remote sensing

MDPI

Article

Flood Inundation Probability Estimation by Integrating Physical and Social Sensing Data: Case Study of 2021 Heavy Rainfall in Henan, China

Wenying Du [1,*], Qingyun Xia [1], Bingqing Cheng [1], Lei Xu [1], Zeqiang Chen [1], Xiang Zhang [1,2,3], Min Huang [3,4] and Nengcheng Chen [1,2]

[1] National Engineering Research Centre of Geographic Information System, School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China; qingyunx@cug.edu.cn (Q.X.); chengbingqing@cug.edu.cn (B.C.); xulei10@cug.edu.cn (L.X.); chenzeqiang@cug.edu.cn (Z.C.); zhangxiang76@cug.edu.cn (X.Z.); chennengcheng@cug.edu.cn (N.C.)
[2] Hubei Luojia Laboratory, Wuhan 430079, China
[3] SongShan Laboratory, Zhengzhou 450046, China; huangm@jxnu.edu.cn
[4] School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China
* Correspondence: duwenying@cug.edu.cn

Abstract: Frequent flooding seriously affects people's safety and economic construction, and assessing the inundation probability can help to strengthen the capacity of emergency management of floods. There are currently two general means of flood sensing: physical and social. Remote sensing data feature high reliability but are often unavailable in disasters caused by persistent heavy rainfall. Social media is characterized by high timeliness and a large data volume but has high redundancy and low reliability. The existing studies have primarily relied on physical sensing data and have not fully exploited the potential of social media data. This paper combines traditional physical sensing data with social media and proposes an integrated physical and social sensing (IPS) method to estimate the probability distribution of flood inundation. Taking the "7·20" Henan rainstorm in 2021 and the study area of Xinxiang, China, as a case study, more than 60,000 messages and 1900 images about this occurrence were acquired from the Weibo platform. Taking filtered water depth points with their geographic location and water depth information as the main input, the inverse distance attenuation function was used to calculate the inundation potential layer of the whole image. Then, the Gaussian kernel was used to weight the physical sensing data based on each water depth point, and finally, the submergence probability layer of the whole image was enhanced. In the validation of the results using radar and social media points, accuracies of 88.77% and 75% were obtained by setting up a threshold classification, demonstrating the effectiveness and usefulness of the method. The significance of this study lies in obtaining discrete social media flood points and achieving space-continuous flood inundation probability mapping, providing decision-making support for urban flood diagnosis and mitigation.

Keywords: flood; social sensing; physical sensing; water depth point; probability estimate

check for updates

Citation: Du, W.; Xia, Q.; Cheng, B.; Xu, L.; Chen, Z.; Zhang, X.; Huang, M.; Chen, N. Flood Inundation Probability Estimation by Integrating Physical and Social Sensing Data: Case Study of 2021 Heavy Rainfall in Henan, China. *Remote Sens.* **2024**, *16*, 2734. https://doi.org/10.3390/rs16152734

Academic Editor: Raffaele Albano

Received: 5 June 2024
Revised: 24 July 2024
Accepted: 24 July 2024
Published: 26 July 2024

Copyright: © 2024 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

According to the Fifth Assessment Report of the United Nations Intergovernmental Panel on Climate Change, global warming will exacerbate the risks associated with extreme events. The intensity and frequency of extreme heavy precipitation will increase in the mid-latitude and tropical humid regions, and flooding disasters will become more frequent [1]. According to the report, flood losses accounted for 29.59% of global disaster losses in 2021, significantly surpassing other disasters, e.g., drought, earthquakes, and fires. With the increased frequency of extreme precipitation events caused by the changing climate and environment, flooding disasters are becoming more frequent [2].

Recent studies in Henan Province have made significant progress in monitoring and understanding the dynamics, causes, and risk assessment of the 2021 Henan rainstorm. Researchers have utilized various methods, including fuzzy mathematics, remote sensing, and satellite data analysis, to evaluate flood risk and impact. For instance, the multiscale dynamic processes of the "7·21" torrential rainfall event were analyzed using advanced diagnostic tools, revealing critical insights into energy conversions and mesoscale influences [3]. Wu et al. [4] conducted an in-depth examination of the extreme rainfall in July 2021 in Henan, utilizing advanced methodologies to reveal the critical influence of a stationary low-level jet stream and regional topography on the event's magnitude, while assessing its predictability and the amplified climate risks anticipated under projected climate change scenarios. Additionally, the verification of the SMS-WARMS forecast model for the "7·19" heavy rainstorm showed the model's capability in accurately predicting precipitation centers, intensities, and timing, which is crucial for early warning and disaster preparedness [5].

There are many models and methods for flood assessment, such as the construction of physical models using physically sensed data [6–9], whereas, in hydrology, it is prevalent to build appropriate models to simulate the flood inundation process, including the commonly used MIKE, SWAT, HEC-RAS, etc. [10–12]. With the advancements in computer technology, integrated software has been developed for each of these models, e.g., MIKE 11, HEC-RAS, Delft3D, etc. [10–13]. However, hydrological models require a large amount of physical sensing data input, and the models are usually applied to specific areas, i.e., a watershed. The hydrological scale affects the evolution of hydrological processes during spatial and temporal variations. In addition, the stochastic nature of hydrological variables and model parameters often leads to uncontrollable uncertainty problems in the model [14].

Remote sensing is a powerful observational means to record information about a target area. It has long been utilized to monitor flood coverage and its dynamic evolution. There are several remotely sensed data, i.e., rainfall, soil moisture, etc., and they can be used to determine probable inundation zones and flood risks [15–18]. Spectral, spatial, and temporal resolutions vary among remote sensing images, and they are usually produced using different equipment, mainly including multispectral sensors, optical sensors, and synthetic aperture radar. Among them, optical images, i.e., multispectral and hyperspectral, are limited by satellites' revisit period and the coverage of thick cumulus clouds during heavy rainfall. This results in a low temporal resolution and the inability to acquire the appropriate data, making it difficult to satisfy the application of real-time flood analysis. However, synthetic aperture radar can penetrate thick cumulus clouds and provide valuable information on shallow water and shaded areas [19]. Although contemporary radar imagery is spatiotemporally sparse, it is undeniable that it can still be utilized to extract flooded water bodies [20–22]. However, radar images are usually used for post-disaster assessment, and it is still of great necessity to require timely ground observations during disasters.

All individuals possess the capacity to observe and transmit information. By engaging citizens, it is possible to broaden access to information and convert it into easy-to-understand maps and status reports. This information can provide effective assistance to responders and emergency managers [23]. As a result, many scholars have begun to use crowdsourced geographic information and social media data (SMD) with geotagged data, e.g., Twitter, Facebook, and Sina Weibo (a popular social media platform in China), as supplement data sources. The data obtained from these platforms are widely used in disaster assessment and analysis due to their substantial volume, rapid transmission speed, and extensive diversity. When people publish images and texts on Weibo with geographic locations, location information can be obtained by analyzing these texts and images [24]. Fang et al. transformed social media data into a proxy variable for flooding to evaluate the performance of the proposed rapid flood model [25]. Resch et al. and Wang et al. employed machine learning models, e.g., latent Dirichlet allocation, SVM (support vector machine),

RF (random forest), KNN (K-nearest neighbor), etc., to extract rescue requests from social media messages and estimated natural disaster damage [26,27].

There are also many studies incorporating social media data sources into flood assessment. Based on a Bayesian model, Rosser et al. [28] combined social media data, remotely sensed imagery, and topographic data to achieve rapid flood mapping. Xu et al. [29] used social media data to refine flooding probability maps generated from digital elevation models and historical flooding points. Panteras et al. [30] used texts and images to overcome the limitations of the low spatial resolution of satellites. The uncertainties of crowdsourced social media data can be estimated by data quality assessment performed manually or by artificial intelligence. The uncertainty evaluation of such data enables them to complement traditional observations and improve flood management activities [31,32].

Taking the limitations of the above-mentioned means into consideration, this study proposes an integrated physical and social sensing (IPS) method to estimate the flood inundation probability. The IPS method uses social media points (SMPs) to generate fundamental inundation potential layers, subsequently enhancing the inundation probability of the entire image based on the weight of the physical sensing data. The contribution of this study is the integration of geo-basic data with SMPs, using submerged points and pre-event physical data to generate flood probability maps with updatability and spatial continuity. Employing IPS methodology, this research integrates traditional physical sensor data with social media-derived information, aiming to advance the field of flood risk assessment. Specifically, the investigation focuses on the analysis of extensive datasets from the "7·20" Henan rainstorm event in 2021, encompassing over 60,000 textual entries and nearly 1900 visual records obtained via the Weibo platform. The overarching goal was to develop a space-continuous model for estimating flood susceptibility, facilitating more effective and timely decision-making in flood management.

The remainder of the paper is structured as follows: Section 2 provides the study area, experimental data, and the proposed method. Section 3 goes into detail about the results of the case study. Section 4 validates the results with social media points and flooded water bodies derived from Gaofen-3 radar images. Section 5 presents the conclusions and outlook.

2. Materials and Methods

2.1. Study Area

From 20 July to 21 July 2021, Henan, China suffered from extreme heavy rainfall, with Zhengzhou and Xinxiang experiencing extraordinarily heavy rainfall. Observations from ten national meteorological stations in Zhengzhou, Xinxiang, and Kaifeng of Henan broke their historic records. Natural disaster relief reached a level I response. The study area, as seen in Figure 1, covers the main watershed area in Xinxiang City. The Wei River is a tributary of the Haihe River, which is the largest river system in North China and one of the seven major rivers in China. Its total length is more than 400 km, of which the main stream is 344.5 km long, and the basin area is 14,970 square kilometers. The study area is the section of the Wei River that runs through Xinxiang, with a total length of about 65 km and a coverage area of about 481 square kilometers.

