



Article

A Residual Neural Network Integrated with a Hydrological Model for Global Flood Susceptibility Mapping Based on Remote Sensing Datasets

Junfei Liu ^{1,2}, Kai Liu ¹ and Ming Wang ^{1,*}

¹ School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875, China; junfeiliu@mail.bnu.edu.cn (J.L.); liukai@bnu.edu.cn (K.L.)

² Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

* Correspondence: wangming@bnu.edu.cn

Abstract: Identifying floods and flood susceptibility mapping are critical for decision-makers and disaster management. Machine learning and deep learning have emerged as powerful tools for flood prevention, whereas they confront the drawbacks of overfitting and biased prediction due to the difficulty in obtaining real data. Therefore, this study presents a novel approach for flood susceptibility prediction by integrating ResNet-18 with a 2D hydrological model for global flood susceptibility mapping using remote sensing datasets. The three main contributions of this study are outlined below. First, a new perspective integrating hydrological simulation and deep learning is presented to overcome the inherent drawbacks of deep learning. Second, the model performance is improved through physics-based initialization. Third, the pretrained model achieves better performance than the original model with incomplete training labels. This experiment demonstrates that the physics-based initialized ResNet-18 model achieves satisfactory prediction performance in terms of accuracy and area under the receiver operating characteristic (ROC) curve (0.854 and 0.932, respectively) and is extremely robust according to a sensitivity analysis.

Keywords: flood susceptibility; deep learning; transfer learning; ResNet-18; XAI; physics-based initialization



Citation: Liu, J.; Liu, K.; Wang, M. A Residual Neural Network Integrated with a Hydrological Model for Global Flood Susceptibility Mapping Based on Remote Sensing Datasets. *Remote Sens.* **2023**, *15*, 2447. <https://doi.org/10.3390/rs15092447>

Academic Editors: Mohammadtaghi Avand, Martina Zelenáková, Jianguang Li and Omid Ghorbanzadeh

Received: 5 April 2023
Revised: 26 April 2023
Accepted: 27 April 2023
Published: 6 May 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Although each country faces its own natural disasters, including tornadoes, earthquakes, and wildfires, floods are one of the major threats to people's livelihoods and affect development prospects around the world [1]. In recent years, the frequency and intensity of floods have increased due to climate change, land-use change, and population growth in flood-prone areas. It was found that 1.81 billion people are directly exposed to 1-in-100-year floods [2]. Flood susceptibility mapping is a critical tool for disaster management planning. Flood susceptibility mapping (FSM) identifies areas that are at risk of flooding according to a series of geoenvironmental conditions [3–5], as it identifies areas that are at risk of flooding and helps in decision-making for disaster risk reduction.

To date, various approaches have been applied for FSM, including physical-based models, physical models and empirical models [6]. Physical-based models are one of the most widely implemented models for predicting flash flood susceptibility [7]. These models simulate the flow of water and the resulting flood inundation using rainfall–runoff modelling based on climatic or remote sensing (RS) data [8]. Physical-based models are numerical models that require intensive effort to preprocess the input data and longer computation times [9]. There are three types of numerical models: 1-, 2-, and 3-dimensional hydrodynamic models [10]. Numerical models are based on fluid motion driven by standard mass, energy and momentum principles.

Physical modelling methods have been a great development in recent years because of the improvement in computing power hardware and software for numerical modelling [6].

However, physical modelling is the least favourable approach to implement because it requires the development and application of extremely accurate experimentations of realistic flood events, complicated manipulation and expensive computing power [7].

Empirical models are data-driven models, referred to as black-box models, which depend on the characteristics and mechanisms of the observed data and hydrological cycle [6]. Machine learning (ML) and deep learning (DL) algorithms are types of empirical models and have emerged as powerful tools for FSM. Some machine learning methods, such as decision trees (DTs) [11], support vector machines (SVMs) [12], K nearest neighbours (KNNs) [13], and artificial neural networks (ANNs) [14], have been applied in FSM and achieved good performance. Multiple ML models have also been integrated and applied in FSM, and such models are typically referred to as ensemble models [15]. The random forest (RF) algorithm is an ensemble algorithm that applies bagging to ensemble a number of DT classifiers [16]. Naïve Bayes trees (NBTs) are an integration of naïve Bayes and DTs for FSM [11]. The Bayesian general linear model (GLM) is an integration of naïve Bayes and GLM [17]. An ensemble of genetic algorithm (GA) models has been applied for FSM [18]. A deep learning neural network (DLNN) that has more than a single hidden layer has been applied in FSM and achieved excellent performance [19].

The ability to quickly train and analyse hydrological data makes ML methods very useful in predicting floods [20], but ML has a number of drawbacks, such as peak-value prediction and overfitting [21,22]. In an effort to address this inherent issue, the integration of machine learning and statistical methods or hydrological models has become a cutting-edge field. Combining bivariate statistics with multiple ML models can result in a coupled model that outperforms individual FSM models [23]. Another proposed approach is to implement more accurate hydrological data generated by hydrological models as labelled data for ML models to improve their prediction performance. For example, a study that integrated a conceptual hydrological model with a monotone composite quantile regression neural network (MCQRNN) improved the model performance in short-term flood probability density forecasting [24]. Global hydrological models (GHMs) integrated with long short-term memory (LSTM) units were applied for global flood simulations and yielded a drastic improvement [25]. A land surface hydrological model integrated with LSTM was applied for streamflow forecasting for a cascade reservoir catchment [26], and it significantly reduced the probabilistic and deterministic forecast errors. A distributed hydrologic model was integrated with LSTM for streamflow forecasting in a medium-sized basin and achieved excellent performance at medium-range timescales [27]. A precipitation-runoff modelling system (PRMS) was integrated with a recurrent graph network and LSTM for streamflow forecasting and achieved excellent performance with fewer training labels [28].

The common conclusion from the previous studies on hybrid/integrated models mentioned above is that these models improved the model prediction ability, while studies that apply the integration of ML and hydrological models for FSM are relatively rare. Therefore, further research with different perspectives is needed to explore new integrated models for FSM.

The residual network (ResNet)-18 is a DLNN that has been widely applied in image recognition and classification tasks [29]. It has 18 layers and utilizes residual connections to address the vanishing gradient problem. Residual connections allow the gradient to flow directly through the network, improving the training process and reducing the number of parameters. Transfer learning is a machine learning technique that involves using a pretrained model to solve a related task [30]. Transfer learning has been used extensively in image recognition and natural language processing tasks.

This study proposes a novel approach to FSM by integrating ResNet-18 with a 2D hydrological model for global flood susceptibility mapping using RS datasets, and an explainable artificial intelligence (XAI) approach, Shapley Additive exPlanations (SHAP) [31], is applied to analyse the difference after hydrologically based transfer learning. The three main contributions of this study are outlined below. First, a new perspective using a ResNet-

18 model, a physics-based initialization rather than a random initialization, is proposed to integrate hydrological simulation and deep learning for FSM. Second, the overfitting tendency of the model can be significantly reduced and the model performance can be improved through physics-based initialization. Third, pre-trained models can achieve better performance with incomplete training labels than the original model.

2. Materials

2.1. Data Preparation

Floods are caused by various environmental factors. A total of 12 flood conditioning factors were considered for FSM in this study based on a literature review (Table 1).

Table 1. Primary sources of the datasets in this study.

