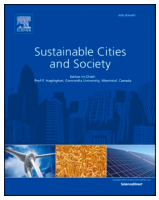
Sustainable Cities and Society 88 (2023) 104307

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs



# Urban flood susceptibility mapping based on social media data in Chengdu city, China

Yao Li [a,*], Frank Badu Osei [a], Tangao Hu [b,c], Alfred Stein [a]

[a] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, the Netherlands
[b] Institute of Remote Sensing and Earth Sciences, Hangzhou Normal University, Yuhangtang Road No. 2318, Hangzhou 311121, China
[c] Zhejiang Provincial Key Laboratory of Urban Wetlands and Regional Change, Hangzhou Normal University, Yuhangtang Road No. 2318, Hangzhou 311121, China

ARTICLE INFO

*Keywords:*
Urban flood susceptibility mapping
Social media data
Naïve Bayes
Standard deviation ellipse
Chengdu city

ABSTRACT

Increase in urban flood hazards has become a major threat to cities, causing considerable losses of life and in the economy. To improve pre-disaster strategies and to mitigate potential losses, it is important to make urban flood susceptibility assessments and to carry out spatiotemporal analyses. In this study, we used standard deviation ellipse (SDE) to analyze the spatial pattern of urban floods and find the area of interest (AOI) based upon related social media data that were collected in Chengdu city, China. We used the social media data as the response variable and selected 10 urban flood-influencing factors as independent variables. We estimated the susceptibility model using the Naïve Bayes (NB) method. The results show that the urban flood events are concentrated in the northeast-central part of Chengdu city, especially around the city center. Results of the susceptibility model were checked by the Receiver Operating Characteristic (ROC) curve, showing that the area under the curve (AUC) was equal to 0.8299. This validation result confirmed that the susceptibility model can predict urban flood with a satisfactory accuracy. The urban flood susceptibility map in the city center area provides a realistic reference for flood monitoring and early warning.

## 1. Introduction

More than 30% of the economic losses and victims from annual natural disasters are flood related (Rodriguez et al., 2021). Flooding is therefore among the most devastating hazards on Earth, posing great threats to a large amount of population in the world (Chang et al., 2021; Zhang et al., 2021, 2022). With its rapid urbanization and development, urban flood has become more and more pervasive and severe in China (Zheng et al., 2016). More than 100 cities have suffered from urban flooding since 2006, and urban flooding has thus become one of the most severe urban problems in China (Liu et al., 2020). Considering the negative influences of urban flooding, it would be highly desirable to better understand the mechanisms of urban flooding and susceptibility mapping for better management.

Consequences of urban flooding include social and economic losses, property damage, loss of life, and include other indirect impacts like health risk. These all highlight the urgent need for a monitoring program (Rosenzweig et al., 2021; Yang et al., 2022). Traditional flood monitoring includes meteorological station/infrastructure observation, manual field survey, and remote sensing observation. Among these, infrastructure observation and manual field surveys have a limited capacity to cover fine-scale areas, because most cities lack the formal infrastructure for effective urban flood monitoring (Helmrich et al., 2021). Moreover, limited cost, manpower, and physical access affect manual field surveys (Li et al., 2019). Nowadays, remote sensing plays an important role in flood monitoring (Bai et al., 2021). Previous researches have used remote sensing images for monitoring flood events, like optical satellite images or specifically synthetic aperture radar (SAR) images (Sun et al., 2016). However, given the satellite revisit limitations, remote sensing images may be unavailable because of cloud cover of optical remote sensing images and the distortion effects of SAR data (Balz et al., 2015).

In recent years, social media such as Weibo and Twitter have emerged as new data sources in natural disaster area. These can offer a big data volume at a high spatiotemporal resolution and with a wide coverage area (Wang et al., 2019). For instance, Zou et al. (2018) used Twitter as a data source to explore the emergency management during the hurricane Sandy event and found that social media data could help improve post disaster damage estimation. Yao et al. (2021) took the Changning earthquake as a case study and proposed a method to assess

* Corresponding author at: ITC, University of Twente, Enschede, the Netherlands.
*E-mail address:* yao.li@utwente.nl (Y. Li).

https://doi.org/10.1016/j.scs.2022.104307
Received 15 July 2022; Received in revised form 15 November 2022; Accepted 15 November 2022
Available online 17 November 2022

the seismic intensity information based upon Weibo data. Li et al. (2021) demonstrated the efficacy of social media data when investigating aid evacuations during the wildfire season in the western United States. Although social media have been used in different kinds of natural disaster management, there are limited applications of urban flood susceptibility modelling by social media data. Regarding flood and social media related research, Songchon et al. (2021) has assessed the effective quality of social media data acquired from Twitter during flood events in Thailand.