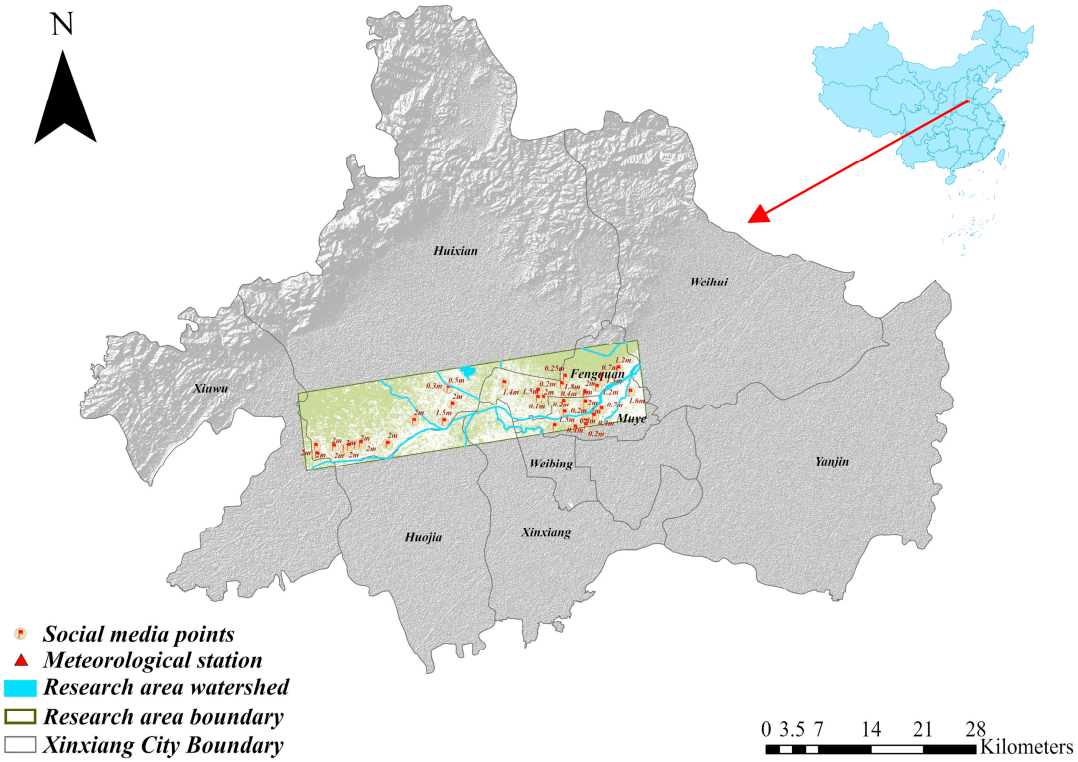


Figure 1. Geographic location of the study area and spatial distribution of SMPs and meteorological stations.

2.2. Experimental Data

2.2.1. Data Collection

The river system in the study area, the National Aeronautics and Space Administration's digital elevation model (DEM) and hour-by-hour observations from ground meteorological stations in China were used as the raw geographic information base data for the flood inundation analysis (Table 1). Data from Sina Weibo during the 2021 rainstorm event in Henan, China were used to extract the geographic names and water depths. Weibo posts and pictures from 00:00 on 17 July to 24:00 on 31 July in 2021 about the flooding in Henan were collected by combining Web Crawler technology and the Weibo Application Programming Interface (API). There were, in total, 63,986 texts and 1901 pictures about Xinxiang, and they were all used in this case study.

Table 1. List of data sources.

Data Type	Item	Data Format	Time (Year)	Sources	Notes
Physical sensing data	River system	Vector	2021	Open Street Map https://www.openstreetmap.org/	Three-level river basin
	DEM	Raster	2020	https://search.earthdata.nasa.gov/	Spatial resolution: 30 m
	Precipitation	Txt	2021	China Meteorological Data Service Centre https://data.cma.cn/	Temporal resolution: 3 h
Social sensing data	Weibo texts	Txt	2021	https://weibo.com/	/
	Weibo pictures	Image	2021		/

2.2.2. Data Processing

Physical Sensing Data Processing

Some conventional parameters, such as the slope, curvature, topographic wetness index (TWI), and stream power index (SPI), were computed using DEM in this work to better evaluate the impact of the topographic elements on the probability of flood inundation. In general, the areas threatened by flooding are smaller in areas with greater topographic relief. In plain areas, the lower the elevation, the greater the danger [33]. The slopes are smaller, the water flows more slowly, and it accumulates more easily [34]. Curvature is detrimental to water discharge because it affects the surface runoff and infiltration more the closer the value is to 0 [33]. The TWI is the influence of the regional topography on the runoff flow direction and the accumulation of physical indicators [33]. The SPI is the erosive force of water flow, with greater SPI values suggesting that the runoff concentration may contribute to soil erosion [33]. The spatial distance analysis module of ArcGIS 10.6 was used to calculate the Euclidean distance from each cell to the nearest source. These vector data from the river network were then transformed into grid image values.

The dataset from 20 July 2021 in Henan Province of the hour-by-hour observations of Chinese surface weather stations was processed. Then, Kriging interpolation was used to input the rainfall data covering the study area into ArcGIS 10.6. The coordinated system of all the data layers was transformed into the WGS_1984_World_Mercator. The size, i.e., the resolution and numbers of rows and columns, of the seven Eigenfactor layers were made consistent using the ArcGIS 10.6 Mask tool. All the data layers were resized to 30 m by using the ArcGIS 10.6 Resampling tool. All the data layers are called the feature information (Figure 2).