Data Type	Subfactor	Resolution	Time	Source and Details
Elevation	Digital elevation model (DEM)	7.5 arc	2010	Google Earth Engine (GEE)
	Slope			
	Slope aspect			
Rainfall	General curvature (GC) Global precipitation measurement (GPM)	10 km	2000–2018	NASA
Soil	Soil type	1:5,000,000	2006	Natural Resources Conservation Service, Department of Agriculture, U.S.
Vegetation	Normalized difference vegetation index (NDVI)	1 km	2000–2018	GEE
Lithological	Lithological types	0.5°	2012	PANGAEA
Climate classification	Köppen-Geiger climate classes (KG)	1 km	2017	Climate Change & Infectious Diseases Group
River network	River network	Variable	2020	Global Runoff Data Centre
Flood inventory map	Historical flood inundation areas	250 m	2000–2018	GEE
Flood inventory map	Historical flood points	Variable	1985–2021	Dartmouth Flood Observatory
Flood inventory map	Hydrological simulation of flood inundation areas	1 km	1980–2013	European Commission

The sediment transport index (*STI*) [32] is widely used in flood sensitivity analysis [33–36]. A detailed characterization of sediment transport that is closely related to roughness coefficients can greatly improve the understanding of fluvial regimes and floods [37,38]. The *STI* is a measure of the amount of sediment that is transported by a river or stream. It is typically expressed as the volume or weight of sediment that is transported over a given period of time [39]. The *STI* reflects the mobility of sediment, and an increase in the *STI* coefficient will increase the frequency of floods [40]. The *STI* is calculated as follows:

$$STI = \left(\frac{A_s}{22.13} \right)^{0.6} \left(\frac{\sin\beta}{0.0896} \right)^{1.3}, \quad (1)$$

where A_s is the area of the basin and β is the slope gradient.

The topographic wetness index (*TWI*) describes soil saturation related to basin runoff [41]. *TWI* integrates water supply and downstream drainage in the upslope catchment area. Although it is straightforward and intuitive, it performs well in a wide range of applications [42]. The *TWI* is calculated as follows:

$$TWI = \ln \left(\frac{\alpha}{\tan\beta} \right), \quad (2)$$

where α and β are the upslope area per unit contour length and slope angle, respectively.

Elevation data were obtained from the Google Earth Engine (GEE) platform. The slope, slope aspect and general curvature (GC) were calculated with the elevation data using

third-order partial derivatives in QGIS. The Euclidean distances to rivers (EDTRs) of the flood points were calculated by the Euclidean distance tool in QGIS and a variable in the Global River Network. In the layer stacking step (Figure 1a), each geoenvironmental factor can be processed as a single-band image of size 3598×1448 , and all the conditioning factor layers are stacked together to form a multiband image. In terms of feature engineering, each image is produced in patches pixel by pixel from the multiband image. Each central pixel and its neighbouring pixels in a 3×3 window are extracted, and the resultant image patch has a size of $3 \times 3 \times 12$ (Figure 1b).

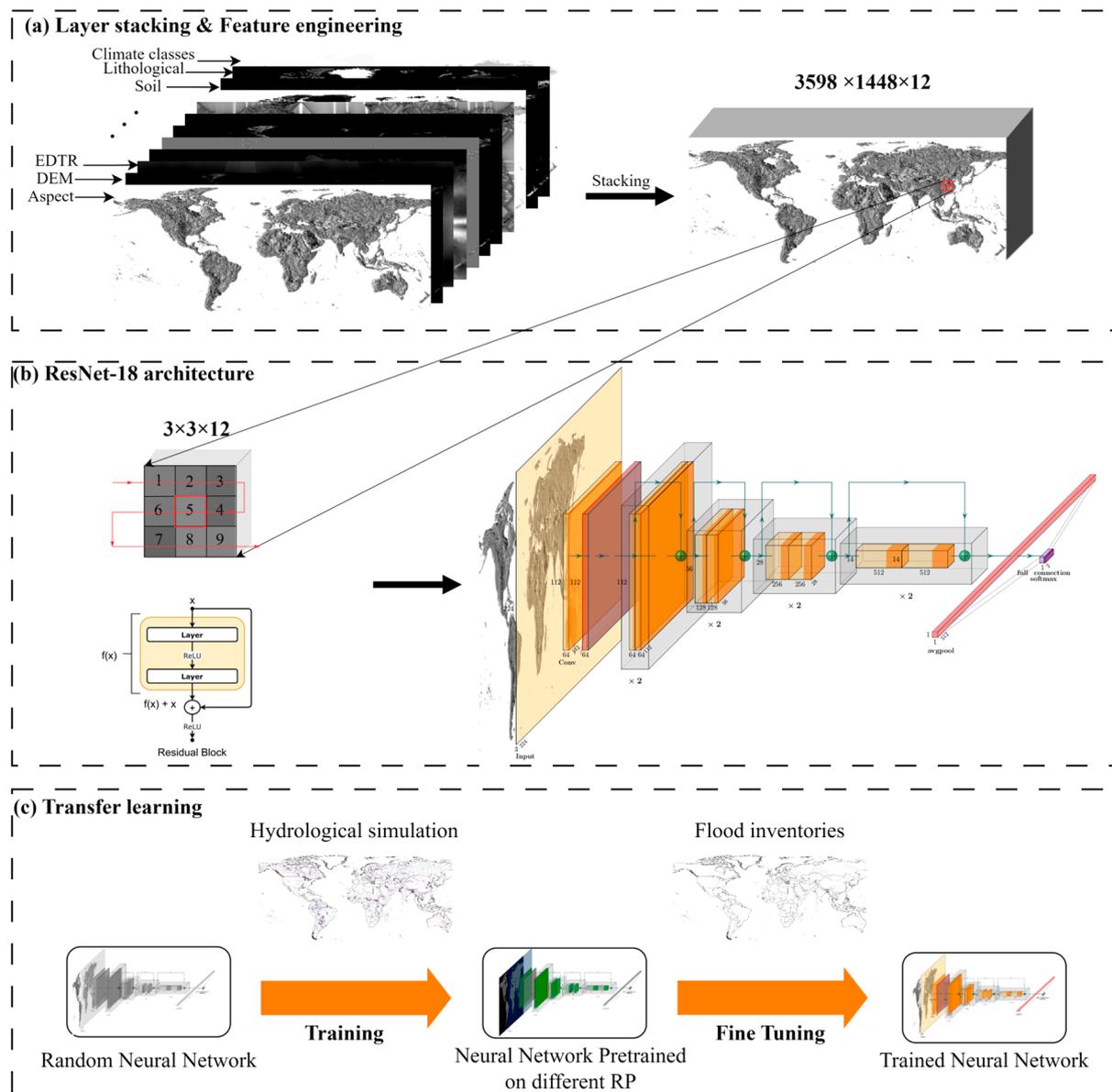


Figure 1. Modelling process in this study: (a) stacking all the conditioning factors together to form a multiband image, (b) sending the resultant image to ResNet, (c) the process of transfer learning.

2.2. Flood Inventory Map

The inventory of flood inundation areas is the benchmark for FSM. Especially for ML, the accuracy of flood event locations directly affects the prediction results of the model. In this study, the satellite-observed inundation maps were obtained from a previous study [43]. The flood event datasets were collected from the Dartmouth Flood Observatory (DFO), which is one of the most comprehensive flood databases [44]. The Terra and

Aqua moderate resolution imaging spectroradiometer (MODIS) sensors were applied to successfully map 913 flood events occurring between 2000 and 2018 according to the DFO database. Each pixel was classified as water or nonwater at a 250-metre resolution. In the task of recognizing floodwater pixels, permanent water pixels were excluded. To make the flood training data more comprehensive, DFO flood event points that were not mapped by MODIS were also included in the flood inventory map.

3. Methods

3.1. ResNet

Building a normal convolutional neural network (CNN) that is sufficiently deep can result in problems such as gradient disappearance, the curse of dimensionality, and degradation issues, and the improvement in accuracy stops at a certain point and eventually begins to degrade [29]. A residual block module is a basic structure in ResNet-18. This residual module allows the model to skip convolutional layers during training, which effectively alleviates the vanishing and exploding gradient problems caused by increasing the network depth [45].