Data shortage and complex urban mechanisms are two major barriers in urban flood susceptibility modelling and mapping research (Zhao et al., 2018). The deficient hydro-meteorological and hydraulic characteristics data make it hard to determine the most important urban flood mechanisms. For example, it is difficult to obtain spatiotemporal rainfall data of a high resolution and detailed sewer drainage data. Complex urban mechanisms are characterized by high-density buildings, winding waterways and culverts, and increasing land fragmentation (Xiang et al., 2021). Several hydrological and hydrodynamic models have been applied with varying degrees of success in urban flooding area, such as SWMM (X. Wu et al., 2017), Mike (Hénonin et al., 2015), SUSIM (Li et al., 2019) and Info Works ICM (Cheng et al., 2017). Time-consuming modeling and complicated parameter estimation, however, are restricting their application (Zhao et al., 2019).

Considering the above difficulties, machine learning models, which have shown advantages in modeling complex systems (Casali et al., 2022), may be useful for urban flood susceptibility modelling with limited data (Zhao et al., 2018). Based upon statistical methods and computer science, machine learning models aim to develop and train mathematical models to find the relationship between explanatory factors and observed flood samples. These models can then be used to urban flood susceptibility modelling or support decision making (Ke et al., 2020). Naïve Bayes (NB) models, comprising classical machine learning algorithms, apply Bayes' theorem with predictor independence assumptions. NB models can perform surprisingly well with effectively handling complex and incomplete data when making large-scale predictions (Tang et al., 2020). In the past, NB models have shown acceptable performances in landslide susceptibility mapping, wildfire susceptibility mapping (Pham et al., 2017) and flood susceptibility. For instance, Khosravi et al. (2019) used a NB model to assess the flood susceptibility in the Ningdu Watershed, one of China's most flood-prone areas, but there has been little use of NB in predicting urban flood.

The main objective of this study is to use social media data to map the urban flood prone area based on NB modeling. To achieve the primary purpose, this study has three sub-objectives: (1) to analyst the dynamic spatial characteristics changes of the urban flood events, and find the area of interest (AOI) based on standard deviation ellipse (SDE); (2) to select the related urban flood influencing factors and analyze the contribution of factors; and (3) to develop an NB model and train this model to map urban flood spatial susceptibility. The modeling will be carried out in Chengdu city, China.

## 2. Study area and data preprocessing

### 2.1. Study area

Chengdu city (30°39′ N, 104°04′ E), the capital city of Sichuan Province, China, covers about 14,335 km$^2$ (Fig. 1a). According to the administrative division standard of 2018 from the Chengdu Civil Affairs Bureau, there are 11 municipal districts, 4 counties and 5 county-level cities in Chengdu city. As the capital of Sichuan Province, Chengdu