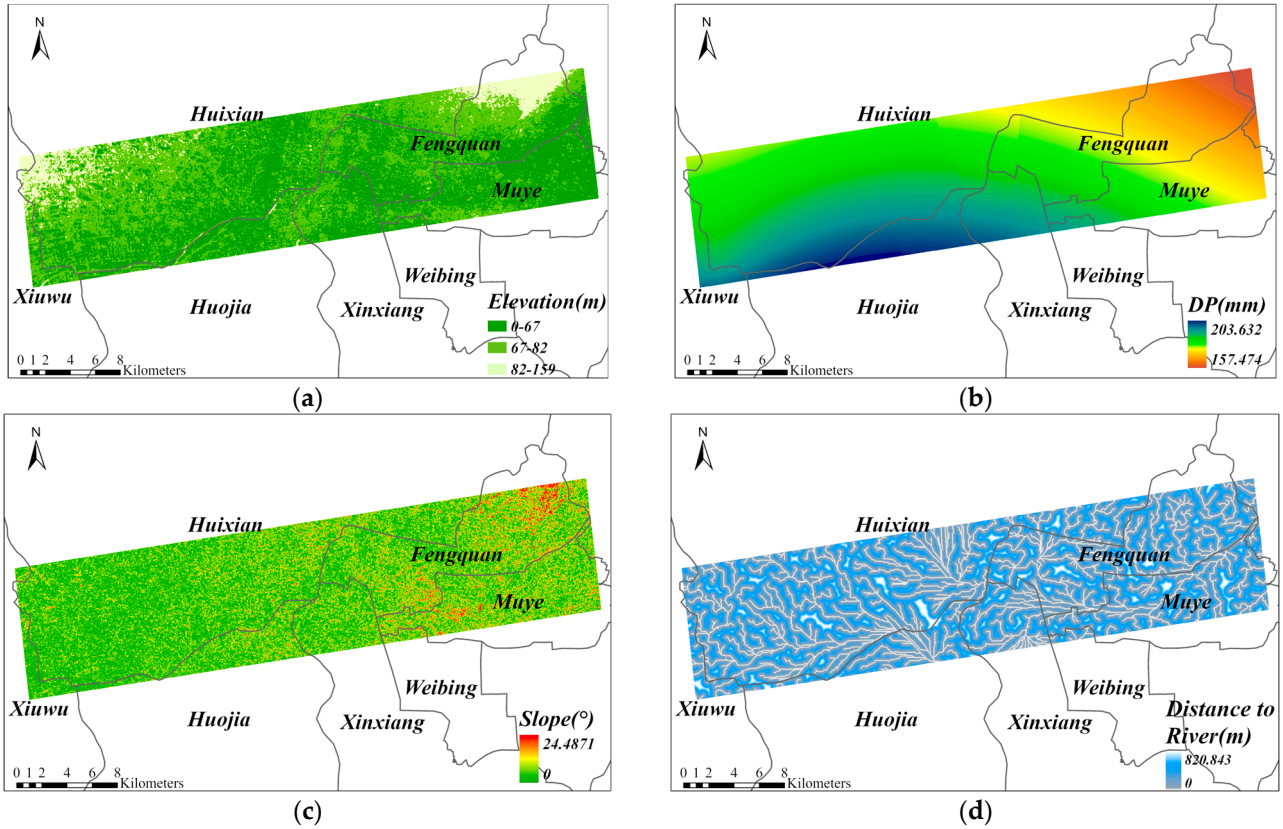


Figure 2. *Cont.*

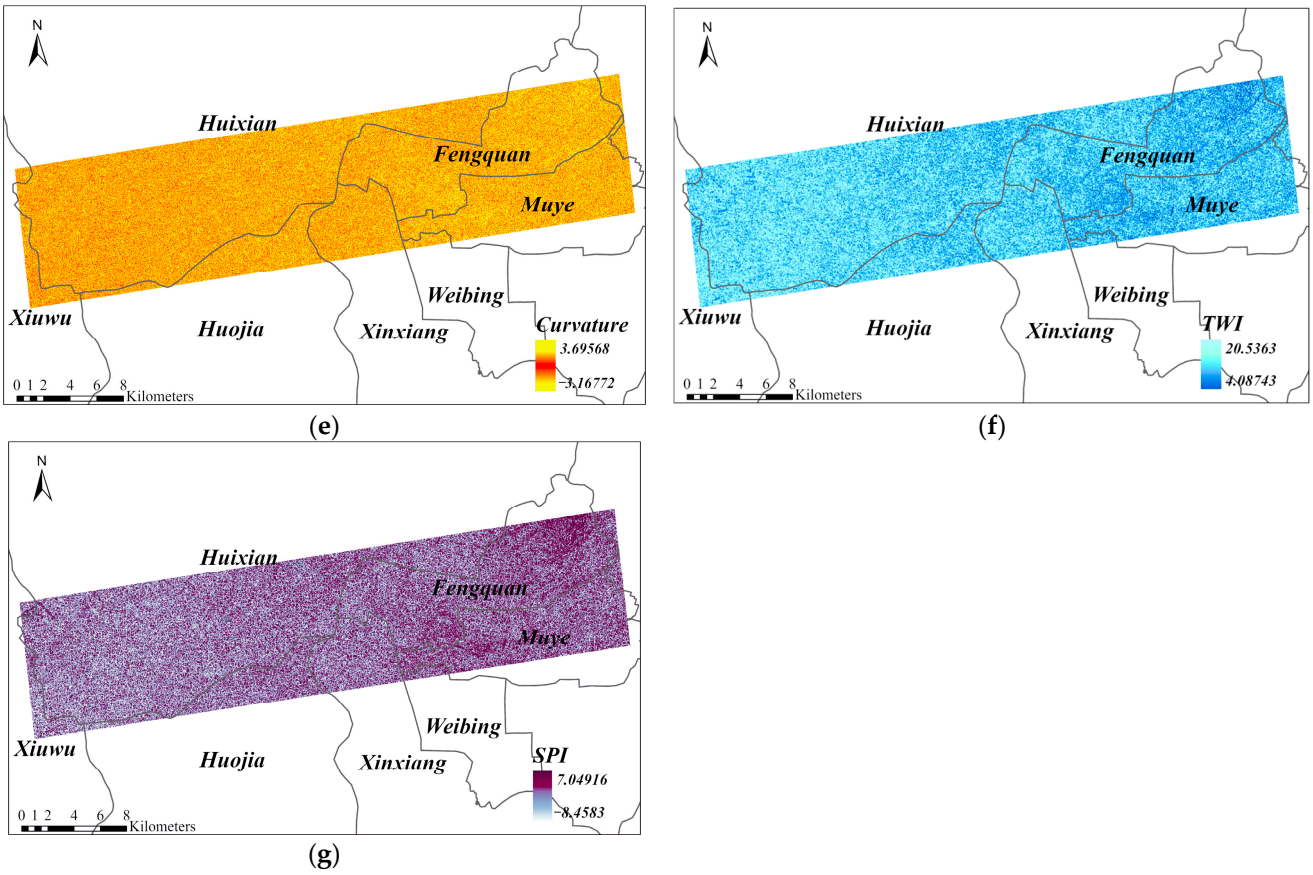


Figure 2. Spatial distribution of feature information: (**a**) elevation (m); (**b**) DP (mm); (**c**) slope (degree); (**d**) distance to river (m); (**e**) curvature; (**f**) TWI; (**g**) SPI.

Typically, not all feature information positively affects the modeling approach, so Geodetector [35] was used for correlation testing to explore whether there was significance of the selected feature information on the water depth values. The *p*-values were less than 0.05, indicating that all the selected features were significant, so no data were eliminated.

Social Sensing Data Processing

(1). Social media texts

To effectively extract beneficial depth and location information from SMD, keyword matching, elimination of spam, and deduplication of the data were applied to the datasets of Xinxiang and Weihui. The postings (noiseless) related to water depth, i.e., keyword(s) containing "water depth" and "* meters", and information on geographic names and water depth were removed. These steps were carried out with the "key word query and match" tools in Excel. The flooded depth and location information was checked by news broadcasts, reporter interviews, police rescues, etc. For instance, information on the Henan rainstorm is available from real-time field interviews by Beijing News reporters on the Tencent website. Real-time field rescue updates were also published by People's Information. Verified information was gathered, as shown in Figure 3. The geographic names were transferred into latitude and longitude coordinates under the appropriate WGS1984 coordinate system utilizing the Baidu API and the coordinate system conversion principle. As shown in Figure 1, these points represent the locations where flooding occurred.

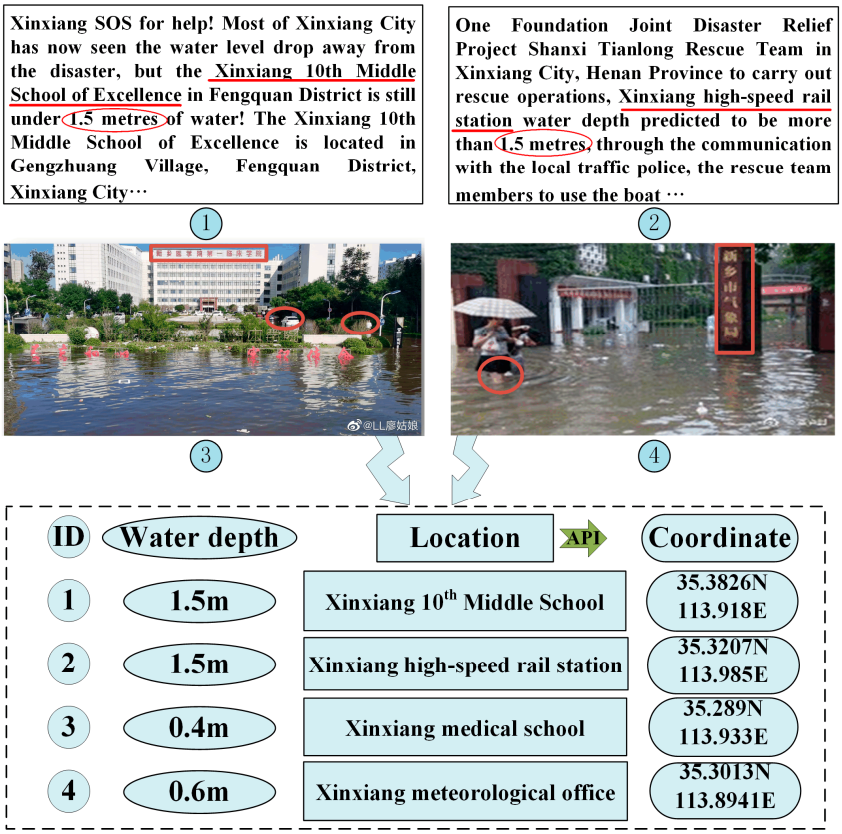


Figure 3. Extraction of the water depth, geographic locations, and latitude/longitude coordinates.

(2). Social media pictures

This part introduces the analysis of the geographical locations and water depths from the images. As shown in Figure 3, images with nameplate texts were selected from the collected images, and then accurate location and water depth information were extracted from them. The locations most similar to the images were obtained by matching the street view of Baidu Map. The acquisition process of the water depth information is presented in Figure 4 and Table 2.

Table 2. Matching water depth values of visual interpretation of water depth case study (adapted from the study by [36]).

Level Name	Range (cm)	Nearest Integer Value (cm)
Level 0	No water	0.0
Level 1	0.0–10.0	10.0
Level 2	10.0–42.5	43.0
Level 3	42.5–85.0	85.0
Level 4	85.0–106.25	106.0
Level 5	106.25–127.25	128.0
Level 6	127.25–148.75	149.0

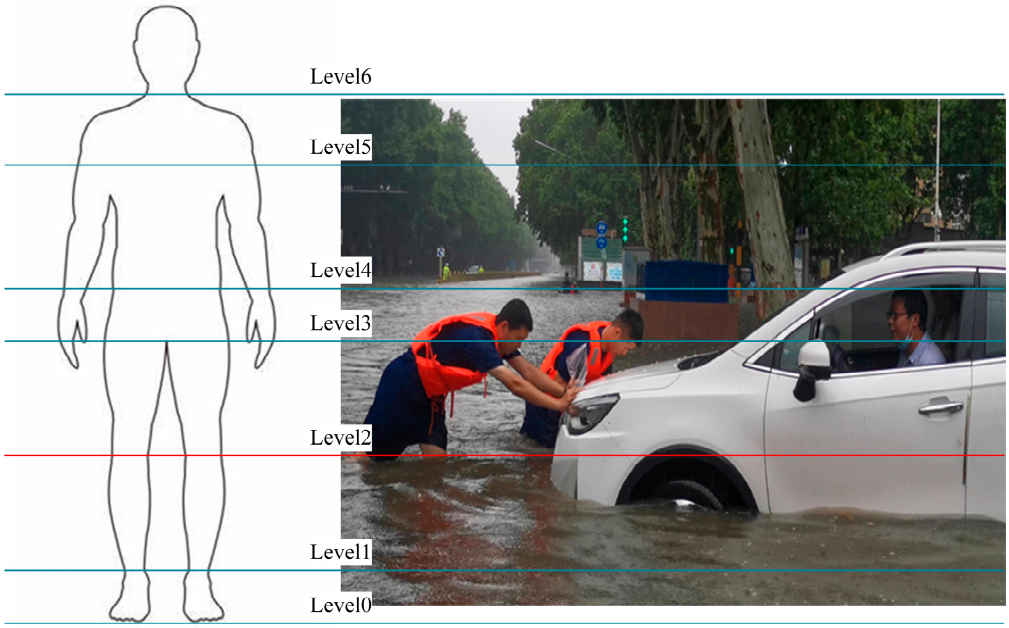


Figure 4. Visual interpretation of water depth case study (adapted from the study by [36]).

2.3. Integrated Physical and Social Sensing (IPS) Method

The IPS method is shown in Figure 5. The inverse distance decay function is introduced for weighting at each flood inundation point by merging the point's inundation water depth data and the research area's DEM data. The method is divided into the following four steps:

(1) The Potential Index Distribution layer is formed centered on the inundation point, which is unique to each individual inundation point, and n sheets of the PID layer are generated for n SMP locations. Normalizing the Potential Index Distribution to the range of [0,1] determines the Probability Index Distribution (PID).
(2) The weight of each generated PID layer is determined by the value of the feature information layer pixel within a certain range around the inundation point location, where the value for the susceptibility to flooding is computed by the feature information using a normalization method, and for the final weight value, a smoother Gaussian surface-based weighting method is introduced.
(3) A thorough flood inundation probability distribution map based on all reported inundation locations in the study area can be created after calculating and averaging the weights of each affecting component in the PID layer.
(4) Finally, the findings of the flood inundation probability distribution estimation for the entire map are validated by integrating the water bodies extracted from the post-event radar images and the social media points used for testing.