ResNet-18 is mainly composed of a basic ResNet module (Figure 1b). The residual building block is composed of convolutional layers (Conv), batch normalization (BN), a rectified linear unit (ReLU) activation function, and a shortcut connection implemented by the residual block. The output of the residual block can be formulated as follows:

$$y = F(x) + x \quad (3)$$

F is the residual equation, where x and y are the input from the previous layer of the neural network and the output of the current layer, respectively. The entire residual network is composed of a convolutional layer and several basic residual modules.

In this study, ResNet-18 was implemented as the architecture. ResNet-18 contains 17 convolutional layers, a max pooling layer with a size of 3×3 , and a fully connected layer, followed by a dropout layer (Figure 1b). ResNet-18 involves 11,193,858 parameters, where the ReLU activation function and BN are applied to the back of all convolution layers in the residual module, and the softmax function is implemented in the final layer.

3.2. Transfer Learning and Pretraining

Transfer learning is a technique in machine learning where the parameters of a pre-trained model are implemented as an informed initialization for neurons, especially for similar tasks [30]. The parameters of a general neural network model are randomly initialized during training, which can not only increase the training time of the model but also easily cause the model to converge to a local optimum rather than a global optimum. Transfer learning takes advantage of the features learned from the previous model, and fine-tuning with a new dataset can significantly reduce the training time and prevent the model from settling into a local optimum [46]. Deep neural networks usually adopt a hierarchical approach to extract meaningful information. The initial layers detect high-level features, such as edges and corners in image recognition tasks, while later layers identify more complex, domain-specific features. The hierarchical structure of neural networks makes them ideal for transfer learning.

The transfer learning approach in this study (Figure 1c) involves two steps: the generation of flood inundation maps and the training of a residual neural network model. This study utilized flood hazard maps with a 1-km resolution generated by 2D hydrological modelling from a previous study [47], which included maps with 20-year, 50-year, 100-year, 200-year, and 500-year return periods (RPs) of flood inundation areas. ResNet-18 was first pretrained using flood maps with different RPs and further trained using data from the flood event database. The performance of the different models was evaluated using metrics such as accuracy and area under the curve (AUC), and the performance of the models was compared.

3.3. Model Performance Measures

The sensitivity and specificity represent the proportion of floods that are classified as flood pixels and the proportion of nonflood pixels that are classified as nonflood, respectively. True positive (TP) and true negative (TN) are the number of pixels that are correctly classified, P is the total number of pixels with floods, and N is the total number of pixels without floods.

$$\text{Sensitivity} = \frac{TP}{TP + FN'} \quad (4)$$

$$\text{Specificity} = \frac{TN}{FP + FN'} \quad (5)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN'} \quad (6)$$

$$\text{AUC} = \frac{\sum TP + \sum TN}{P + N} \quad (7)$$

For a binary classification problem, the confusion matrix defines the basis for performance metrics, and the receiver operating characteristic (ROC) is a plot generated with sensitivity and specificity information [48]. The AUC calculates the probability P value of each entity as a positive class iteratively when a model classifies which category an entity belongs to, and it sorts these values and calculates them as a threshold θ to weigh a positive category.

3.4. Model Interpretation

SHAP is a post hoc method for black box models [31]. SHAP explains the decision-making rules of the model by calculating the contribution of different features to the predicted value of an instance X . The SHAP explanation method calculates the Shapley values according to coalitional game theory, and the feature values of each instance act as players in a coalition. A player can be an individual feature value or a group of feature values, such as tabular data or image pixels. One innovation brought by SHAP is that the Shapley value interpretation is represented as an additional feature attribution method, a linear model. The SHAP explainer is as follows:

$$g(z') = \phi_0 + \sum_{j=1}^M \phi_j z'_j, \quad (8)$$

where g is the explanatorily linear model, $z' \in \{0,1\}$ is a coalition vector that equals 1 if a feature is observed, M is the number of input features and $\phi_j \in \mathbb{R}$ is the set of Shapley values that represent the feature attributes of a feature j .

Shapely interaction values evaluate the interaction effects between pairs of features after considering the effects of individual features. This index indicates how different pairs of features affect the model output. The Shapley interaction value from game theory is defined as follows:

$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{i,j}(S) \quad (9)$$

In this formula, the main effect of the features is subtracted, and the pure interaction between the features can be computed after considering the individual effects. Values for all possible feature coalition S were averaged, which is similar to the calculation of Shapley values. When the SHAP interaction values for all features were computed, a matrix for each instance was generated with dimensions of $M \times M$, where M is the number of features.

4. Results

4.1. Model Performance

Table 2 shows the prediction performance. All experiments were performed using Python with the Kera framework and scikit-learn.

Table 2. Prediction performance of models with the validation dataset.

Models	Accuracy	Specificity	Sensitivity	TP	TN	FP	FN	SD	RMSE
ResNet	0.851	0.905	0.742	26242	10758	2761	3745	0.463	0.387
ResNet (RP20)	0.853	0.893	0.773	25906	11217	3097	3286	0.470	0.383
ResNet (RP50)	0.854	0.907	0.748	26317	10842	2686	3661	0.463	0.382
ResNet (RP100)	0.853	0.909	0.742	26368	10759	2635	3744	0.462	0.383
ResNet (RP200)	0.855	0.900	0.765	26095	11090	2908	3413	0.467	0.381
ResNet (RP500)	0.853	0.906	0.749	26265	10861	2738	3642	0.464	0.383

In this experiment, the performance of the pretrained models generated by different flood maps was not consistent. The model pretrained using RP50 had the best performance, with accuracy increasing from 0.851 to 0.854, and specificity and sensitivity increasing from 0.905 to 0.907 and 0.742 to 0.748, respectively, compared to those of the ResNet model initialized with random parameters. The other models showed varying degrees of improvement compared to the performance of the ResNet model, and the second-best performing model was that with ResNet (RP100), with accuracy and specificity increasing from 0.851 to 0.853 and 0.905 to 0.909, respectively. The remaining models performed similarly, all performing better than the ResNet model but not as well as ResNet (RP50) and ResNet (RP100).

It should be noted that simply mixing hydrological simulation data (RP200 in this experiment) with flood event datasets does not improve the performance of the model. Compared to that of ResNet, the performance of the model trained with the mixed dataset was significantly lower, with AUC and accuracy decreasing to 0.910 and 0.823, respectively. This indicates that supplementing the data with hydrological simulation data alone does not improve the model performance compared with that of pretraining.

4.2. ROC curve

The success rate curve that was computed using the training dataset is shown in Figure 2a. From these success rate curves, the ResNet (RP50) model had the best performance (AUC = 0.947), followed by the ResNet (RP20) model (AUC = 0.946), the ResNet (RP100) (AUC = 0.944) model, the ResNet (RP500) model (AUC = 0.942) and the ResNet (RP200) model (AUC = 0.933). Compared to those of the ResNet model, the ResNet (RP50), ResNet (RP20) and ResNet (RP100) models had higher AUC values.

Similarly, the prediction rate was graphically represented as a curve using the validation dataset (Figure 2b). Thus, the ResNet (RP200) model had the best performance (AUC = 0.933), followed by the ResNet (RP50) model (AUC = 0.932), the ResNet (RP500) (AUC = 0.932) model, the ResNet (RP100) model (AUC = 0.931) and the ResNet (RP20) model (AUC = 0.931). Compared to those of the ResNet model, all pretrained models had significantly higher AUCs than ResNet.

These results indicate that transfer learning using hydrological simulation flood maps can significantly reduce the overfitting tendency of the model (except for RP200), preventing the model from falling into a local optimum.