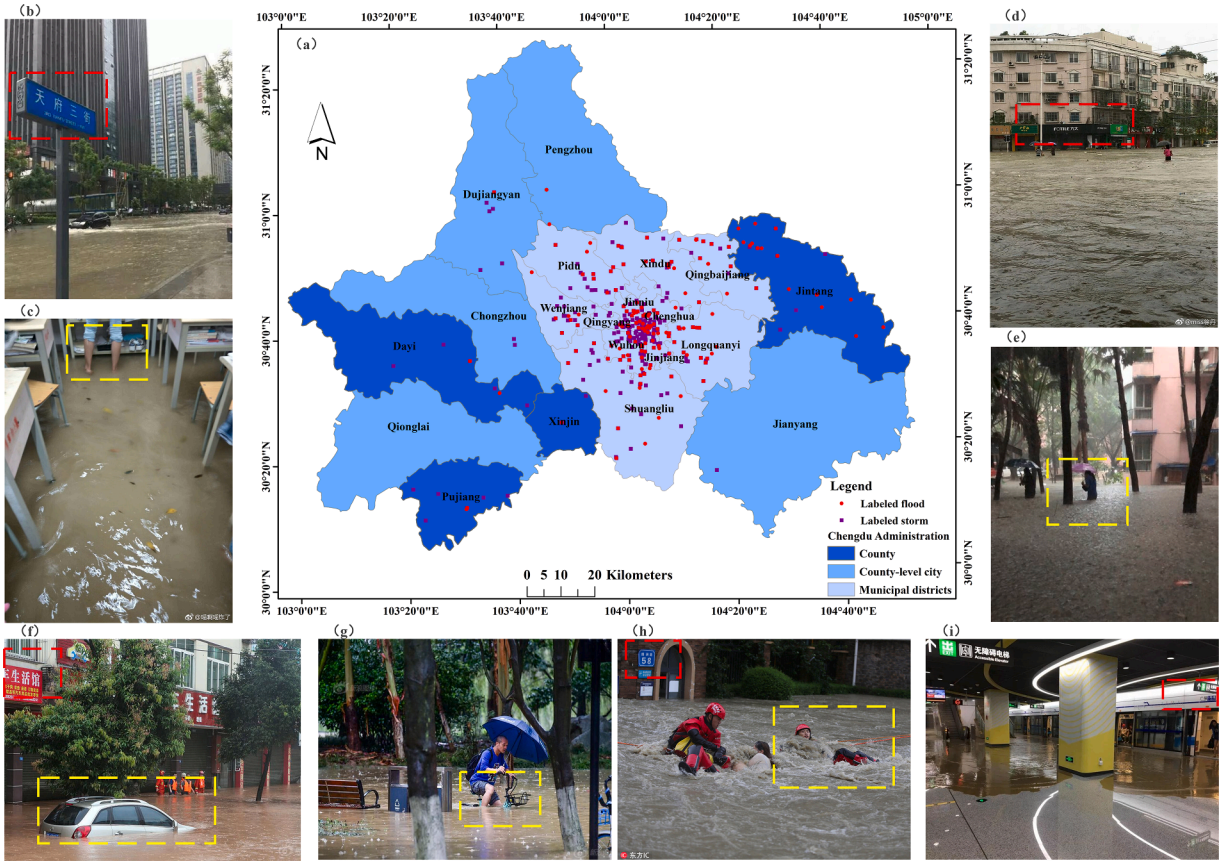



**Fig. 1.** Location of study area: (a) the administrative division of study area with the social media data; (b-e) the urban flood pictures from social media blogs, and (f-i) the urban flood pictures from mainstream media. The red boxes in the pictures can show the location information and the yellow boxes can reflect the severity of the flood.

city is the most important economic center in southwest China. Frequent urban flood disasters, however, have caused large economic losses and have seriously threatened the safety of people's lives and property. Affected by heavy rainfalls in July 2018, Sichuan Province suffered from severe flooding that was not seen in 50 years. Many rivers had over-alerted and over-protected water levels, causing heavy losses in many places. Chengdu city, one of the most severely affected cities, experienced 5 floods in the city on July 2, 9-11, 14-16, 15-17, and 19-21, all severely affecting citizens' lives.

### 2.2. Data sources

This study requires the use of urban flooding records, topography, precipitation, remote sensing images, and other surface information. Flooding records were used to construct flood inventory maps, while the remaining datasets were utilized as the influencing factors of urban flooding. Detailed information of the datasets and their sources is shown in Table 1. All data are available at http://www.sciencedb.cn/dataSet/handle/712 (Li et al., 2018).

### 2.3. Social media data cleaning and mining

The historical urban flood records during July, 2018 were derived from a published social media dataset (Li et al., 2018). The open Application Programming Interface (API) of the Sina Weibo platform (https://open.weibo.com/wiki/API) provides an opportunity to capture blog data (Deng et al., 2016). The original dataset included 1432 blogs which were captured from the Weibo API by the searching strategy:

[location = Chengdu; keywords = {'flood disaster', 'flood', 'inundation', 'overflow', 'submerge', 'heavy storm'}; time = 01-07-2018 to 31-07-2018].