2.3.1. IDW-Based Inundation Potential Calculation from SMPs

The inundation potential index is calculated for the entire study area based on each SMP to form an inundation potential distribution map. The calculation method is based on the following guidelines:

(1). The verified flood inundation locations published on the Weibo platform reflect the flood inundation in a certain range of areas, and the closer the area is to the inundation location, the higher the likelihood of flooding is, and conversely, the further the area is from the inundation location, the lower the likelihood of inundation is.
(2). Within a certain range around the SMP, the lower the terrain, the higher the potential of inundation. On the contrary, the higher the terrain, the lower the potential of inundation. The terrain is mainly combined with the elevation data and the inundation water depth data.

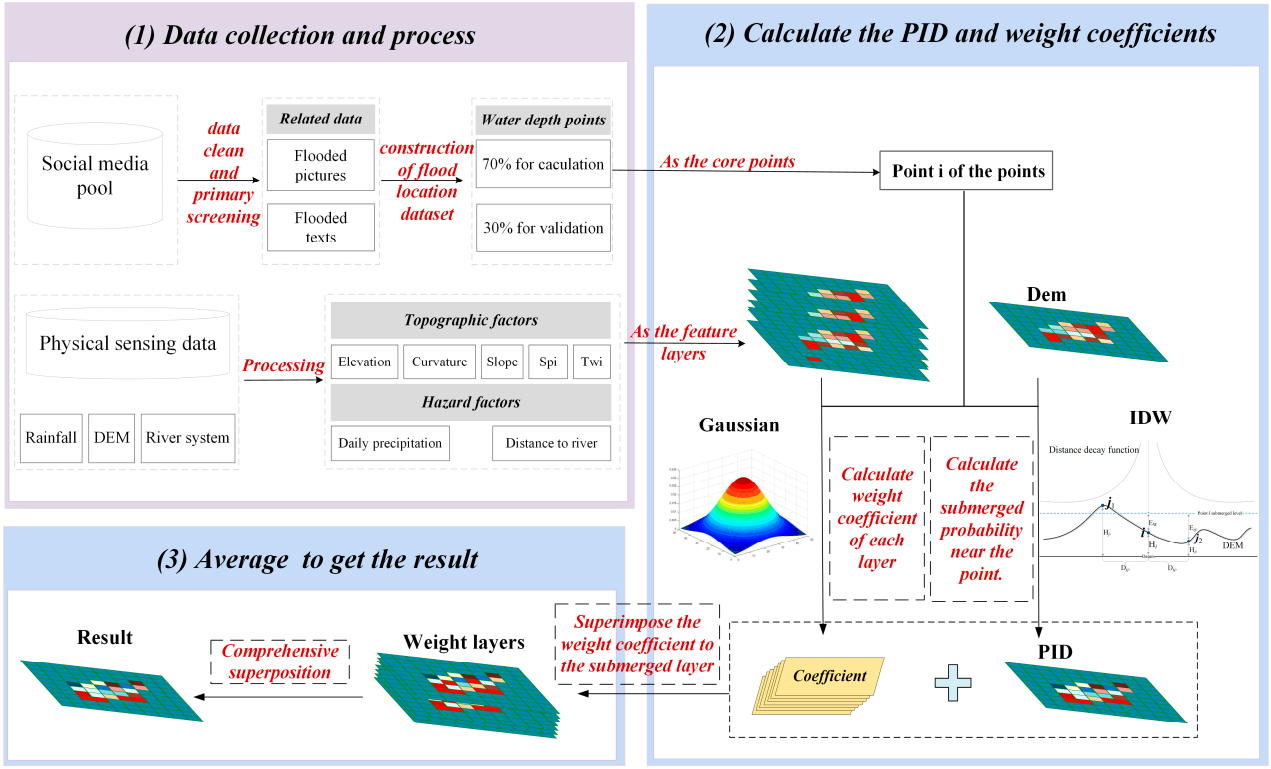


Figure 5. The overall framework of flood inundation probability estimation.

In the three-dimensional space of the study area range, the coordinates of the SMP are set to be $i(x_i, y_i)$, The remaining arbitrary point in the study area is denoted as $j(x_j, y_j)$, In the spatial distribution of the DEM, the elevation of point i is H_i and the elevation of point j is H_j. The depth of inundation at inundation point i is E_w.

Then, the potential of flooding at point j can be expressed as:

$$P_{ij} = (E_{ij})^{\alpha} \times \frac{1}{(D_{ij})^{\beta}} \quad (1)$$

where E_{ij} denotes the depth of inundation at point j and is calculated using the following equation:

$$E_{ij} = \begin{cases} 0, E_w + H_i - H_j < 0 \\ E_w + H_i - H_j, E_w + H_i - H_j > 0 \end{cases} \quad (2)$$

and D_{ij} denotes the Euclidean distance between the submerged point i and any point j:

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3)$$

where the indices α and β are the weighting parameters to regulate the intensity of the effects of E_{ij} and D_{ij} on the results. In this study, the empirical values of $\alpha = 1$ and $\beta = 1/2$ were used.

2.3.2. Gaussian Kernel-Based Integration of Physical Sensing Data and SMPs

As mentioned earlier, based on each SMP location, a normalized inundation PID map can be generated for the extent of the study area, and n probability maps with consistent coverage are ultimately generated based on n point locations. The following section focuses on the methodology and process of synthesizing all the inundation probability maps.

In this study, five topographic factors of elevation—the slope, curvature, SPI, TWI, and rainfall and distance to river, were used as input variables. These can reflect, to a certain

extent, the inundation of the area around the mapped points during floods. This study was based on the following basic guidelines:

(1). It is indisputable that a great extent of flooding in a study area depends on the presence or absence of heavy rainfall events over a long period of time. When rainfall exceeds a certain range, the greater the amount of rainfall is, and the greater the probability that these areas will be inundated.

(2). The topographic factors in the study area greatly affect the rate and direction of flood flow. The areas cannot be drained in a timely manner after rainfall, and the degree of soil erosion will be increased. So, the contribution of the feature information to the flood susceptibility within the range around the SMD can reflect, to a certain extent, the probability of flood inundation in the region.

This study used the probability density function of a two-dimensional Gaussian normal distribution to define the weight of the feature information.

Let the joint probability density of the two-dimensional continuous random variable (x, y) be:

$$\begin{cases} f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{x-\mu_2}{\sigma_2} + \frac{(x-\mu_2)^2}{\sigma_2^2}\right]\right\} \\ -\infty < x < +\infty, -\infty < y < +\infty \end{cases} \quad (4)$$

where μ_1, μ_2 denote the mean values of variable x, y, respectively, σ_1, σ_2 denote the standard deviations of variables x, y; μ_1, μ_2, σ_1, σ_2, and ρ are constants, and $\sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$; thus, it is claimed that (x, y) obeys the two-dimensional normal distribution with the parameters of μ_1, μ_2, σ_1, σ_2, and ρ. The density function of the two-dimensional normal distribution shows an inverted bell-shaped distribution in three-dimensional space and satisfies:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1 \quad (5)$$

Taking the SMP location i as the origin, a 2D Gaussian surface centered at point i is symmetrically distributed along the x and y axes in 3D geospatial space with the mean $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2$ and using the constant ρ, with $\rho = 0$. The joint probability density function is simplified to the following equation:

$$f(x, y) = \frac{1}{2\pi\lambda}\exp\left(-\frac{x^2 + y^2}{2\lambda}\right) \quad (6)$$

where $\lambda = {\sigma_1}^2 = {\sigma_2}^2$ is the parameter of the probability density function, which is used to control the extent of the region of interest around point i to be included in the computation. The smaller λ is, the steeper the inverted-bell shaped two-dimensional Gaussian surface is, and the higher the contribution is of the pixel values to the computation; vice versa, if λ is larger, the number of pixels participating in the computation is larger, and the contribution of the pixels to point i is relatively small.

To determine the optimal value of λ, in this paper, the weight values of the corresponding weighting factors were calculated for the five SMP locations by setting up the radius of the region of interest from 1 cell to 30 cells (900 m). Figure 6 shows the calculation of the saturation value of λ at 25,000 m. With the increase in the region's radius, the factor weighting value reached the saturation value at around 420 m (corresponding to the number of cells of 14). So, the region's radius was set to be 420 m to calculate the factor weighting value around the SMP. Therefore, in this paper, λ = 25,000 and radius = 420 m were input into the function to calculate the weighted probability value.

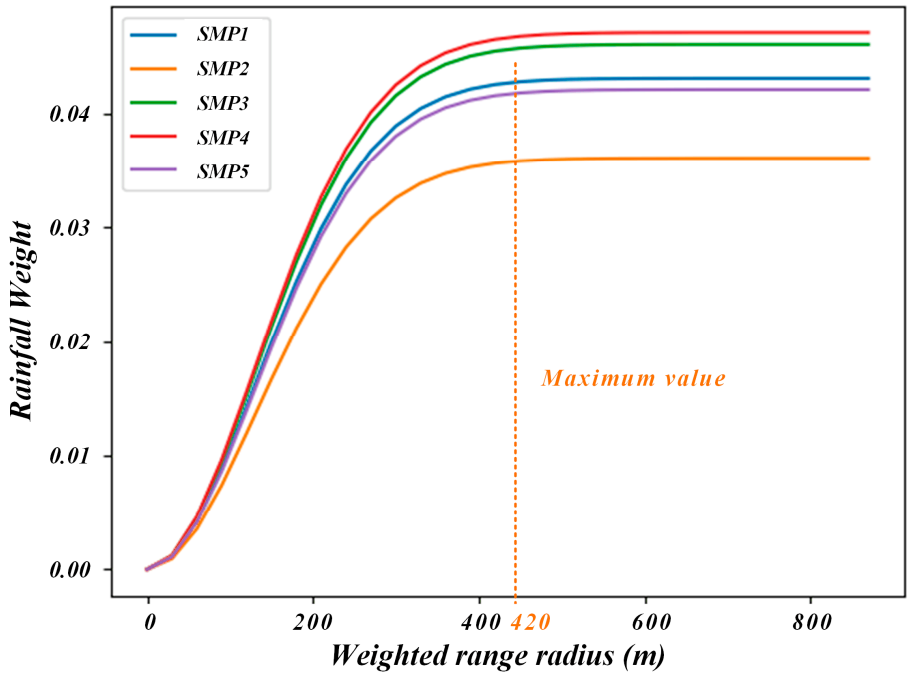


Figure 6. Saturation value test.