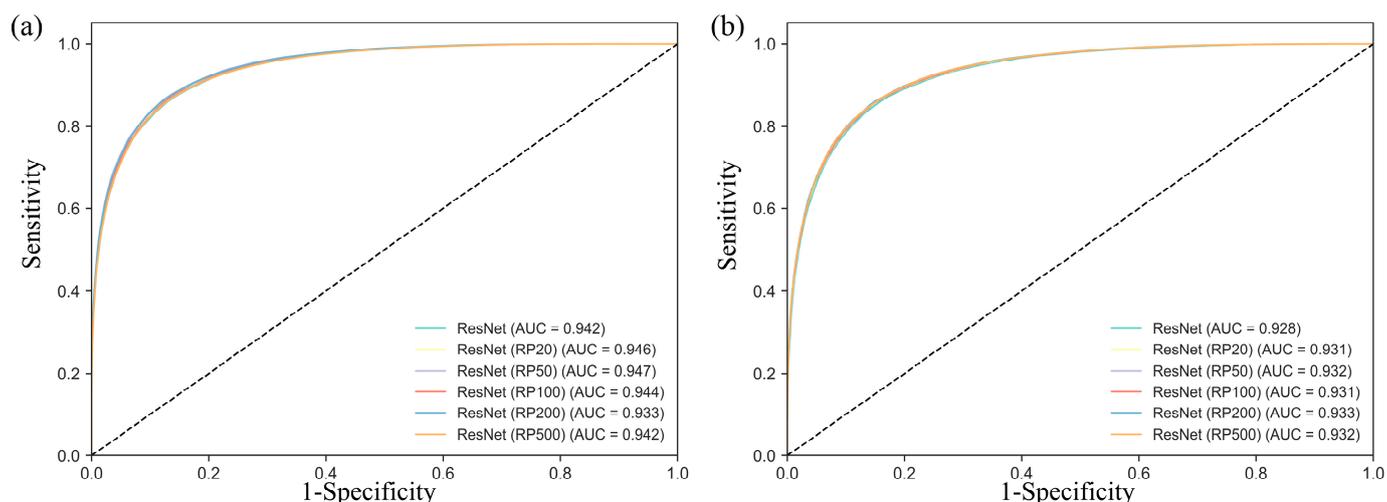


Figure 2. The ROC curves and AUC values of six models: (a) success rate curves, (b) prediction rate curves.

5. Discussion

5.1. Model Performance with Fewer Training Labels

A previous study found that the pretrained model had a significant improvement in streamflow prediction with fewer training labels [28]. It is highly valuable to investigate whether pretraining also yields this improvement in FSM because of the difficulty of acquiring real flood data.

We randomly sampled 30%, 50%, and 70% of the original training data five times each and trained the model with both random initialization and physics initialization. It can be observed that the model with physics initialization outperformed the model with random initialization by a considerable margin (Table 3). All pretrained models had better performance in terms of AUC and accuracy. ResNet (RP50) performed the best with 30% and 50% of the training labels, but ResNet (RP500) had the best performance with 70% of the training labels.

Table 3. Prediction performance of models trained by incomplete training labels with the same validation dataset.

30% Training Labels	AUC	Accuracy	50% Training Labels	AUC	Accuracy	70% Training Labels	AUC	Accuracy
ResNet	0.879	0.795	ResNet	0.894	0.813	ResNet	0.913	0.834
ResNet (RP20)	0.887	0.808	ResNet (RP20)	0.909	0.827	ResNet (RP20)	0.924	0.844
ResNet (RP50)	0.906	0.824	ResNet (RP50)	0.913	0.831	ResNet (RP50)	0.920	0.837
ResNet (RP100)	0.886	0.803	ResNet (RP100)	0.901	0.813	ResNet (RP100)	0.916	0.834
ResNet (RP200)	0.888	0.805	ResNet (RP200)	0.910	0.829	ResNet (RP200)	0.923	0.843
ResNet (RP500)	0.895	0.815	ResNet (RP500)	0.914	0.831	ResNet (RP500)	0.927	0.849

These results indicate that pretraining can significantly improve the model performance with incomplete training data by leveraging hydrological knowledge to learn representative latent variables without risking overfitting a fewer number of observations. Incomplete flood datasets are very common in FSM studies due to the difficulty in obtaining real data. We have found that it is best to integrate labelled data (such as RS data) with hydrological simulation when implementing deep learning for FSM studies, allowing neural networks to leverage hydrological knowledge. This approach has the potential to train a model with better accuracy and robustness.

5.2. Comparison with a Global Flood Dataset

In our experiment, we calculated the susceptibility index of 1,469,173 grid cells to build a global FSM. All susceptible indices were sorted in ascending order and divided into five classes using the natural (Jenks) breaks method [34]. In terms of ResNet, the first class of values (0–0.109) identified zones with very low flood susceptibility, accounting for 46.31% of the study area. The low (0.106–0.269), moderate (0.269–0.457), and high (0.457–0.673) susceptible pixels accounted for 20.62%, 14.46%, and 11.22% of the study area, respectively. Approximately 7.39% of the study area had a very high (0.673–1) flood susceptibility. In terms of ResNet (RP50), the very low (0–0.102) and low (0.102–0.262) susceptible classes accounted for 43.16% and 21.57%, respectively, of the study area, and the moderate (0.262–0.454) flood susceptibility class accounted for 17.64%. The high (0.454–0.678) and very high (0.678–1) classes accounted for 10.81% and 6.82%, respectively.

To evaluate the accuracy of the flood susceptibility map and the pretraining improvement in this study, the flood susceptibility map was compared with a global river flood event map which shows the number of global flood events from 1980 to 2019 (please refer to Figure 10b in [49]).

The ResNet (RP50) map more closely resembled the flood event map from a global perspective (Figure 3), especially in North America and eastern Asia.

