk map generated by the AHP model (Figure 10a) presents a broad and diffuse distribution pattern. Although it assigned the highest weight (0.356) to runoff depth (Q_{depth}), constrained by the traditional linear superposition principle, this model cannot effectively

handle nonlinear threshold interactions among conditioning factors. Consequently, it erroneously classifies large flat areas lacking hydrological connectivity as high-risk zones solely based on the low-lying attributes of static topography. Statistical results (Figure 10e) further confirm this “risk overestimation” phenomenon: AHP designated as much as 26.3% of the entire basin as high or very high-risk zones (17.1% + 9.2%). However, overlay analysis with historical flood points (black dots in the figure) reveals that a large number of actual disaster points fell into the “moderate” or even “low” risk zones predicted by AHP. This phenomenon of “casting a wide net but catching few fish” reveals the structural defect of linear models in complex terrain, characterized by the simultaneous existence of over-prediction and severe under-detection [13].

In contrast, the machine learning model group represented by SVM, Standard RF, and PSO-RF (Figure 10b–d) exhibited a more precise dendritic risk structure, where high-risk zones are strictly controlled by valley geometry and develop linearly along the main stream and major tributaries of the Lhasa River. Quantitative statistics indicate that the delineation of high-risk zones by these three models is more convergent and conservative, with the combined proportion of “high” and “very high” risk zones being only 12.2%, 10.1%, and 10.9%, respectively. Among them, the spatial distribution of the PSO-RF model shows the highest agreement with historical flood points; the vast majority of black dots are precisely encompassed within the deep red very high-risk bands, while broad bedrock mountains and high-altitude watersheds are correctly identified as very low-risk zones. This indicates that after introducing the nonlinear mapping mechanism, the model can more acutely capture the physical disaster-causing conditions where “low elevation” and “high runoff depth” coexist, thereby achieving superior discrimination efficiency at the macro scale. To further investigate the internal generalization differences within machine learning models, the following text conducts a locally magnified comparative analysis focusing on local river valleys and rugged mountainous areas.

4.5.2. Correction of Topographic Artifacts

Local magnified analysis of the valley bottom (Figure 11) reveals essential differences in decision boundary construction between Support Vector Machine (SVM) and PSO-RF. As shown, the SVM model (Figure 11a) predicted a continuous blocky “red zone” spanning the entire wide valley bottom, engulfing all flat terrain from the riverbank to the foot of the mountain. However, overlaid historical flood points clearly show that actual disasters are sparsely distributed only in narrow areas immediately adjacent to the river channel, and the red zone delineated by SVM includes a large number of dry floodplains with no historical records. This “topographic artifacts” phenomenon is attributed to the characteristics of the RBF kernel in the SVM algorithm: mapping based on feature space distance tends to generate overly smooth and generous decision boundaries, leading to the “spillover” of risk probabilities to surrounding areas with similar geographical features (e.g., equally flat) but different hydrological features (far from water bodies) [5]. Conversely, the PSO-RF model (Figure 11b) delineated a tightly constrained high-risk corridor, where the red zone is strictly limited to the immediate vicinity of the riverbank, perfectly matching historical points, while adjacent flatlands were correctly downgraded to “moderate” or “low” risk. This demonstrates that tree-based ensemble models, utilizing a hierarchical hard thresholding mechanism and enhanced by the physical constraints of Q_{depth} , successfully constructed steeper and more precise risk boundaries, thereby achieving higher Precision [65].