For the SMP $(i_x,\ i_y)$ with the Gaussian function as the kernel, the weights of the feature information on its PID layer can be defined as:

$$WoF(i_x,\ i_y) = \iint G\left(j_x,\ j_y\right) F\left(j_x,\ j_y\right) dR \tag{7}$$

where the Gaussian function is established with the SMP $(i_x,\ i_y)$ as the center, position $(j_x,\ j_y)$ is the random pixel within the coverage of the kernel function, i.e., R. $F\left(j_x,\ j_y\right)$ represents the pixel value of certain feature information at this position, and $G\ (j_x,\ j_y)$ is the value of the Gaussian function at this position.

Using the two weighting components (PID and Gaussian kernel) described above, the weight of each feature is integrated into each PID layer and summed to obtain the IPE belonging to each weight feature. Finally, the final IPE is obtained by averaging. The IPE of each weight feature at any location j in the study area is defined as follows:

$$IPE_j = \sum_i WoF_i \times PID_{ij} \tag{8}$$

3. Results

3.1. Extraction Results and Spatial Distribution of SMP

After cleaning the data of the keyword matching, eliminating spam and duplications, 39 available and verified SMP locations in the study area were obtained, all stored with attributes, including their IDs, water depth values, and latitude and longitude coordinates. As shown in Figure 7, 70% of these 39 points were used for training and prediction, and the remaining 30% were used to verify the performance of the model.

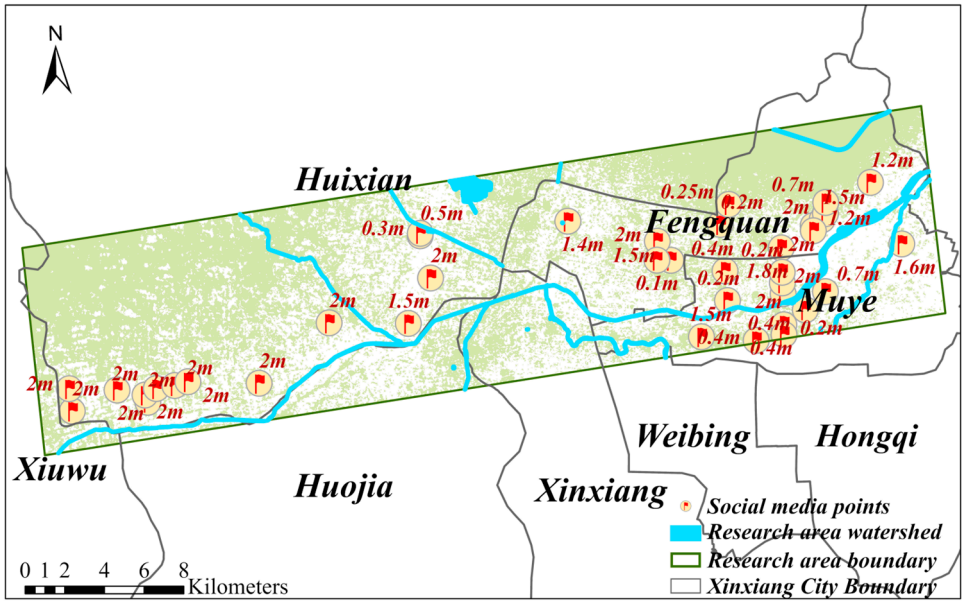


Figure 7. Extraction results and spatial distribution of SMPs.

3.2. Probability Estimation Using SMP

The first step is to generate the inundation potential distribution based on the SMP locations. With each SMP location as the center, based on their water depth and the elevation data of the study area, the potential distribution of the flood inundation in the whole study area can be estimated. In this study, a total of 39 SMP in the study area were selected, and flood inundation PID maps for the same study area based on 27 points were finally generated.

As shown in Figure 8, the inundation probability result maps of the four representative SMPs distributed along the Weihe River in the study area were selected. The change in the raster color represents the change in the probability of flood inundation from low to high and the flood risk as well. It was found that the inundation probability is mainly centered on the point from high to low change, which is related to the inverse distance weighting. The distribution of inundation probability is closely related to the distribution and change in elevation. Around the SMP, the higher the elevation of the region, the lower the probability of being inundated. For example, the northwestern part of the study area is the Taishan Mountain, with higher elevation, so this region has a lower probability of being inundated.

3.3. Weighted Probability by Incorporating Physical Sensing Data and SMP

Based on the layer calculated by each SMP, a Gaussian function is introduced to calculate the weights of each of the seven feature information layers within the radius of interest. Finally, the weights of all the SMP layers are superimposed to obtain the probability layer of all feature information. Figure 9 shows the distribution of pixel values in each weight layer, that is, the probability distribution. Combined with Equation (8), it can be seen that when the PID is the same, the greater the weight of the feature information, and the greater the final probability value. It is concluded that the more the curve moves to the upper right, the greater the weight of the feature information is. It can be concluded that:

(1). Regarding the probability of all feature information, the trend of the number of pixels on the curve is basically the same, which is related to the weight allocation calculation formula (Gaussian function).

(2). The probability–image curve of the rainfall factor is shifted to the right, which indicates that the weight of rainfall is higher than that of the other factors, confirming that the rainfall factor was indeed the most important causal factor in this flood event, and that the method presented assigns the largest weight value to it.

(3). The flow power index also shows a rightward shift, indicating that, according to the Weibo site, it also contributes to the generation of flooding, which is related to the flooding triggers (river flooding) that occur.

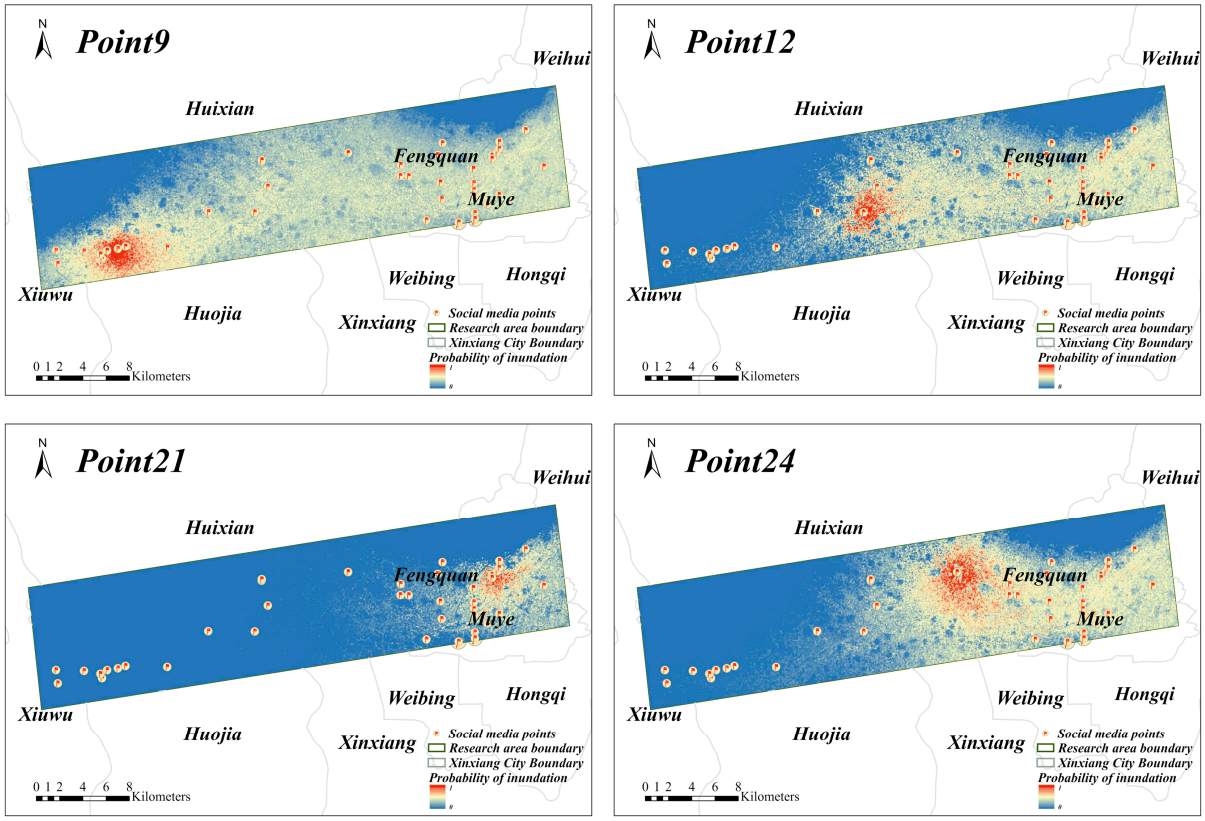


Figure 8. Index layers of the probability distribution of representative points along the Wei River in the study area.