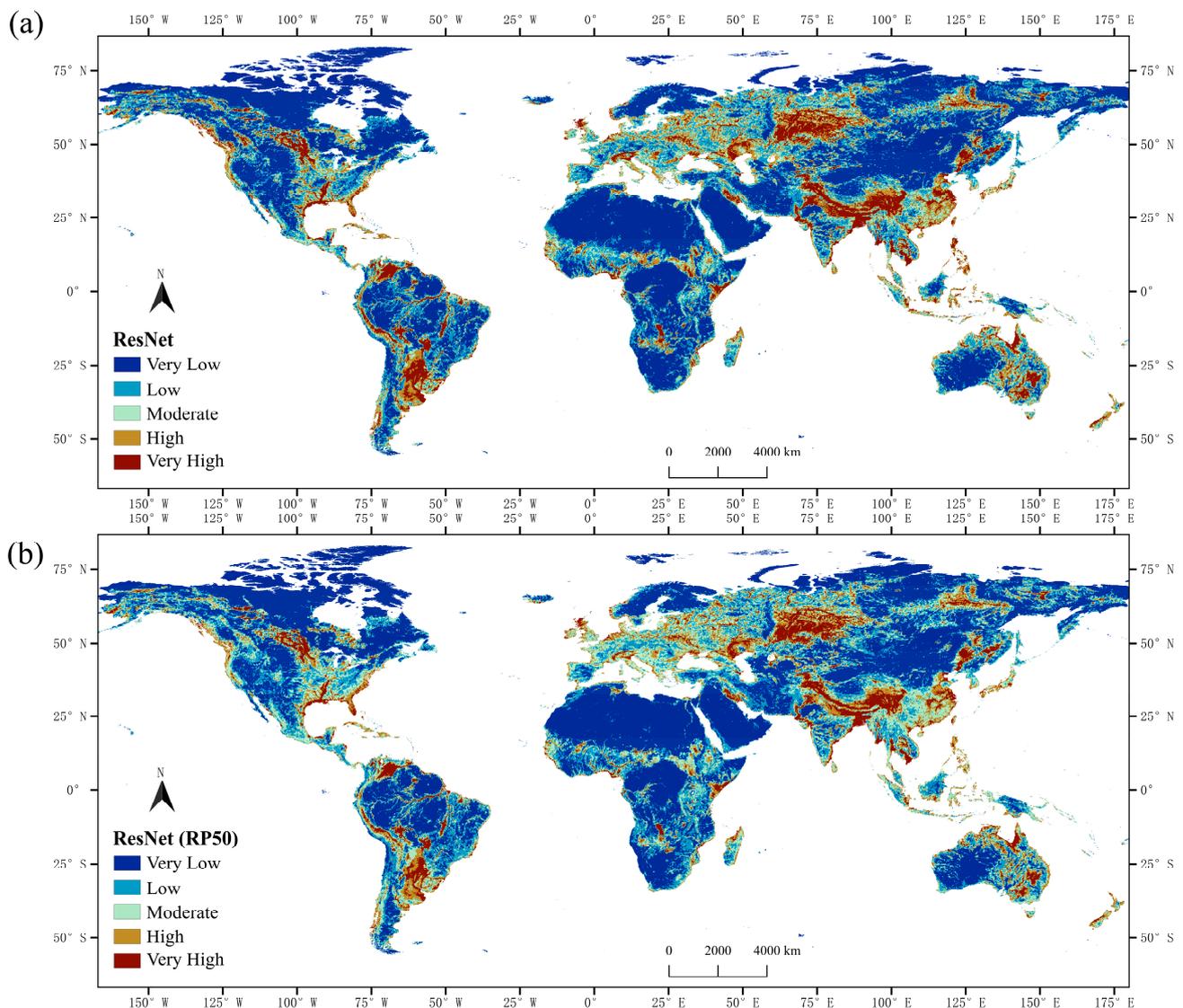


Figure 3. Comparison of (a) global FSM of ResNet; (b) global FSM of ResNet (RP50).

In eastern Asia, the ResNet (RP50) map showed more areas of low, moderate and high susceptibility in the Korean Peninsula, Hokkaido Island, and Sakhalin Island. These trends were consistent with the flood event maps of these regions. In Far East Russia, there were multiple rivers with a comparatively high incidence of flooding. ResNet (RP50) showed more areas of moderate susceptibility in these regions, and the map was consistent with the flood event map.

In North America, the ResNet (RP50) map showed more areas of low and moderate susceptibility in America, Mexico, and Quebec, a province of Canada. These map trends were consistent with those of the flood event map.

5.3. Comparison among ResNet (RP50), ResNet (RP100), and ResNet (RP200)

RPs of 100 and 200 years are highly related to flood occurrence in flood studies. Therefore, a comparison was made in regions with prediction differences among ResNet, ResNet (RP50), ResNet (RP100), and ResNet (RP200), specifically in eastern Asia, Europe, and North America (Figure 4).

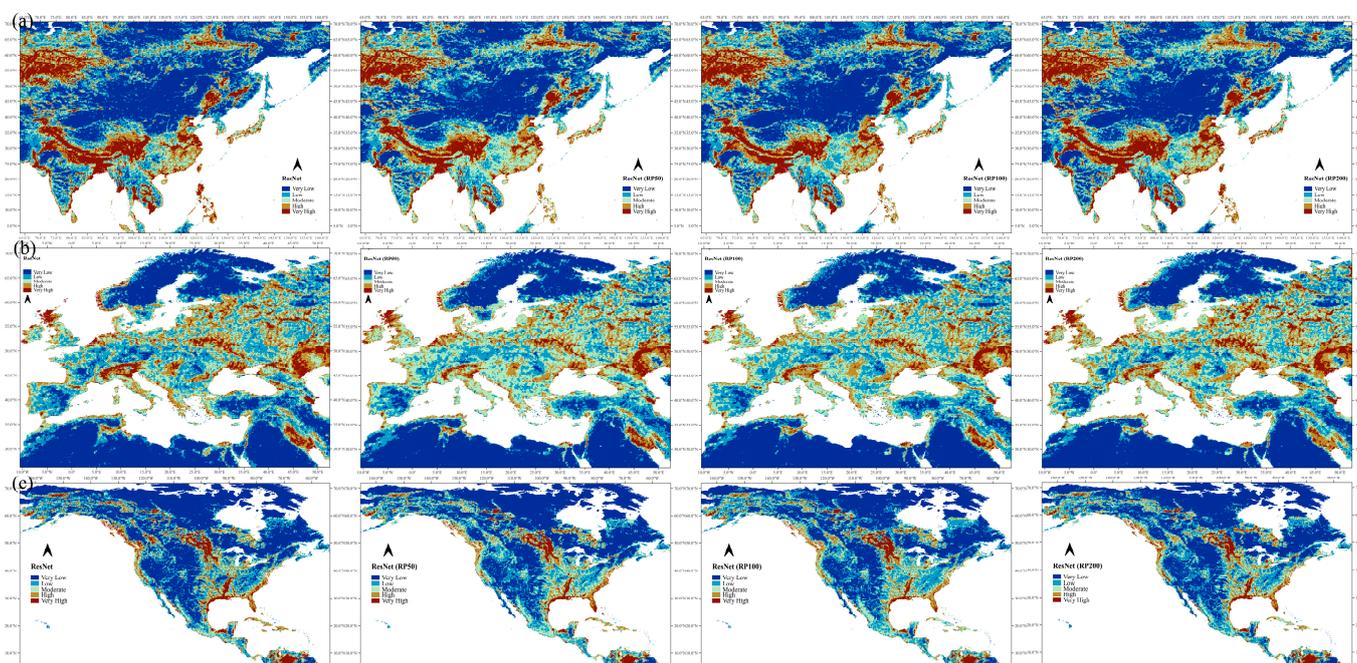


Figure 4. Comparison of ResNet, ResNet (RP50), ResNet (RP100), and ResNet (RP200) in eastern Asia (a), Europe (b), and North America (c).

In eastern Asia, all maps of the pretrained ResNet showed more areas of moderate susceptibility in southeastern Asia, central Asia, the Korean Peninsula, Hokkaido Island, and Sakhalin Island. Compared with the maps of the 3 other pretrained ResNets, the map of ResNet (RP200) showed high and very high susceptibility in South Korea, Philippines, and Sakhalin Island.

In Europe, all maps of the pretrained ResNet showed more areas of moderate and high susceptibility, specifically in Ireland, the UK, and France. The map of ResNet (RP200) showed more areas of moderate and high susceptibility in Ireland, the UK, Norway, and Finland.

In North America, all maps of the pretrained ResNet showed more areas of low and moderate susceptibility in the western US, northern Mexico, and eastern Canada. Compared with the maps of the other pretrained ResNets, the maps of ResNet (RP100) and ResNet (RP200) showed more areas of moderate and high susceptibility in eastern Canada.

5.4. Model Interpretation with SHAP

In this study, 1000 points were randomly sampled for model interpretability analysis using SHAP, exploring how pretraining adjusts the feature importance to improve the

model performance. Figure 4 shows the feature importance ranking of all models. The slope, DEM and *TWI* are the most important factors causing floods in all models, and the difference between ResNet and ResNet (RP50) lies mainly in the different contributions of the DEM, GPM, and NDVI to the model output (Figure 4a,b). Compared with those of ResNet, the importance of the slope, NDVI and *TWI* decreased for ResNet (RP50), but the importance of the DEM, GPM, KG, and EDTR increased.

The feature importances of the pretrained models are mostly similar, while the importance of the STI increased significantly for ResNet (200) compared with the other models (Figure 5e). the difference between ResNet (RP50) and ResNet (RP100) lies mainly in the different contributions of *TWI* (Figure 5c,d). The importance of the GPM comparatively increased for ResNet (20) compared with the other models (Figure 5b). The slope was the most important factor in ResNet (500).