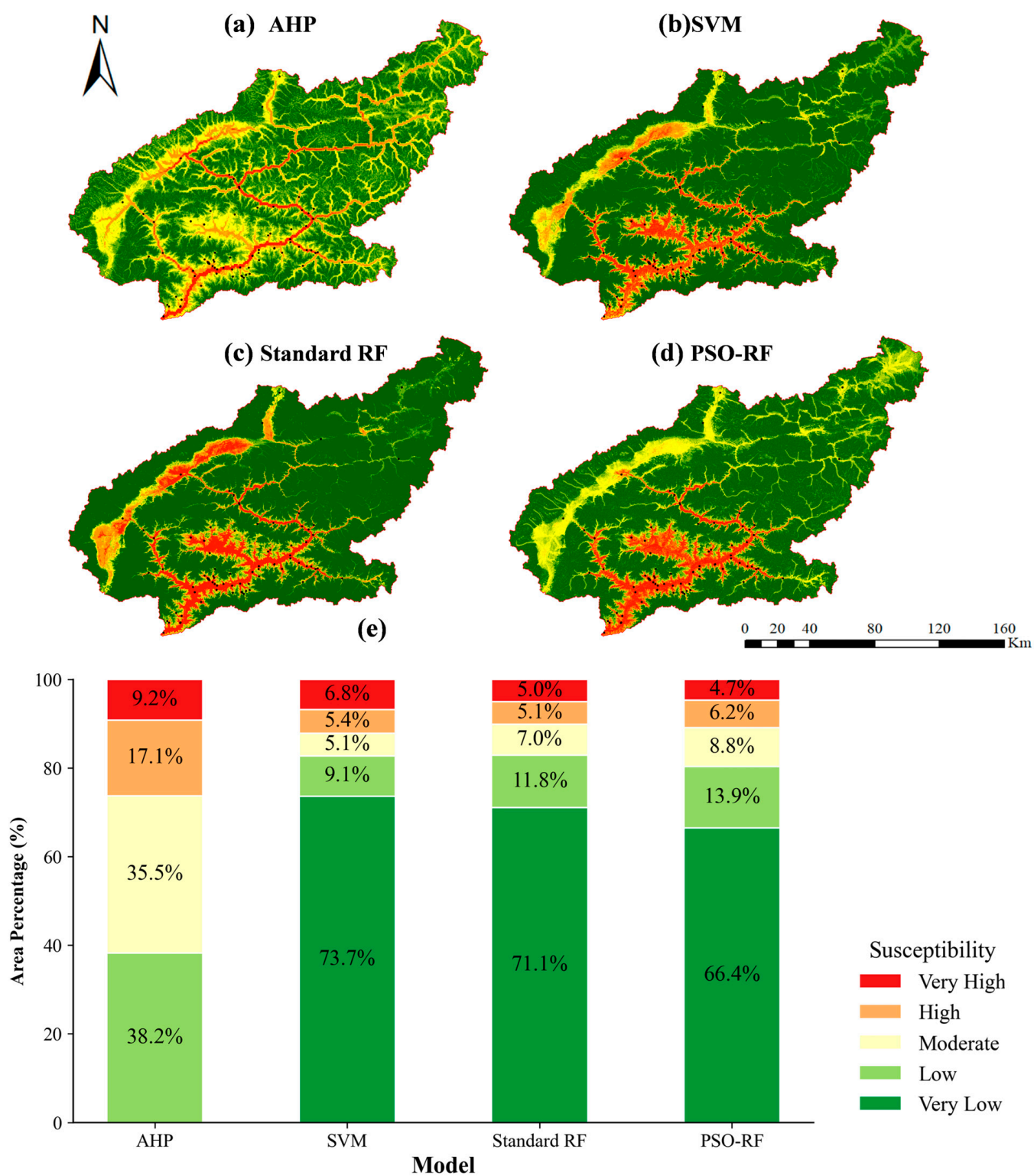


Figure 10. Flood susceptibility maps generated by the four models: (a) AHP; (b) SVM; (c) Standard RF; (d) PSO-RF. The maps are classified into five risk levels using the Natural Breaks (Jenks) method; (e) Statistical distribution of basin area percentage across five susceptibility levels for the four models.