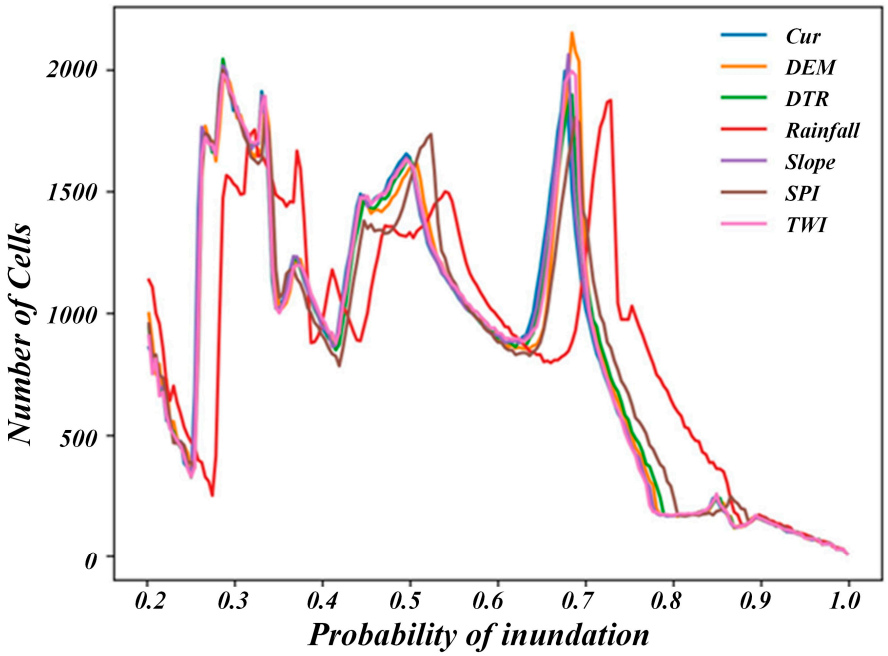


Figure 9. The pixel probability and pixel number distribution of all feature information weighted layers.

The weights were applied to the IDW-PID layer of this SMP, and finally, the flood inundation PID prediction map integrating the flooding SMPs and the physical sensing data within the study area was synthesized, as shown in Figure 10. The following discoveries were made: The distribution of the flood inundation probability is roughly the same as that of the SMPs. In addition, it is closely related to the topographic trend of the DEM within the study area. The closer the pixel is to the point and the lower the elevation is, the higher the probability of flood inundation, that is, the greater the accessibility of the flood. On the contrary, the further the pixel is from the point and the higher the elevation is, the lower the probability of flood inundation.

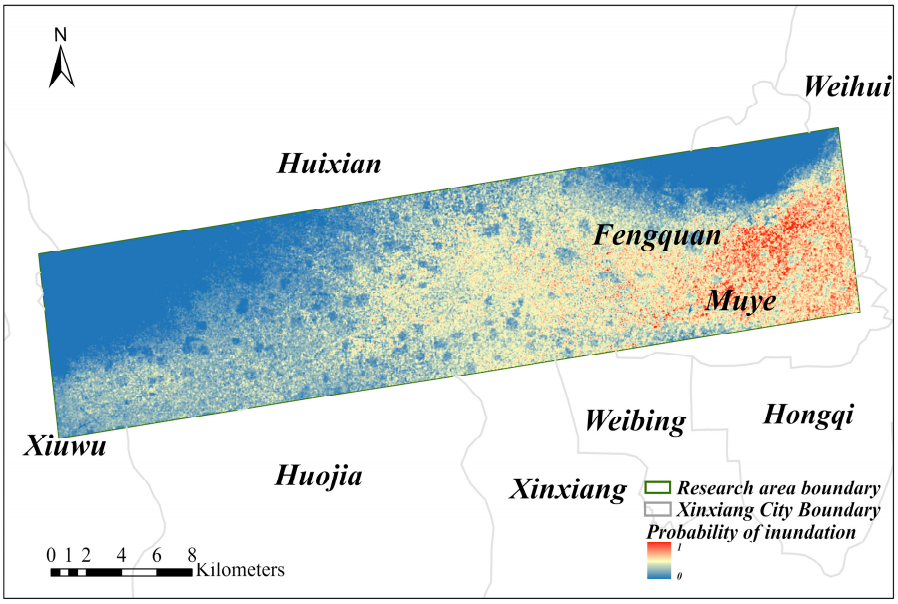


Figure 10. The final flood probability distribution map.

4. Discussion

4.1. Comparison

4.1.1. Comparison with Physical Sensing Method

Existing studies often rely solely on physical sensing data without incorporating social media. For example, an integrated method based on particle swarm optimization and weakly labeled support vector machine has been used to assess urban flooding susceptibility [6–9]. The main difference between the two approaches is that the IPS approach uniquely incorporates real-time data from social media, which provides near real-time flood information and improves the accuracy of flood prediction. Existing models typically assess the susceptibility of spatial areas and are not event-specific. By focusing on specific events, the IPS method facilitates immediate decision-making and emergency response, whereas existing methods are more suitable for general susceptibility analysis and long-term urban planning.

4.1.2. Comparison with Radar Image

It is always difficult to grasp the scientific scale to calculate the probability through the category of events, and to verify the probability events, it is necessary to test them through actual observation and statistical data. So, the obtained flood probability distribution map was validated with the water bodies extracted from the post-disaster Gaofen-3 radar images. Gaofen-3 is China's first C-band multi-polarization synthetic aperture radar (SAR) satellite. It is capable of 12 imaging modes, covering the traditional Strip map and Scan SAR modes, as well as Spotlight and Global Observation modes. The image used in this validation came from GF-3 satellite data (FSII), with a spatial resolution of 10 m. The observation time was

22 July 2021, which fell within the disaster period, and the HV polarization method, which is more sensitive to the water body, was chosen.

In this paper, SARscape radar image plug-in in ENVI 5.6 was used to process the GF-3 satellite L1 A-level single-look complex (SLC) data. The preprocessing process included importing SAR data, multi-look processing, filtering processing, geocoding, and radiometric calibration. Finally, radar backscattering coefficient data with a 10 m spatial resolution were obtained. The backscattering coefficient in SAR images is mainly affected by the roughness of the surface of the ground object and the system parameters of the radar. The surfaces of water bodies are smooth and mirror-scattering, and their backscattering coefficient is low. The surface roughness of non-water bodies, such as vegetation and towns, is diffused reflection, and their backscattering coefficient is high. The threshold segmentation method uses the principle of the low backscattering coefficient of water to divide the pixels lower than a certain threshold in the SAR image into water bodies, with the remaining pixels determined to be non-water bodies. However, due to the spatial and temporal resolution of the radar image, a comparison of the whole map was not possible, and the parts that could be verified were selected as much as possible for visual and quantitative analysis.

GF-3 image and threshold segmentation method were used to extract the flood range of the day. We used this range to clip the final flood probability distribution map and obtained the probability estimation results only in the flood range. This was helpful for the subsequent quantitative statistical verification of the range to evaluate the effectiveness of the method. As shown in Figure 11, the submerged area of the water body is still relatively large. According to the colors of the submerged areas shown in the legend, it was found that most of the probability estimates in this area were relatively large. The methodology of this study shows some applicability in most of the areas, and the corresponding degree of inundation risk is also reflected by the different colors.

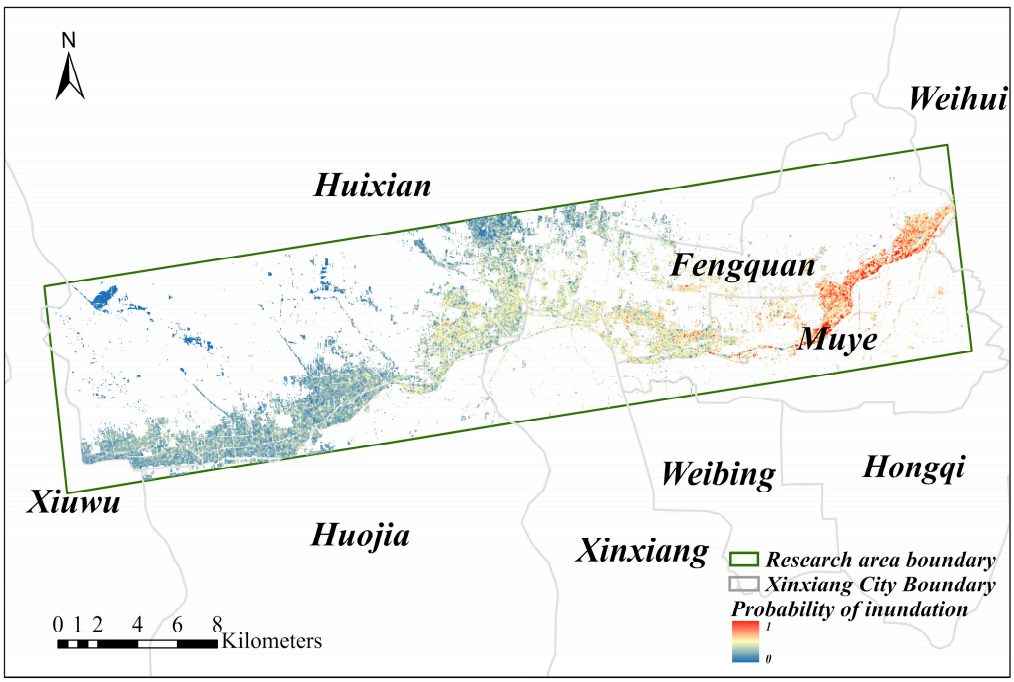


Figure 11. Map of the results for the areas of the extracted water body.

4.1.3. Comparison with Social Media Data

Combined with Figure 12, it can be seen that most of the points are located in areas with high probability, which may be related to the distribution along the river. The terrain is low and vulnerable to river intrusion. If the flooding river is not dredged as soon as possible, it will affect the area closest to the river. A probability value between 0 and 1 can

be regarded as the level of flood risk. The greater the value, the greater the risk. According to the different colors shown in the area in the figure, the different risk levels of floods can be seen.