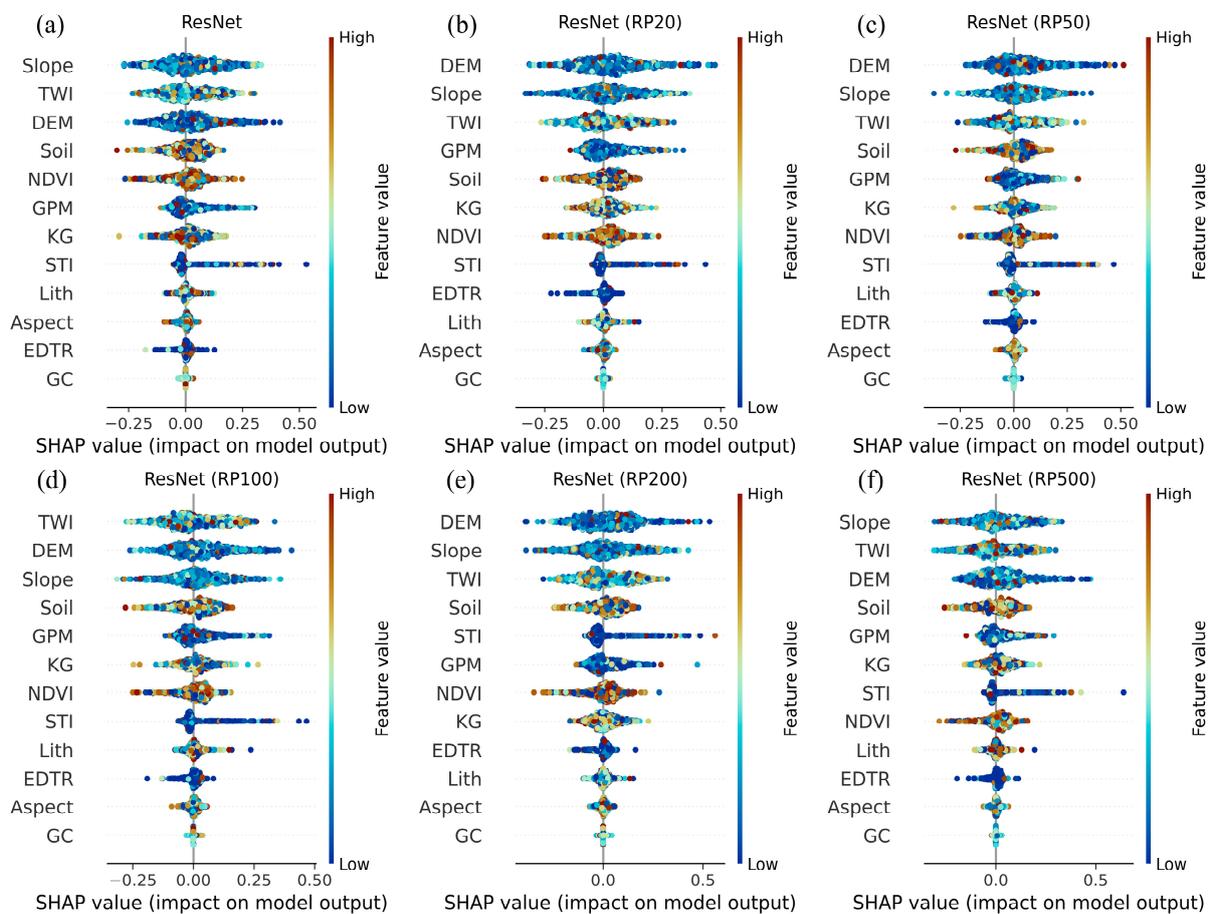


Figure 5. The importance of environmental factors in (a) ResNet, (b) ResNet (RP20), (c) ResNet (RP50), (d) ResNet (RP100), (e) ResNet (RP200), (f) ResNet (RP500).

In comparing the interaction value matrices of the two models (Figure 6), the interaction value commonly increased in ResNet (RP50). It can be seen that the interaction values of the DEM and KG (0.094 to 0.105), GPM and slope (0.036 to 0.05), slope and NDVI (0.053 to 0.069), NDVI and KG (0.126 to 0.136), and Lith and KG (0.04 to 0.058) significantly increased in ResNet (RP50), indicating that the influence of these five pairs of features on the model output increased.

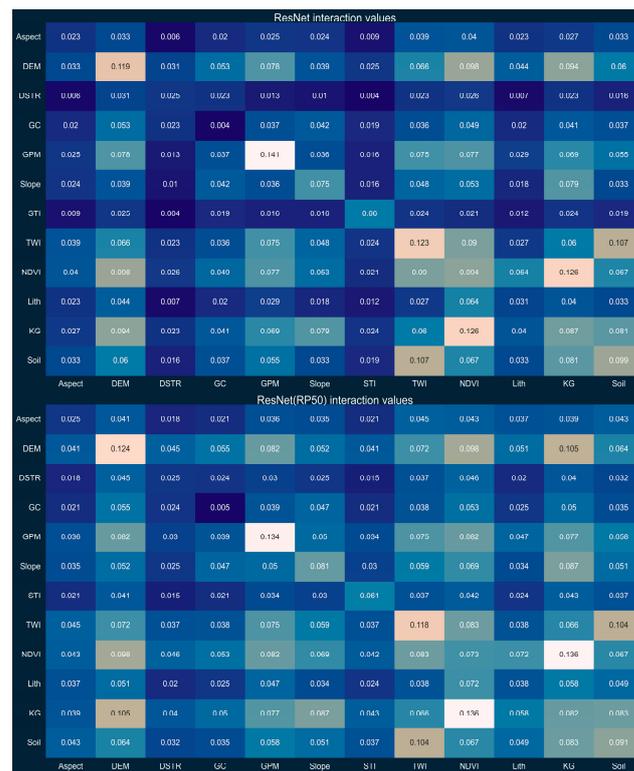


Figure 6. Interaction value matrices of ResNet and ResNet (RP50).

5.5. Model Sensitivity and Uncertainty Analysis

A robust FSM model should have results that do not change significantly within a reasonable range of variations in the input data. To effectively demonstrate that the prediction results of the physics-initialization method are generalizable, two random operations that occurred during the modelling process were analysed to measure the sensitivity and uncertainty of the model. For sensitivity testing, ten random selections of training and testing sets were made (Table 4), and the stacking orders of the flood conditioning factors were randomly changed (Table 5). A total of 81,861 grid cells were selected based on systematic sampling with a periodic interval of 20 of the total grids.

Table 4. Results of ResNet (RP50) carried out 10 times with different training/test sets.

Experiment	1	2	3	4	5	6	7	8	9	10	Min	Max	SD	Average
AUC	0.932	0.935	0.937	0.934	0.934	0.935	0.935	0.935	0.932	0.933	0.932	0.937	0.001	0.934
Accuracy	0.854	0.856	0.861	0.856	0.858	0.858	0.857	0.860	0.854	0.854	0.854	0.861	0.002	0.857
Specificity	0.918	0.910	0.909	0.909	0.914	0.919	0.919	0.923	0.907	0.903	0.903	0.923	0.006	0.913
Sensitivity	0.733	0.761	0.746	0.750	0.810	0.795	0.791	0.793	0.748	0.758	0.733	0.810	0.026	0.768

Table 5. Results of ResNet (RP50) carried out 10 times with different stacking orders of flood conditioning factors.

Experiment	1	2	3	4	5	6	7	8	9	10	Min	Max	SD	Average
AUC	0.932	0.925	0.927	0.929	0.930	0.928	0.931	0.931	0.928	0.927	0.925	0.932	0.002	0.929
Accuracy	0.854	0.846	0.846	0.849	0.851	0.849	0.852	0.852	0.849	0.848	0.846	0.854	0.003	0.850
Specificity	0.907	0.928	0.933	0.915	0.919	0.883	0.913	0.909	0.893	0.898	0.883	0.933	0.015	0.910
Sensitivity	0.748	0.682	0.673	0.718	0.717	0.780	0.730	0.738	0.761	0.747	0.673	0.780	0.033	0.729

The experiment showed that all the evaluation criteria have reasonable fluctuations. The mean and standard deviation (SD) of the AUC were 0.934 and 0.001, and those of

accuracy were 0.857 and 0.002, respectively. Compared with the results of ResNet (RP50) with different training/test sets, the results of ResNet (RP50) with different factor stacking orders fluctuated slightly more.