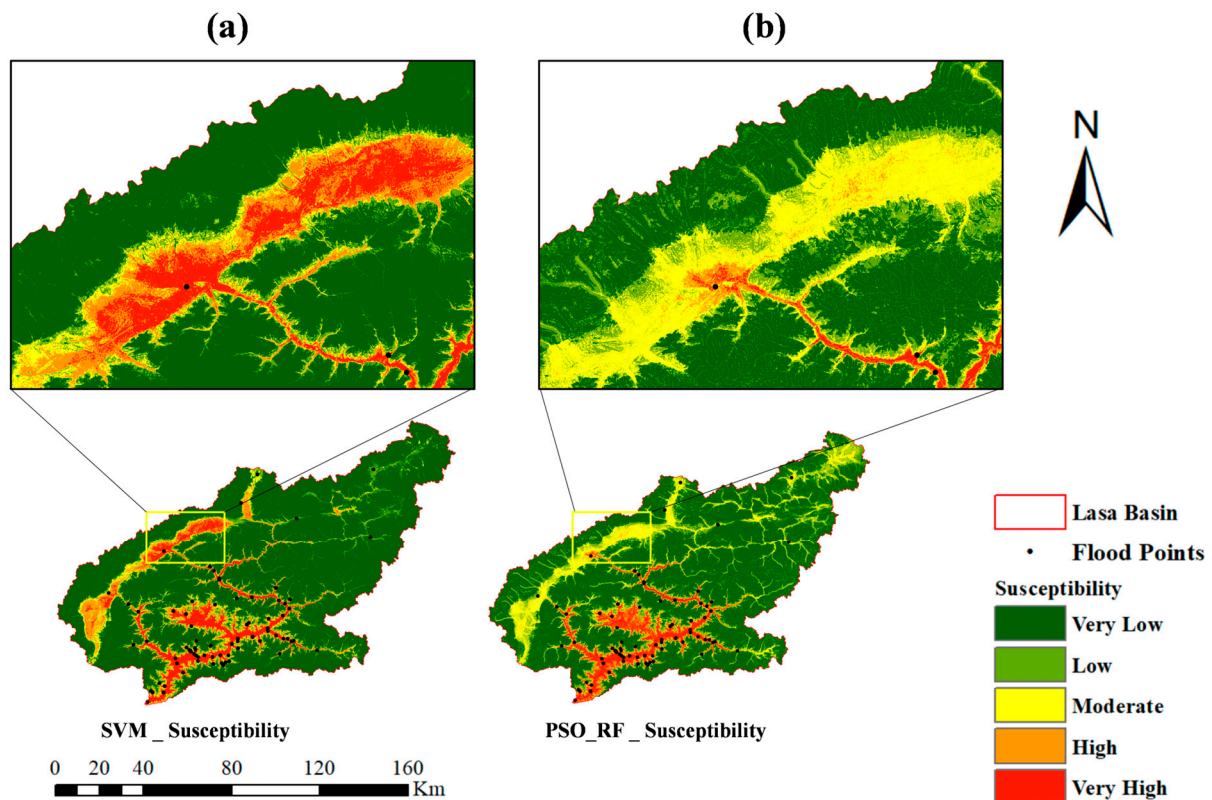


Figure 11. Zoom-in comparison of spatial patterns and artifacts. (a) SVM map showing overestimation (broad red zones) in the valley floor; (b) PSO-RF map showing precise delineation of high-risk corridors along the river channel. Yellow boxes indicate the detailed comparison.

4.5.3. Suppression of Random Noise

Texture analysis focusing on rugged terrain sub-regions (Figure 12) provides direct visual evidence of the impact of model complexity on spatial continuity. Under unconstrained depth ($\text{max\_depth} = \text{None}$), the prediction map of the Standard Random Forest (Std RF) model (Figure 12a) is deeply plagued by “salt-and-pepper noise”—high-risk zones in the valley exhibit an irregular fragmented texture distribution. This fragmented spatial texture destroys the spatial autocorrelation inherent in geographical phenomena, confirming that an overly deep model structure leads to overfitting of local random noise or outliers in the training data. In contrast, benefiting from the “shallow tree” strategy ($d \approx 4$) identified by PSO, the PSO-RF model (Figure 12b) exhibits exceptional smoothness, with continuous and natural transitions in risk levels, and eliminates isolated noise pixels. This spatial consistency proves that structural regularization forces the model to capture broad regional physical laws (such as “high altitude equals safety”) rather than fitting high-frequency local variance, thereby generating a risk zoning map that is more interpretable and suitable for land use planning [46].

4.6. Uncertainty Analysis and Limitations

Although the proposed P-PDRF framework successfully mitigates overfitting and physical inconsistency, the integration of physical simulations with statistical learning in ungauged basins inevitably involves a cascading propagation of uncertainties. To ensure the transparent application of this framework, uncertainties are critically evaluated across four interrelated dimensions.