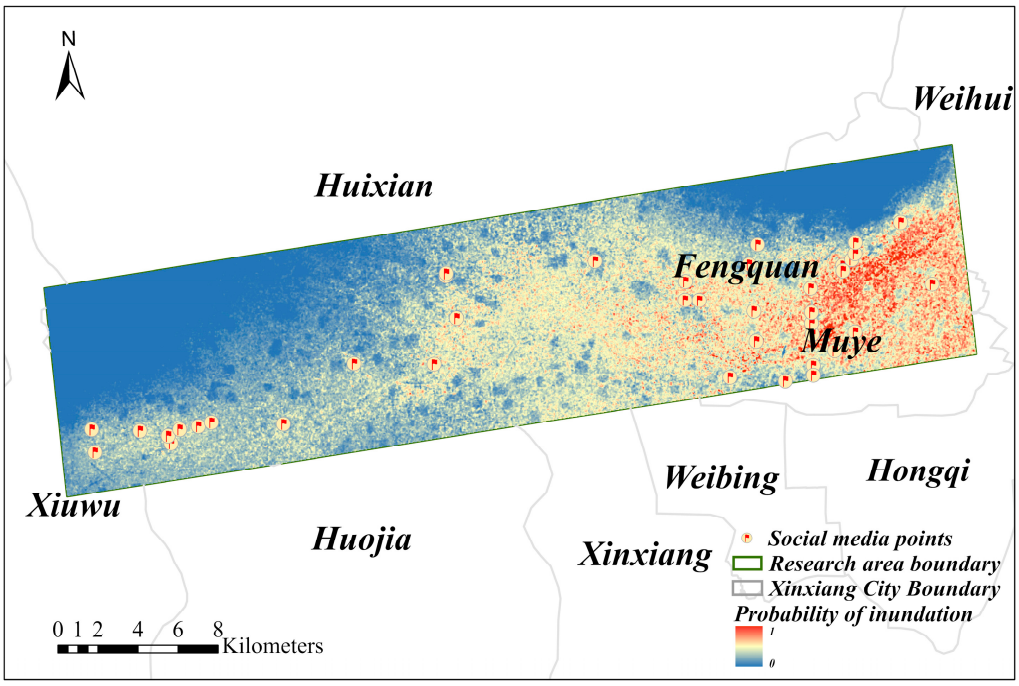


Figure 12. Validation using points extracted from social media data.

4.2. Quantitative Evaluation

When the radius of interest had a maximum value of 420 m, the water depth was judged to be 2 m. Within the radius, an estimated probability value greater than 0.098 was judged to be submerged. Both a greater depth of water and a smaller distance make the probability greater than 0.098.

Therefore, in this paper, 0.098 was chosen as the threshold for judging whether inundation occurred or not, and the size of the value represents the size of the inundation degree. By numerically counting the results of the above validation, the probability of all the pixels were obtained in the region of the water body, as shown in the Table 3.

Table 3. Predicted probability values for water body regions—cell count statistics.

Range	No. of Pixels	Percentage of Pixels
0–0.2	54,958	54.5559%
0.2–0.4	38,967	38.6819%
0.4–0.6	4834	4.7986%
0.6–0.8	366	0.3633%
0.8–1	14	0.0139%

The category determination of the pixels using threshold segmentation revealed that 88.77% were identified as submerged. The probability distribution of the submerged pixels is shown in Table 4.

In the validation, using 30% of the total dataset, it was found that 75% of points were located in pixels that were recognized as submerged at this threshold; their probability distribution is shown in Table 5.

Table 4. Statistics of the number of cell numbers with predicted probability values greater than 0.098 in water body regions.

Range	Percentage of Pixels
0.098–0.278	70.3773%
0.278–0.459	15.8949%
0.459–0.639	2.312%
0.639–0.82	1.787%
0.82–1	0.0079%

Table 5. Estimation probabilities and judgment categories of validation dataset.

ID	Estimation Probability	Category (Flooded (F)/Unflooded (UF))
1	0.034669	UF
2	0.04841	UF
3	0.168542	F
4	0.122249	F
5	0.170505	F
6	0.341228	F
7	0.246976	F
8	0.380073	F
9	0.093704	UF
10	0.22975	F
11	0.181383	F
12	0.194581	F

Overall, the method presented in this paper shows good results in most of the study area, in addition to the predicted result maps that provide a continuous probabilistic view to the whole map. Compared to such post-disaster water extraction assessment methods, the proposed method can reduce the utilization of resources while achieving updatability and spatial continuity.

4.3. Result Interpretation

During the specified calculation period, i.e., 22–26 July 2021, the areas of Fengquan and Muye were identified as at high flooding risk. Figure 10 illustrates that these areas are significantly at risk, primarily due to their lower elevation and higher urbanization rate. High flooding probability areas are mainly clustered along the river, indicating that river areas are more susceptible to flooding. This finding is consistent with what is indicated in official reports about the distribution of flooded areas. Figure 9 shows that the probability–image curve for the rainfall factor is shifted to the right, indicating its higher weight. This confirms that rainfall was the main driver of the flood event. Additionally, the flow power index (SPI) also shows a rightward shift, indicating its influence on flooding. The combination of the high rainfall and the insufficient flood storage capacity of the river resulted in severe flooding.

Given the high flood risk in Fengquan and Muye, it is crucial to enhance monitoring and response strategies in these areas. Authorities should prioritize real-time tracking of rainfall and river water levels to anticipate and mitigate further developments of flooding. Implementing advanced flood prediction models that integrate both physical and social sensing data can improve the accuracy and timeliness of flood warnings. Additionally, reinforcing infrastructure and drainage systems in vulnerable areas and conducting regular risk assessments can significantly reduce potential damages. Public awareness campaigns should also be launched to educate residents about emergency procedures and safety measures during flood events.

4.4. Advantages and Limitations

The IPS method proposed in this paper can extract submergence depth and flooded location information from social sensing data and combines the socially sensed flooded information with physical factors to achieve space-continuous flood estimation. Due to the high data availability and large data volume of social sensing data, IPS could achieve low-cost and rapid flood inundation probability estimation for large areas.

The overall timeliness of IPS is about 24 h–48 h after the occurrence of floods. The time lag originates from the four aspects of social sensing data collection, submerged location and depth extraction, social sensed data verification by local news and reports, and parameter calculation. The time lag of the IPS method is mainly caused by the socially sensed flooded information verification by local news and reports, usually taking half or a whole day. Compared to radar and available optical remote sensing, IPS enables flood inundation estimation in a relatively timely manner. In comparison to ground-based monitoring, the IPS method proposed in this paper could achieve space-continuous flood inundation estimation and facilitate valuable decision-making insights for large-scale and long-term (at least 3–5 days) severe flood events caused by continuous rainfall, such as the "7·20" Henan rainstorm in 2021 and the "7·31" Beijing–Tianjin–Hebei heavy rainstorm in 2023.

IPS also has its limitations. First, it has three higher requirements for the socially sensed urban flooding data, including the data quantity, quality, and spatial distribution [25]. In terms of data quantity, there should be at least one social media inundation point within every 400 km^2 in order to ensure reliable estimation results. With respect to quality, the water depth and location information of the social media points used needs to be carefully checked and verified. For spatial distribution, the social media points used need to be spatially distributed as evenly as possible, with better coverage of the locations with higher inundation. Second, elevation is significant for IPS, and better results can be obtained if elevation data with higher accuracy is used. Finally, due to the high timeliness requirements of flood events, this paper may not be significant for waterlogging events that stop within a few hours.

4.5. Possible Improvements

Although IPS has its limitations, this approach offers innovative methods for flood probability estimation compared to traditional physical methods. Future improvements can focus on the following: (1) With the popularity of social media platforms and the rapid development of mobile internet, more flood-related social media data can be acquired. (2) More advanced textual information extraction methods can be used to more accurately extract flood social media data. This will enhance the reliability and timeliness of the data used for flood probability estimation. (3) Higher resolution DEM data can be considered to improve the accuracy of probability estimation. Higher resolution DEM can capture more detailed topographic changes, which are critical for floodplain mapping. (4) Incorporating additional data sources, such as weather forecasts and historical flood records, can enhance the robustness of our flood probability estimation. By integrating these diverse datasets, more accurate estimation of flood risks can be achieved.

5. Conclusions and Outlook

This paper proposes a method that integrates historical geographic information base data with real-time social media data. By employing the inverse distance function and Gaussian function, corresponding weights were assigned to different data sources to estimate the probability of flooding in areas experiencing heavy rainfall. Additionally, a series of multi-source data feature extraction techniques were introduced, particularly for social media data. This approach allows us to analyze the advantages and disadvantages of social and physical sensing, enabling them to complement each other as information sources for disaster assessment. Using the 2021 heavy rainfall event in Henan Province as a case study, the feasibility of the IPS method was demonstrated by verifying the

inundation probability estimation maps against the flood range of GF-3 radar imagery and validation sets.

In particular, the inundation potential across the entire map was generalized by applying the distance attenuation effect in the dimensions of the point water depth. Furthermore, the Gaussian function was utilized to capture physically sensed geographic information data and assign corresponding weights to enrich the detailed probability values of the entire map. Ultimately, an inundation probability distribution map was obtained for the overall area. Overall, the IPS method offers new insights into the estimation of flooding probability, particularly in scenarios lacking real-time updates and the availability of remote sensing images. Future research will focus on integrating advanced NLP and machine learning methods for social media data, along with utilizing higher resolution DEM data to significantly enhance the accuracy and reliability of flood probability estimation.

Author Contributions: Conceptualization, W.D. and Q.X.; data curation, Q.X.; formal analysis, Q.X.; funding acquisition, W.D.; investigation, Q.X.; methodology, W.D. and Q.X.; resources, W.D.; supervision, W.D.; validation, Q.X. and L.X.; visualization, Q.X.; writing—original draft, Q.X.; writing—review and editing, W.D., Q.X., B.C., L.X., Z.C., X.Z., M.H. and N.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China under Grant No. 2021YFF0704400; the Natural Science Foundation of Hubei Province under Grant No. 2023AFB107; the National Natural Science Foundation of China Program under Grant Nos. 41971351 and 42201438; the Special Fund of Hubei Luojia Laboratory under Grant No. 220100034; and the Open Fund of the National Engineering Research Centre for Geographic Information System under Grant Nos. 2021KFJJ06 and 2022KFJJ02.