Comparing the mean value of 10 susceptibility estimates and a single susceptibility estimate (Figure 7a,b) proved that the correlation between them is very high, and the two uncertainty scenarios showed a very high correlation ($r^2 = 0.98$ and $r^2 = 0.98$) between the single susceptibility estimate and the average susceptibility estimate, indicating that the susceptibility predicted by ResNet (RP50) is extremely robust.

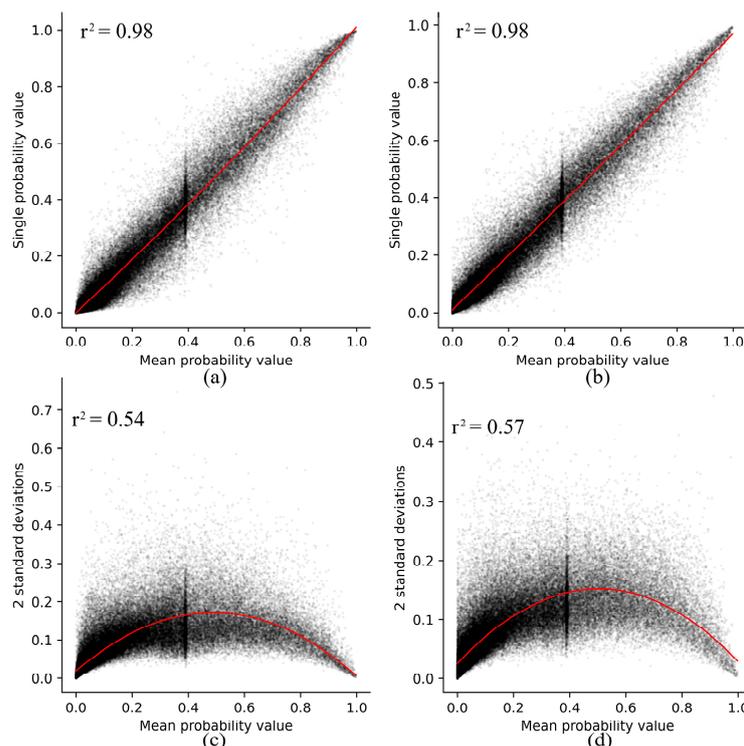


Figure 7. Comparison between a single susceptibility estimate and the mean susceptibility estimate: (a) The mean susceptibility estimate was calculated based on 10 estimates acquired from different training/test sets. (b) The mean susceptibility estimate was calculated from 10 estimates acquired from different stacking orders of the factors. (c) The x-axis shows the mean susceptibility estimate of the 10 estimates carried out with different training/test sets. The y-axis shows the 2SD of the susceptibility estimate. (d) The x-axis denotes the mean susceptibility estimate of 10 estimates derived from different stacking orders of flood conditioning factors. The y-axis shows the 2SD of the susceptibility estimate.

To quantify the uncertainty of flood prediction methods, the measure strategy proposed in [50] was applied. Figure 6c,d show a plot of the mean susceptibility estimate on the x-axis against two standard deviations (2SD) of the susceptibility estimate on the y-axis. The 2SD value increases from very low to moderate susceptibility and then decreases to very high susceptibility (Figure 7c,d). Specifically, the 2SD value is relatively low for the low (<0.3) and high susceptibility zones (<0.25), which indicates that ResNet (RP50) can be stably predicted in these two susceptible zones.

6. Conclusions

The hybridization of ML and hydrological models is necessary to enhance the robustness and reliability of flood prediction. In this study, a global flood susceptibility map was made by combining a 2D hydrological model with ResNet-18 using a new framework according to RS datasets. The models were trained using flood inundation maps generated by a 2D hydrological model with RPs of 20 years, 50 years, 100 years, 200 years, and 500 years

as pretraining data for transfer learning. The models with RPs of 20 years, 100 years, 200 years, and 500 years for physics initialization avoided settling into a local optimum and showed improved accuracy, particularly the model pretrained with RPs of 50 years, which showed the most significant improvement. The AUC and accuracy were improved from 0.928 and 0.851 to 0.932 and 0.854, respectively. Transfer learning could significantly improve the model performance with incomplete training data, which are common in the real world. The flood susceptibility map generated by this hybrid model was also closer to that in another global flood event dataset, which indicates an improvement in addressing the problem of biased prediction for ML. It was found that the feature importance of the models using transfer learning changed according to the SHAP post hoc test. The model was insensitive to the randomness of the training and test set splitting process and the stacking order of the factors. Finally, this paper provides a perspective on combining hydrological models with deep learning, which can be extended to other disaster studies.

Author Contributions: All authors contributed to the study's design. Data collection, experimental execution and analysis were performed by J.L. under the supervision of K.L. and M.W. All authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated and analysed during the current study are available in public data repositories and do not require any licences. Table 1 provides the details for accessing each dataset.

Conflicts of Interest: The authors have no financial or proprietary interests in any material discussed in this article.

References

- Jevrejeva, S.; Jackson, L.P.; Grinsted, A.; Lincke, D.; Marzeion, B. Flood damage costs under the sea level rise with warming of 1.5 °C and 2 °C. *Environ. Res. Lett.* **2018**, *13*, 074014. [\[CrossRef\]](#)
- Rentschler, J.; Salhab, M.; Jafino, B.A. Flood exposure and poverty in 188 countries. *Nat. Commun.* **2022**, *13*, 3527. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kwak, Y.; Park, J.; Fukami, K. Near Real-Time Flood Volume Estimation From MODIS Time-Series Imagery in the Indus River Basin. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 578–586. [\[CrossRef\]](#)
- Ahmadlou, M.; Karimi, M.; Alizadeh, S.; Shirzadi, A.; Parvinnejhad, D.; Shahabi, H.; Panahi, M. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto Int.* **2018**, *34*, 1252–1272. [\[CrossRef\]](#)
- Swain, K.C.; Singha, C.; Nayak, L. Flood Susceptibility Mapping through the GIS-AHP Technique Using the Cloud. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 720. [\[CrossRef\]](#)
- Mudashiru, R.B.; Sabtu, N.; Abustan, I.; Balogun, W. Flood hazard mapping methods: A review. *J. Hydrol.* **2021**, *603*, 126846. [\[CrossRef\]](#)
- Bellos, V. Ways for flood hazard mapping in urbanised environments: A short. *Water Util. J.* **2012**, *4*, 25–31.
- Ramírez, J.A. Chapter 11: Prediction and modeling of flood hydrology and hydraulics. In *Inland Flood Hazards: Human, Riparian and Aquatic Communities*; Wohl, E.E., Ed.; Cambridge University Press: Cambridge, UK, 2000; Volume 498, pp. 293–333.
- Hong, H.; Tsangaratos, P.; Ilija, 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. [\[CrossRef\]](#)
- Anees, M.T.; Abdullah, K.; Nordin, M.N.M.; Rahman, N.N.N.A.; Syakir, M.I.; Kadir, M.O.A. One- and Two-Dimensional Hydrological Modelling and Their Uncertainties. *Flood Risk Manag.* **2017**, *11*, 221–244.
- Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Tien Bui, D. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total Environ.* **2018**, *627*, 744–755. [\[CrossRef\]](#)
- Tehrany, M.S.; Pradhan, B.; Mansor, S.; Ahmad, N. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA* **2015**, *125*, 91–101. [\[CrossRef\]](#)
- Liu, K.; Li, Z.; Yao, C.; Chen, J.; Zhang, K.; Saifullah, M. Coupling the k-nearest neighbor procedure with the Kalman filter for real-time updating of the hydraulic model in flood forecasting. *Int. J. Sediment Res.* **2016**, *31*, 149–158. [\[CrossRef\]](#)
- Pan, T.-Y.; Yang, Y.-T.; Kuo, H.-C.; Tan, Y.-C.; Lai, J.-S.; Chang, T.-J.; Lee, C.-S.; Hsu, K.H. Improvement of watershed flood forecasting by typhoon rainfall climate model with an ANN-based southwest monsoon rainfall enhancement. *J. Hydrol.* **2013**, *506*, 90–100. [\[CrossRef\]](#)