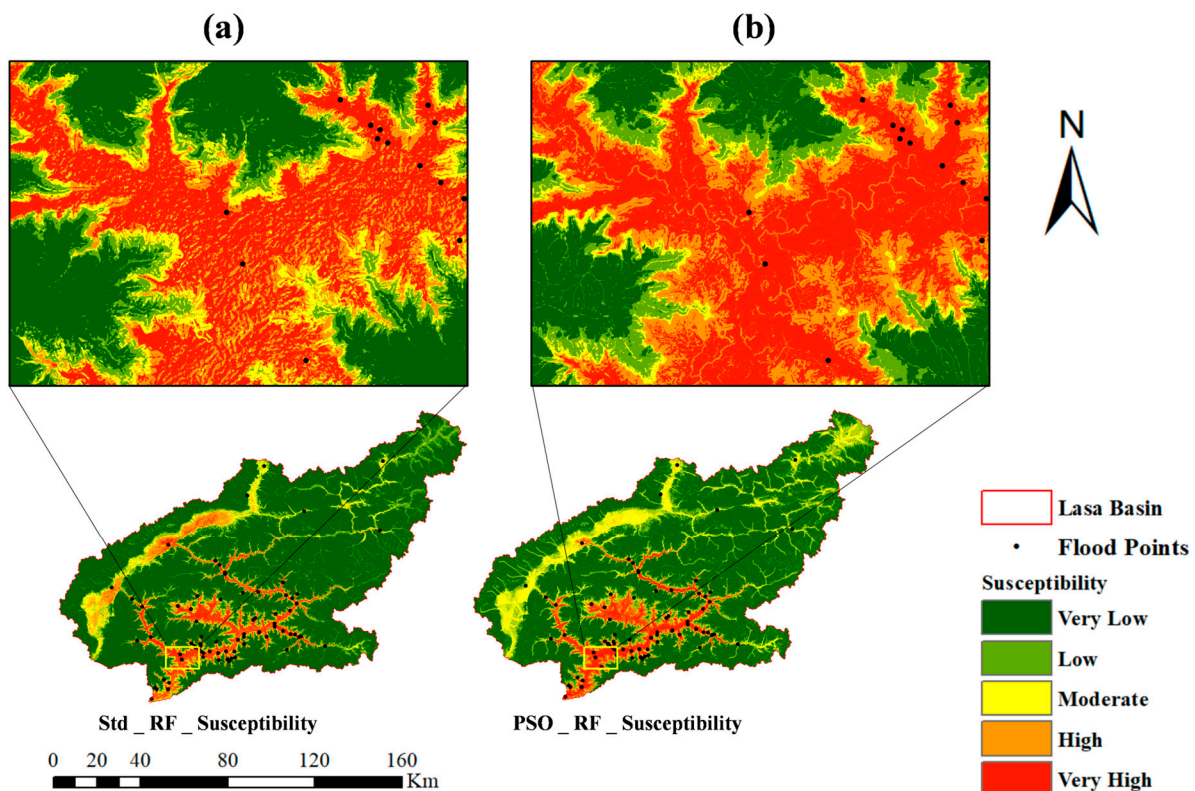


Figure 12. Zoom-in comparison of spatial texture. (a) Standard RF map exhibiting “salt-and-pepper” noise in highland areas; (b) PSO-RF map displaying smooth and continuous risk zones, demonstrating the regularization effect of shallow trees.

The first dimension concerns hydro-meteorological forcing and spatial scale discrepancies. The primary bottleneck in alpine hydrology is the severe scarcity of ground observations. As discussed in Section 4.1, reliance on satellite precipitation (CMFD) introduces significant amplitude errors and spatial smoothing during localized extreme convective events. Furthermore, a spatial scale mismatch exists between the physical simulation and statistical learning. The BTOP model operates at a resolution of 1 km, whereas the static ML factors are at 30 m. Although bilinear interpolation ensures a smooth mathematical transition of Q_{depth} , it cannot deterministically resolve sub-grid micro-topographic routing variations, thereby introducing spatial representation uncertainty into the high-resolution predictions.

The second dimension involves physical equifinality versus topological robustness. The BTOP model, similar to most hydrological models, is subject to structural simplifications and parameter equifinality [12]. Multiple distinct parameter sets calibrated via the SCE-UA algorithm may yield similar NSE values at the basin outlet while generating slightly different internal spatial distributions of runoff depth. However, it is crucial to note that the P-PDRF framework is intentionally designed to be partially immune to these absolute parameter uncertainties. Because the raw discharge is transformed into a relative [0, 1] index, the ML classifier relies strictly on the topological sequence of flow accumulation rather than absolute magnitudes. Consequently, provided that the fundamental gravity-driven routing pathways remain structurally correct, the propagation of parameter uncertainty is substantially dampened.

The third dimension addresses sampling bias and generalization boundaries. The quality of data-driven boundaries heavily depends on the representativeness of training data. In the extreme terrain of the QTP, the 119 historical flood records are inevitably subject to anthropogenic sampling bias, as they are largely clustered along populated main river valleys and accessible road networks. Consequently, remote and uninhabited

deep gorges may be undersampled. Although the proposed PCNS strategy effectively mitigates trivial learning by deliberately mining hard negatives from dry terraces, the regional transferability of the model to fundamentally different geomorphological contexts, such as broad unconfined alluvial plains, remains a dimension of generalization uncertainty that requires cross-basin validation.

The final dimension pertains to algorithmic sensitivity and temporal non-stationarity. The hybrid model exhibits significant sensitivity to algorithmic configurations. As revealed by the optimization landscape (Figure 6), the model is highly sensitive to the maximum tree depth (max_depth); exceeding the optimal shallow tree threshold rapidly induces spatial overfitting. Additionally, converting continuous ML probabilities into discrete risk zones relies on statistical thresholding, such as Natural Breaks, which inherently introduces subjective boundary uncertainty. Finally, from a long-term perspective, this assessment relies on historical data and static topography under the assumption of stationarity. Given that the Lhasa River Basin is highly sensitive to climate change and cryosphere degradation, including glacier retreat and permafrost thawing, the fundamental mechanisms of rainfall and runoff may exhibit non-stationarity in the future. The integration of long-sequence dynamic climate projections into the PIML framework remains a critical direction for subsequent research.

5. Summary and Conclusions

This study aims to resolve the core contradiction in flood susceptibility mapping in data-scarce alpine basins: the dichotomy between physical models (mechanistically strong but parameter-sensitive) and machine learning algorithms (highly accurate but weak in physical consistency). By constructing the Physically constrained Particle Swarm Optimization–Random Forest (P-PDRF) framework, this study confirms that this “accuracy-mechanism” tradeoff is not irreconcilable. By integrating the BTOP to construct relative physical constraints and employing Particle Swarm Optimization (PSO) to implement structural regularization, P-PDRF achieved superior generalization performance ($\text{AUC} = 0.942$) while maintaining physical plausibility. Distinct from the “black box” paradigm of traditional data-driven models, our results demonstrate that combining “imperfect” physical priors (such as relative runoff depth) with “constrained” statistical learning models (shallow tree) is an effective pathway to solve the Prediction in Ungauged Basins (PUB) challenge. By establishing a feedback loop between the BTOP and structured statistical learning, this study successfully operationalized the paradigm of Physics-Guided Feature Engineering (PGFE) [19]. The main conclusions are as follows:

- (1) Dynamic Constraints Correct Topographic Artifacts: Despite the systematic amplitude bias ($R_{\text{bias}} = -29.5\%$) in the physical simulation, the BTOP-derived Runoff Depth (Q_{depth}) proved to be a robust “Relative Hydraulic Intensity Index.” Ablation studies confirmed that Q_{depth} significantly outperforms the static Topographic Wetness Index (TWI) in feature importance (0.15 vs. 0.08). It acts as a critical “Refinement Factor,” incorporating hydrological connectivity to correctly identify “dry lowlands” (hydrologically isolated depressions) that static models often misclassify.
- (2) A crucial finding was revealed: The “shallow tree” ($d \approx 4$) selected via PSO significantly outperformed structurally complex unconstrained models. Although the deep learning field generally posits that increasing model complexity helps capture nonlinear patterns, in geological hazard assessment scenarios characterized by sample scarcity ($N = 119$) and significant noise, excessive complexity leads the model to “rote memorization” of high-frequency spatial noise rather than learning generalizable physical laws [18]. We hypothesize that the core mechanisms controlling flood occurrence (gravity-driven runoff, topographic convergence) inherently belong to

a “low-dimensional physical manifold.” Limiting tree depth effectively imposes a structured prior, forcing the model to make decisions based only on the most dominant physical factors (elevation, Q_{depth}), which aligns with the Parsimony Principle in environmental modeling.

- (3) The P-PDRF framework achieved a significant improvement in specificity (0.861), providing robust validation for the effectiveness of the Physiographically Constrained Negative Sampling (PCNS) strategy. Unlike traditional random sampling, the PCNS strategy intentionally retained a large number of “hard negatives”—safe points located in low-altitude, gentle regions (e.g., river terraces)—within the training set. This strategy artificially reduced the distinguishability of topographic factors between positive and negative samples, forcing the classifier to rely on hydraulic factors (Q_{depth}) to identify the physical differences between “absolute safe zones” (terraces) and “absolute flood-prone zones” (floodplains). The over-generalization of Support Vector Machines (SVM) at valley bottoms (large areas of red false alarms) highlights the limitations of kernel-based methods when facing such hard samples, conversely underscoring the superiority of rule-based tree models combined with the PCNS strategy in precisely delineating high-risk corridors.

This study provides a transferable operational framework for the PUB initiative, offering a robust pathway for disaster assessment in data-scarce alpine basins. It confirms the existence of “Orthogonality of Errors” between process-based physical knowledge and data-driven models—where physical models provide causal structure, and statistical models correct amplitude bias.

Subsequent work will focus on upgrading this framework into a “Multi-scale Spatiotemporal Dynamic Assessment System.” We will move beyond single static susceptibility mapping and plan to utilize long-sequence dynamic hydrological variables generated by the BTOP model to delve into risk evolution characteristics at different temporal scales: including identifying the dynamic expansion and recession processes of risk zones during individual flood events, seasonal floodplains within the year (e.g., differences between monsoon floods and spring snowmelt floods), and risk boundary evolution under extreme climate scenarios. By extending the P-PDRF framework in the temporal dimension, we aim to capture the spatiotemporal non-stationarity laws of flood risk under climate change, thereby providing more refined and dynamic full-cycle disaster management evidence for data-scarce alpine basins.

Author Contributions: Conceptualization: C.Y. and L.Z.; Methodology: C.Y., L.W. and L.Z.; Data curation: C.Z., C.F. and Y.G.; Formal analysis: P.H., C.Y. and J.Y.; Visualization: H.L., C.Y. and L.W.; Validation: Y.G. and C.F.; Funding acquisition: L.Z.; Supervision: L.Z.; Writing—original draft: C.Y.; Writing—review and editing: L.Z. and C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Department of Xizang Autonomous Region, grant number XZ202501ZY0145, and the Natural Science Foundation Youth Project of the Science and Technology Department of Sichuan Province, grant number 2024NSFSC0984.

Data Availability Statement: The streamflow observation data used for model calibration and validation are available on request from the corresponding author due to data sharing restrictions of the Hydrology and Water Resources Survey Bureau of the Tibet Autonomous Region. All other datasets used in this study are publicly available from the following sources: Historical flood inventory data are available from the National Tibetan Plateau Data Center (TPDC) (<http://data.tpdc.ac.cn>); DEM data (ASTER GDEM V3) are available from the NASA Earth Science Data (<https://www.gscloud.cn/sources/accessdata/310?pid=302>, accessed on 11 December 2025); Fractional Vegetation Cover (FVC) data are available from TPDC (<http://data.tpdc.ac.cn>); China Soil Dataset for Land Surface Modeling (CSDLv2) is available from TPDC (<http://data.tpdc.ac.cn>); China Meteorological

Forcing Dataset (CMFD) is available from TPDC (<http://data.tpdc.ac.cn>); ERA5-Land reanalysis data are available from ECMWF (<https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5-land>, accessed on 11 December 2025).

Acknowledgments: We gratefully acknowledge the National Tibetan Plateau Data Center (TPDC) for providing the datasets used in this study, and the Hydrology and Water Resources Survey Bureau of the Tibet Autonomous Region for providing the streamflow observation data.