Data Availability Statement: The data and code are accessible at the following link: https://github.com/WUHANQINGY/IPS-Flooded (accessed on 28 March 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Pachauri, R.K.; Mayer, L.; Intergovernmental Panel on Climate Change (Eds.) *Climate Change 2014: Synthesis Report*; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2015; ISBN 978-92-9169-143-2.
2. Global Natural Disaster Assessment Report. 2022. Available online: https://www.preventionweb.net/publication/2022-global-natural-disaster-assessment-report (accessed on 12 October 2023).
3. Liu, J.; Tao, L.; Yang, Y. Dynamical Analysis of Multi-Scale Interaction during the "21·7" Persistent Rainstorm in Henan. *Atmos. Res.* **2023**, *292*, 106857. [CrossRef]
4. Wu, P.; Clark, R.; Furtado, K.; Xiao, C.; Wang, Q.; Sun, R. A Case Study of the July 2021 Henan Extreme Rainfall Event: From Weather Forecast to Climate Risks. *Weather Clim. Extrem.* **2023**, *40*, 100571. [CrossRef]
5. Wu, Y.; Jiang, N.; Xu, Y.; Yeh, T.-K.; Guo, A.; Xu, T.; Li, S.; Gao, Z. Revealing the Water Vapor Transport during the Henan "7.20" Heavy Rainstorm Based on ERA5 and Real-Time GNSS. *Egypt. J. Remote Sens. Space Sci.* **2024**, *27*, 165–177. [CrossRef]
6. Hong, H.; Panahi, M.; Shirzadi, A.; Ma, T.; Liu, J.; Zhu, A.-X.; Chen, W.; Kougias, I.; Kazakis, N. Flood Susceptibility Assessment in Hengfeng Area Coupling Adaptive Neuro-Fuzzy Inference System with Genetic Algorithm and Differential Evolution. *Sci. Total Environ.* **2018**, *621*, 1124–1141. [CrossRef]
7. Abebe, Y.; Kabir, G.; Tesfamariam, S. Assessing Urban Areas Vulnerability to Pluvial Flooding Using GIS Applications and Bayesian Belief Network Model. *J. Clean. Prod.* **2018**, *174*, 1629–1641. [CrossRef]
8. Arabameri, A.; Rezaei, K.; Cerdà, A.; Conoscenti, C.; Kalantari, Z. A Comparison of Statistical Methods and Multi-Criteria Decision Making to Map Flood Hazard Susceptibility in Northern Iran. *Sci. Total Environ.* **2019**, *660*, 443–458. [CrossRef] [PubMed]
9. 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]
10. Tansar, H.; Babur, M.; Karnchanapaiboon, S. Flood Inundation Modeling and Hazard Assessment in Lower Ping River Basin Using MIKE FLOOD. *Arab. J. Geosci.* **2020**, *13*, 934. [CrossRef]
11. Ntanganedzeni, B.; Nobert, J. Flood Risk Assessment in Luvuvhu River, Limpopo Province, South Africa. *Phys. Chem. Earth Parts A/B/C* **2021**, *124*, 102959. [CrossRef]
12. Yu, W.; Jiang, C.; Liu, J.; Zhou, Q. Hydrologic-hydrodynamic model and its application in flood risk analysis. *J. Hydroelectr. Eng.* **2019**, *38*, 87–97.
13. Jaya, A.; Sari, F.; Saragih, I.; Dafitra, I. Sea-Level Prediction for Early Warning Information of Coastal Inundation in Belawan Coastal Area Using Delft3D Model. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *893*, 012034. [CrossRef]

14. Wang, Y.; Liu, W.; Li, X.; Xu, J. Research Progress of Rainfall-runoff Simulation Based on Land-atmosphere Coupling Model. *J. Chang. River Sci. Res. Inst.* **2024**, *41*, 26–35.
15. Huang, X.; Wang, C.; Li, Z. A near Real-Time Flood-Mapping Approach by Integrating Social Media and Post-Event Satellite Imagery. *Ann. GIS* **2018**, *24*, 113–123. [CrossRef]
16. Du, W.; Gong, Y.; Chen, N. PSO-WELLSVM: An Integrated Method and Its Application in Urban Waterlogging Susceptibility Assessment in the Central Wuhan, China. *Comput. Geosci.* **2022**, *161*, 105079. [CrossRef]
17. Brocca, L.; Ciabatta, L.; Massari, C.; Moramarco, T.; Hahn, S.; Hasenauer, S.; Kidd, R.; Dorigo, W.; Wagner, W.; Levizzani, V. Soil as a Natural Rain Gauge: Estimating Global Rainfall from Satellite Soil Moisture Data. *J. Geophys. Res. Atmos.* **2014**, *119*, 5128–5141. [CrossRef]
18. Souto, J.; Beltrão, N.; Teodoro, A. Performance of Remotely Sensed Soil Moisture for Temporal and Spatial Analysis of Rainfall over São Francisco River Basin, Brazil. *Geosciences* **2019**, *9*, 144. [CrossRef]
19. Su, L.; Li, Z.; Gao, F.; Yu, M. A review of remote sensing image water extraction. *Remote Sens. Nat. Resour.* **2021**, *33*, 9–19.
20. Giustarini, L.; Hostache, R.; Matgen, P.; Schumann, G.J.-P.; Bates, P.D.; Mason, D.C. A Change Detection Approach to Flood Mapping in Urban Areas Using TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2417–2430. [CrossRef]
21. Martinis, S.; Rieke, C. Backscatter Analysis Using Multi-Temporal and Multi-Frequency SAR Data in the Context of Flood Mapping at River Saale, Germany. *Remote Sens.* **2015**, *7*, 7732–7752. [CrossRef]
22. Martinis, S.; Twele, A.; Voigt, S. Towards Operational near Real-Time Flood Detection Using a Split-Based Automatic Thresholding Procedure on High Resolution TerraSAR-X Data. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 303–314. [CrossRef]
23. Goodchild, M.F.; Glennon, J.A. Crowdsourcing Geographic Information for Disaster Response: A Research Frontier. *Int. J. Digit. Earth* **2010**, *3*, 231–241. [CrossRef]
24. Qi, L.; Li, J.; Wang, Y.; Gao, X. Urban Observation: Integration of Remote Sensing and Social Media Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4252–4264. [CrossRef]
25. Fang, L.; Zhang, Z.; Huang, J. Rapid Flood Modelling Using HAND-FFA-SRC Coupled Approach and Social Media-Based Geodata in a Coastal Chinese Watershed. *Environ. Model. Softw.* **2023**, *170*, 105862. [CrossRef]
26. Wang, Z.; Lam, N.S.N.; Sun, M.; Huang, X.; Shang, J.; Zou, L.; Wu, Y.; Mihunov, V.V. A Machine Learning Approach for Detecting Rescue Requests from Social Media. *Int. J. Geo-Inf.* **2022**, *11*, 570. [CrossRef]
27. Resch, B.; Usländer, F.; Havas, C. Combining Machine-Learning Topic Models and Spatiotemporal Analysis of Social Media Data for Disaster Footprint and Damage Assessment. *Cartogr. Geogr. Inf. Sci.* **2018**, *45*, 362–376. [CrossRef]
28. Rosser, J.F.; Leibovici, D.G.; Jackson, M.J. Rapid Flood Inundation Mapping Using Social Media, Remote Sensing and Topographic Data. *Nat. Hazards* **2017**, *87*, 103–120. [CrossRef]
29. Xu, L.; Ma, A. Coarse-to-Fine Waterlogging Probability Assessment Based on Remote Sensing Image and Social Media Data. *Geo-Spat. Inf. Sci.* **2021**, *24*, 279–301. [CrossRef]
30. Panteras, G.; Cervone, G. Enhancing the Temporal Resolution of Satellite-Based Flood Extent Generation Using Crowdsourced Data for Disaster Monitoring. *Int. J. Remote Sens.* **2018**, *39*, 1459–1474. [CrossRef]
31. Songchon, C.; Wright, G.; Beevers, L. Quality Assessment of Crowdsourced Social Media Data for Urban Flood Management. *Comput. Environ. Urban Syst.* **2021**, *90*, 101690. [CrossRef]
32. Li, H.; Zech, J.; Ludwig, C.; Fendrich, S.; Shapiro, A.; Schultz, M.; Zipf, A. Automatic Mapping of National Surface Water with OpenStreetMap and Sentinel-2 MSI Data Using Deep Learning. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102571. [CrossRef]
33. Cao, C.; Xu, P.; Wang, Y.; Chen, J.; Zheng, L.; Niu, C. Flash Flood Hazard Susceptibility Mapping Using Frequency Ratio and Statistical Index Methods in Coalmine Subsidence Areas. *Sustainability* **2016**, *8*, 948. [CrossRef]
34. Tang, X.; Shu, Y.; Lian, Y.; Zhao, Y.; Fu, Y. A Spatial Assessment of Urban Waterlogging Risk Based on a Weighted Naïve Bayes Classifier. *Sci. Total Environ.* **2018**, *630*, 264–274. [CrossRef]
35. Wang, J.; Li, X.; Christakos, G.; Liao, Y.; Zhang, T.; Gu, X.; Zheng, X. Geographical Detectors-Based Health Risk Assessment and Its Application in the Neural Tube Defects Study of the Heshun Region, China. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 107–127. [CrossRef]
36. Chaudhary, P.; D'Aronco, S.; Leitao, J.P.; Schindler, K.; Wegner, J.D. Water Level Prediction from Social Media Images with a Multi-Task Ranking Approach. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 252–262. [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.