15. La Salandra, M.; Colacicco, R.; Dellino, P.; Capolongo, D. An Effective Approach for Automatic River Features Extraction Using High-Resolution UAV Imagery. *Drones* **2023**, *7*, 70. [[CrossRef](#)]
16. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
17. Hosseini, F.S.; Choubin, B.; Mosavi, A.; Nabipour, N.; Shamshirband, S.; Darabi, H.; Haghghi, A.T. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Sci. Total Environ.* **2020**, *711*, 135161. [[CrossRef](#)]
18. Razavi Termeh, S.V.; Kornejady, A.; Pourghasemi, H.R.; Keesstra, S. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Sci. Total Environ.* **2018**, *615*, 438–451. [[CrossRef](#)]
19. Bui, Q.-T.; Nguyen, Q.-H.; Nguyen, X.L.; Pham, V.D.; Nguyen, H.D.; Pham, V.-M. Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *J. Hydrol.* **2020**, *581*, 124379. [[CrossRef](#)]
20. Schumann, G.J. Preface: Remote sensing in flood monitoring and management. *Remote Sens.* **2015**, *7*, 17013–17015. [[CrossRef](#)]
21. Mosavi, A.; Ozturk, P.; Chau, K.-w. Flood Prediction Using Machine Learning Models: Literature Review. *Water* **2018**, *10*, 1536. [[CrossRef](#)]
22. Yang, L.; Cervone, G. Analysis of remote sensing imagery for disaster assessment using deep learning: A case study of flooding event. *Soft Comput.* **2019**, *23*, 13393–13408. [[CrossRef](#)]
23. Liu, J.; Wang, J.; Xiong, J.; Cheng, W.; Sun, H.; Yong, Z.; Wang, N. Hybrid Models Incorporating Bivariate Statistics and Machine Learning Methods for Flash Flood Susceptibility Assessment Based on Remote Sensing Datasets. *Remote Sens.* **2021**, *13*, 4945. [[CrossRef](#)]
24. Zhou, Y.; Cui, Z.; Lin, K.; Sheng, S.; Chen, H.; Guo, S.; Xu, C.-Y. Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. *J. Hydrol.* **2022**, *604*, 127255. [[CrossRef](#)]
25. Yang, T.; Sun, F.; Gentine, P.; Liu, W.; Wang, H.; Yin, J.; Du, M.; Liu, C. Evaluation and machine learning improvement of global hydrological model-based flood simulations. *Environ. Res. Lett.* **2019**, *14*, 114027. [[CrossRef](#)]
26. Liu, J.; Yuan, X.; Zeng, J.; Jiao, Y.; Li, Y.; Zhong, L.; Yao, L. Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 265–278. [[CrossRef](#)]
27. Sharma, S.; Raj Ghimire, G.; Siddique, R. Machine learning for postprocessing ensemble streamflow forecasts. *J. Hydroinform.* **2023**, *25*, 126–139. [[CrossRef](#)]
28. Jia, X.; Zwart, J.; Sadler, J.; Appling, A.; Oliver, S.; Markstrom, S.; Willard, J.; Xu, S.; Steinbach, M.; Read, J. Physics-guided recurrent graph model for predicting flow and temperature in river networks. In Proceedings of the 2021 SIAM International Conference on Data Mining (SDM), Virtual Event, 29 April–1 May 2021; pp. 612–620.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [[CrossRef](#)]
31. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
32. De Roo, A. Modelling runoff and sediment transport in catchments using GIS. *Hydrol. Process.* **1998**, *12*, 905–922. [[CrossRef](#)]
33. 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](#)]
34. Chen, W.; Hong, H.; Li, S.; Shahabi, H.; Wang, Y.; Wang, X.; Ahmad, B.B. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* **2019**, *575*, 864–873. [[CrossRef](#)]
35. Tehrani, M.S.; Jones, S.; Shabani, F. Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *Catena* **2019**, *175*, 174–192. [[CrossRef](#)]
36. Fang, Z.; Wang, Y.; Peng, L.; Hong, H. Predicting flood susceptibility using LSTM neural networks. *J. Hydrol.* **2021**, *594*, 125734. [[CrossRef](#)]
37. Podhorányi, M.; Unucka, J.; Bobál, P.; Říhová, V. Effects of LIDAR DEM resolution in hydrodynamic modelling: Model sensitivity for cross-sections. *Int. J. Digit. Earth* **2013**, *6*, 3–27. [[CrossRef](#)]
38. La Salandra, M.; Roseto, R.; Mele, D.; Dellino, P.; Capolongo, D. Probabilistic hydro-geomorphological hazard assessment based on UAV-derived high-resolution topographic data: The case of Basento river (Southern Italy). *Sci. Total Environ.* **2022**, *842*, 156736. [[CrossRef](#)] [[PubMed](#)]
39. Werner, M.G.F.; Hunter, N.M.; Bates, P.D. Identifiability of distributed floodplain roughness values in flood extent estimation. *J. Hydrol.* **2005**, *314*, 139–157. [[CrossRef](#)]
40. Billi, P. Flash flood sediment transport in a steep sand-bed ephemeral stream. *Int. J. Sediment Res.* **2011**, *26*, 193–209. [[CrossRef](#)]
41. 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. J.* **1979**, *24*, 43–69. [[CrossRef](#)]
42. Kopecký, M.; Macek, M.; Wild, J. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Sci. Total Environ.* **2021**, *757*, 143785. [[CrossRef](#)]
43. Tellman, B.; Sullivan, J.A.; Kuhn, C.; Kettner, A.J.; Doyle, C.S.; Brakenridge, G.R.; Erickson, T.A.; Slayback, D.A. Satellite imaging reveals increased proportion of population exposed to floods. *Nature* **2021**, *596*, 80–86. [[CrossRef](#)]
44. Brakenridge, G.; Karnes, D. The Dartmouth Flood Observatory: An electronic research tool and electronic archive for investigations of extreme flood events. *Geosci. Inf. Soc. Proc.* **1996**, *27*, 31–36.

45. He, F.; Liu, T.; Tao, D. Why resnet works? residuals generalize. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 5349–5362. [[CrossRef](#)] [[PubMed](#)]
46. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
47. Dottori, F.; Salamon, P.; Bianchi, A.; Alfieri, L.; Hirpa, F.A.; Feyen, L. Development and evaluation of a framework for global flood hazard mapping. *Adv. Water Resour.* **2016**, *94*, 87–102. [[CrossRef](#)]
48. Hand, D.J.; Till, R.J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **2001**, *45*, 171–186. [[CrossRef](#)]
49. Pan, M.; Lin, P.; Beck, H.E.; Zeng, Z.; Yamazaki, D.; David, C.H.; Lu, H.; Yang, K.; Hong, Y.; Wood, E.F. Global Reach-Level 3-Hourly River Flood Reanalysis (1980–2019). *Bull. Am. Meteorol. Soc.* **2021**, *102*, E2086–E2105. [[CrossRef](#)]
50. Guzzetti, F.; Galli, M.; Reichenbach, P.; Ardizzone, F.; Cardinali, M. Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. *Nat. Hazards Earth Syst. Sci.* **2006**, *6*, 115–131. [[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.