Conflicts of Interest: Author Peng Huang was employed by the company Sichuan Road and Waterway Construction Engineering Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Intergovernment Panel on Climate Change. *Climate Change 2007: The Physical Science Basis*; IPCC: Geneva, Switzerland, 2007.
2. Hirabayashi, Y.; Mahendran, R.; Koirala, S.; Konoshima, L.; Yamazaki, D.; Watanabe, S.; Kim, H.; Kanae, S. Global flood risk under climate change. *Nat. Clim. Change* **2013**, *3*, 816–821. [[CrossRef](#)]
3. Winsemius, H.C.; Aerts, J.C.; Van Beek, L.P.; Bierkens, M.F.; Bouwman, A.; Jongman, B.; Kwadijk, J.C.; Ligtoet, W.; Lucas, P.L.; Van Vuuren, D.P. Global Drivers of Future River Flood Risk. *Nat. Clim. Change* **2016**, *6*, 381–385. [[CrossRef](#)]
4. Costache, R.; Pham, Q.B.; Sharifi, E.; Linh, N.T.T.; Abba, S.I.; Vojtek, M.; Vojteková, J.; Nhi, P.T.T.; Khoi, D.N. Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques. *Remote Sens.* **2019**, *12*, 106. [[CrossRef](#)]
5. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F.; Mosavi, A. An Ensemble Prediction of Flood Susceptibility Using Multivariate Discriminant Analysis, Classification and Regression Trees, and Support Vector Machines. *Sci. Total Environ.* **2019**, *651*, 2087–2096. [[CrossRef](#)] [[PubMed](#)]
6. Immerzeel, W.W.; Van Beek, L.P.H.; Bierkens, M.F.P. Climate Change Will Affect the Asian Water Towers. *Science* **2010**, *328*, 1382–1385. [[CrossRef](#)]
7. Hrachowitz, M.; Savenije, H.H.G.; Blöschl, G.; McDonnell, J.J.; Sivapalan, M.; Pomeroy, J.W.; Arheimer, B.; Blume, T.; Clark, M.P.; Ehret, U.; et al. A Decade of Predictions in Ungauged Basins (PUB)—A Review. *Hydrol. Sci. J.* **2013**, *58*, 1198–1255. [[CrossRef](#)]
8. Blöschl, G. *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*; Cambridge University Press: Cambridge, UK, 2013.
9. Yue, J.; Zhou, L.; Du, J.; Zhou, C.; Nimai, S.; Wu, L.; Ao, T. Runoff Simulation in Data-Scarce Alpine Regions: Comparative Analysis Based on LSTM and Physically Based Models. *Water* **2024**, *16*, 2161. [[CrossRef](#)]
10. Aronica, G.; Bates, P.D.; Horritt, M.S. Assessing the Uncertainty in Distributed Model Predictions Using Observed Binary Pattern Information within GLUE. *Hydrol. Process.* **2002**, *16*, 2001–2016. [[CrossRef](#)]
11. Beven, K.J.; Kirkby, M.J. A Physically Based, Variable Contributing Area Model of Basin Hydrology/Un Modèle à Base Physique de Zone d'appel Variable de l'hydrologie Du Bassin Versant. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [[CrossRef](#)]
12. Beven, K. A Manifesto for the Equifinality Thesis. *J. Hydrol.* **2006**, *320*, 18–36. [[CrossRef](#)]
13. Yalcin, A. GIS-Based Landslide Susceptibility Mapping Using Analytical Hierarchy Process and Bivariate Statistics in Ardesen (Turkey): Comparisons of Results and Confirmations. *Catena* **2008**, *72*, 1–12. [[CrossRef](#)]
14. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
15. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Flood Susceptibility Mapping Using a Novel Ensemble Weights-of-Evidence and Support Vector Machine Models in GIS. *J. Hydrol.* **2014**, *512*, 332–343. [[CrossRef](#)]
16. Khosravi, K.; Shahabi, H.; Pham, B.T.; Adamowski, J.; Shirzadi, A.; Pradhan, B.; Dou, J.; Ly, H.-B.; Gróf, G.; Ho, H.L. A Comparative Assessment of Flood Susceptibility Modeling Using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *J. Hydrol.* **2019**, *573*, 311–323. [[CrossRef](#)]
17. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* **2019**, *566*, 195–204. [[CrossRef](#)] [[PubMed](#)]
18. Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillerá-Aroita, G. Modelling Species Presence-only Data with Random Forests. *Ecography* **2021**, *44*, 1731–1742. [[CrossRef](#)]
19. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [[CrossRef](#)]
20. Nearing, G.S.; Krazert, F.; Sampson, A.K.; Pelissier, C.S.; Klotz, D.; Frame, J.M.; Prieto, C.; Gupta, H.V. What Role Does Hydrological Science Play in the Age of Machine Learning? *Water Resour. Res.* **2021**, *57*, e2020WR028091. [[CrossRef](#)]

21. Shafizadeh-Moghadam, H.; Valavi, R.; Shahabi, H.; Chapi, K.; Shirzadi, A. Novel Forecasting Approaches Using Combination of Machine Learning and Statistical Models for Flood Susceptibility Mapping. *J. Environ. Manag.* **2018**, *217*, 1–11. [\[CrossRef\]](#)
22. Lin, X.; Zhang, Y.; Yao, Z.; Gong, T.; Wang, H.; Chu, D.; Liu, L.; Zhang, F. The Trend on Runoff Variations in the Lhasa River Basin. *J. Geogr. Sci.* **2008**, *18*, 95–106. [\[CrossRef\]](#)
23. He, J.; Yang, K.; Tang, W.; Lu, H.; Qin, J.; Chen, Y.; Li, X. The First High-Resolution Meteorological Forcing Dataset for Land Process Studies over China. *Sci. Data* **2020**, *7*, 25. [\[CrossRef\]](#)
24. Gao, J.; Shi, Y.; Zhang, H.; Chen, X.; Zhang, W.; Shen, W.; Xiao, T.; Zhang, Y. *China Regional 250m Fractional Vegetation Cover Data Set (2000–2023)*; National Tibetan Plateau/Third Pole Environment Data Center: Beijing, China, 2022.
25. Shi, G.; Sun, W.; Shangguan, W.; Wei, Z.; Yuan, H.; Zhang, Y.; Liang, H.; Li, L.; Sun, X.; Li, D. A China Dataset of Soil Properties for Land Surface Modeling (Version 2). *Earth Syst. Sci. Data Discuss.* **2024**, *2024*, 1–35.
26. Yang, J.; Huang, X. 30 m Annual Land Cover and Its Dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data Discuss.* **2021**, *13*, 3907–3925. [\[CrossRef\]](#)
27. Takeuchi, K.; Hapuarachchi, P.; Zhou, M.; Ishidaira, H.; Magome, J. A BTOP Model to Extend TOPMODEL for Distributed Hydrological Simulation of Large Basins. *Hydrol. Process. Int. J.* **2008**, *22*, 3236–3251. [\[CrossRef\]](#)
28. Tianqi, A.; Takeuchi, K.; Ishidaira, H.; Yoshitani, J.; Fukami, K. Development and Application of a New Algorithm for Automated Pit Removal for Grid DEMs. *Hydrol. Sci. J.* **2003**, *48*, 985–997. [\[CrossRef\]](#)
29. Duan, Q.; Sorooshian, S.; Gupta, V.K. Optimal Use of the SCE-UA Global Optimization Method for Calibrating Watershed Models. *J. Hydrol.* **1994**, *158*, 265–284. [\[CrossRef\]](#)
30. Rahmati, O.; Pourghasemi, H.R.; Zeinivand, H. Flood Susceptibility Mapping Using Frequency Ratio and Weights-of-Evidence Models in the Golestan Province, Iran. *Geocarto Int.* **2016**, *31*, 42–70. [\[CrossRef\]](#)
31. Saxton, K.E.; Rawls, W.J. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Sci. Soc. Am. J.* **2006**, *70*, 1569–1578. [\[CrossRef\]](#)
32. Cosby, B.J.; Hornberger, G.M.; Clapp, R.B.; Ginn, T.R. A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resour. Res.* **1984**, *20*, 682–690. [\[CrossRef\]](#)
33. Cronshey, R. *Urban Hydrology for Small Watersheds*; US Department of Agriculture, Soil Conservation Service, Engineering Division: Washington, DC, USA, 1986.
34. Mishra, S.K.; Singh, V.P. *Soil Conservation Service Curve Number (SCS-CN) Methodology*; Springer Science & Business Media: Dordrecht, The Netherlands, 2013; Volume 42.
35. Han, J.; Pei, J.; Tong, H. *Data Mining: Concepts and Techniques*; Morgan Kaufmann: San Francisco, CA, USA, 2022.
36. Tien Bui, D.; Pradhan, B.; Nampak, H.; Bui, Q.-T.; Tran, Q.-A.; Nguyen, Q.-P. Hybrid Artificial Intelligence Approach Based on Neural Fuzzy Inference Model and Metaheuristic Optimization for Flood Susceptibility Modeling in a High-Frequency Tropical Cyclone Area Using GIS. *J. Hydrol.* **2016**, *540*, 317–330. [\[CrossRef\]](#)
37. Bui, D.T.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Omidav, E.; Pham, B.T.; Asl, D.T.; Khaledian, H.; Pradhan, B.; Panahi, M.; et al. A Novel Ensemble Artificial Intelligence Approach for Gully Erosion Mapping in a Semi-Arid Watershed (Iran). *Sensors* **2019**, *19*, 2444. [\[CrossRef\]](#)
38. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009.
39. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Montreal, QC, Canada, 20–25 August 1995; Volume 14, pp. 1137–1145.
40. Saaty, T.L. *The Analytic Hierarchy Process*; McGraw Hill: New York, NY, USA, 1980.
41. Hong, H.; Tsangaratos, P.; Ilia, I.; Liu, J.; Zhu, A.-X.; Chen, W. Application of Fuzzy Weight of Evidence and Data Mining Techniques in Construction of Flood Susceptibility Map of Poyang County, China. *Sci. Total Environ.* **2018**, *625*, 575–588. [\[CrossRef\]](#)
42. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: New York, NY, USA, 2013.
43. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
44. Kavzoglu, T.; Colkesen, I. A Kernel Functions Analysis for Support Vector Machines for Land Cover Classification. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 352–359. [\[CrossRef\]](#)
45. Yao, X.; Tham, L.G.; Dai, F.C. Landslide Susceptibility Mapping Based on Support Vector Machine: A Case Study on Natural Slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [\[CrossRef\]](#)
46. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Flood Susceptibility Analysis and Its Verification Using a Novel Ensemble Support Vector Machine and Frequency Ratio Method. *Stoch. Environ. Res. Risk Assess.* **2015**, *29*, 1149–1165. [\[CrossRef\]](#)
47. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks; IEEE: New York, NY, USA, 1995; Volume 4, pp. 1942–1948.
48. Poli, R.; Kennedy, J.; Blackwell, T. Particle Swarm Optimization: An Overview. *Swarm Intell.* **2007**, *1*, 33–57. [\[CrossRef\]](#)

49. Geman, S.; Bienenstock, E.; Doursat, R. Neural Networks and the Bias/Variance Dilemma. *Neural Comput.* **1992**, *4*, 1–58. [[CrossRef](#)]
50. Nguyen, P.T.; Ha, D.H.; Jaafari, A.; Nguyen, H.D.; Van Phong, T.; Al-Ansari, N.; Prakash, I.; Le, H.V.; Pham, B.T. Groundwater Potential Mapping Combining Artificial Neural Network and Real AdaBoost Ensemble Technique: The DakNong Province Case-Study, Vietnam. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2473. [[CrossRef](#)]
51. Pham, B.T.; Jaafari, A.; Prakash, I.; Bui, D.T. A Novel Hybrid Intelligent Model of Support Vector Machines and the MultiBoost Ensemble for Landslide Susceptibility Modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 2865–2886. [[CrossRef](#)]
52. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
53. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [[CrossRef](#)]
54. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)]
55. Hossin, M.; Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1–11.
56. Youden, W.J. Index for Rating Diagnostic Tests. *Cancer* **1950**, *3*, 32–35. [[CrossRef](#)] [[PubMed](#)]
57. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
58. Molnar, C. *Interpretable Machine Learning*; Lulu. Com: Morrisville, NC, USA, 2020.
59. Štrumbelj, E.; Kononenko, I. Explaining Prediction Models and Individual Predictions with Feature Contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665. [[CrossRef](#)]
60. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE* **2007**, *50*, 885–900. [[CrossRef](#)]
61. Sun, M.; Liu, A.; Zhao, L.; Wang, C.; Yang, Y. Evaluation of Multi-Source Precipitation Products in the Hinterland of the Tibetan Plateau. *Atmosphere* **2024**, *15*, 138. [[CrossRef](#)]
62. Kratzert, F.; Klotz, D.; Herrnegger, M.; Sampson, A.K.; Hochreiter, S.; Nearing, G.S. Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. *Water Resour. Res.* **2019**, *55*, 11344–11354. [[CrossRef](#)]
63. Mallick, J.; Hang, H.T.; Das, A.; Poddar, S.; Singh, C.K. Integrating Ensemble Machine Learning and SAR-Based Geospatial Modelling for Inclusive and Equitable Urban Flood Resilience. *Sustain. Cities Soc.* **2026**, *137*, 107158. [[CrossRef](#)]
64. Tien Bui, D.; Hoang, N.-D.; Martínez-Álvarez, F.; Ngo, P.-T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A Novel Deep Learning Neural Network Approach for Predicting Flash Flood Susceptibility: A Case Study at a High Frequency Tropical Storm Area. *Sci. Total Environ.* **2020**, *701*, 134413. [[CrossRef](#)]
65. Costache, R.; Tin, T.T.; Arabameri, A.; Crăciun, A.; Ajin, R.S.; Costache, I.; Islam, A.R.M.T.; Abba, S.I.; Sahana, M.; Avand, M.; et al. Flash-Flood Hazard Using Deep Learning Based on H₂O R Package and Fuzzy-Multicriteria Decision-Making Analysis. *J. Hydrol.* **2022**, *609*, 127747. [[CrossRef](#)]
66. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* **1988**, 837–845. [[CrossRef](#)]
67. Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Bui, D.T.; Pham, B.T.; Khosravi, K. A Novel Hybrid Artificial Intelligence Approach for Flood Susceptibility Assessment. *Environ. Model. Softw.* **2017**, *95*, 229–245. [[CrossRef](#)]
68. Zhao, G.; Pang, B.; Xu, Z.; Cui, L.; Wang, J.; Zuo, D.; Peng, D. Improving Urban Flood Susceptibility Mapping Using Transfer Learning. *J. Hydrol.* **2021**, *602*, 126777. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.