Every record had the attributes: post location, blog content, post time, and forwarding times. During data pre-processing, we took five steps to control the data quality to obtain two datasets: First, 42 duplicate records were removed, resulting in dataset-1 which was used for precipitation-heavy storm time analysis. Second, retweeted blogs were removed by selecting the blogs from people that used Weibo to express their own conditions during the flood period, rather than to share content generated by others (Kankanamge et al., 2020). Third, blog posts were selected with detailed information on the geographical location of the street, as several records in dataset-1 provided too coarse location information in city/district level. We thus removed 318 records with locations that were at the city/district level only. Fourth, we used the news and reports from different sources like the mainstream media such as Chinanews.com, sina.com, sohu.com, scnews.newssc.org, tianqi.com, and official government website, e.g., gk.chengdu.gov.cn to cross-validate the flooding locations to assure the reliability of the data. Fifth, we selected blogs with photos and videos documenting the flood to cross-validate the flooding locations. The geotagged photos and videos posted in the blogs usually provided location specific, more trustworthy information than the text messages (Fig. 1b-i). This resulted into 325 ultimately recognized flood sites as dataset-2. Their distribution is shown in. Fig. 1a. For the subsequent analysis, also 325 non-flooded sites were chosen according to the field survey and local news, which were located at high elevations and in high drainage capacity areas (Hosseini et al., 2020; Zhao et al., 2019). Moreover, the first author knows the city personally very well, as she was grown up there, and she used this rather subjective knowledge to help selecting the non-flooded sites.

**Table 1**
Datasets used in this study.

| Data and format | Format | Time | Source |
|---|---|---|---|
| Flood records | Shapefile-Point | July, 2018 | Social Media data from Weibo |
| Administrative divisions | Shapefile-Polygon | 2018 | Chengdu Civil Affairs Bureau |
| Digital Elevation Model (DEM) | Raster-30 m | - | The Geospatial Data Cloud |
| Precipitation data | Text- District | July, 2018 | Chengdu Meteorological Bureau |
| River | Shapefile-Polygon | 2018 | National Catalogue service For Geographic Information |
| Sentinel 2 images | Raster-10 m | 2018 | Google Earth Engine Platform |

### 2.4. Urban flood influencing factors

According to previous research and data availability (Hosseini et al., 2020; Tien Bui et al., 2020; Zhao et al., 2019), we initially selected ten urban flood influencing factors: elevation (E), aspect (As), slope (S), curvature (C), average precipitation/Day (P), land cover types (LC), normalized difference built-up index (NDBI), the fraction of vegetation cover (FVC), distance to river (DR) and population density (PD). We divided these influencing factors into three categories: topographic factors, precipitation factors, and land surface factors.

#### 2.4.1. Topographic factors

Topographic factors are important drivers that influence urban flood occurrences. Based upon the DEM, the topographical factors are derived either directly or indirectly. Among these factors, aspect, slope, and curvature were determined from the DEM, using ArcGIS 10.8.

(1) Elevation (E): Elevation is an important influencing factor of urban flood, as runoff will flow from high to lower elevation areas (Coulthard & Frostick, 2010). The elevation in Chengdu ranges from 322 to 7134 m (Fig. 2a).
(2) Aspect (As): Aspect influences the soil humidity and affects the flow directions (Costache, 2019). Therefore, it may have an indirect influence on urban flood occurrence (Tien Bui et al., 2020).
(3) Slope (S): Slope represents the elevation change rate. Slope can influence the runoff speed (Kassogué et al., 2017). Steeper slopes can cause rapid flows, while floods tend to occur on gentle slopes.
(4) Curvature (C): The profile curvature affects the acceleration and deceleration of flow and the planform curvature influences convergence and divergence of flow (Shahabi et al., 2021).

#### 2.4.2. Precipitation

Precipitation (P) is the direct cause of urban flood in Chengdu city (Ke et al., 2020). The precipitation amount and duration is key in urban flooding, especially precipitation of a high intensity in a short duration (Hong et al., 2018; Wu et al., 2020). Precipitation data were available as text data at a district-level resolution. The average daily precipitation at the district-level was used as precipitation influencing factor.

#### 2.4.3. Land surface factors

(1) Land cover types (LC): Runoff is influenced by different land cover characteristics due to its impact on infiltration and flow (Darabi et al., 2019). The Chengdu 2018 LC map at a 10 m resolution was obtained from finer resolution observation and monitoring (Gong et al., 2020). It contains eight classes: cropland, forest, grassland, shrubland, wetland, water, bare land, and impervious surface in Fig. 2e.

(2) Normalized difference built-up index (NDBI): Built-up area represents an impervious surface and therefore plays an important role in urban flooding. The NDBI map was derived from the Google Earth Engine (GEE) platform. As shown in Fig. 2f, higher NDBI values correspond with a higher building density. The NDBI values were obtained by Eq. (1):

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad (1)$$

where SWIR is the short-wave infrared band, and NIR is the near-

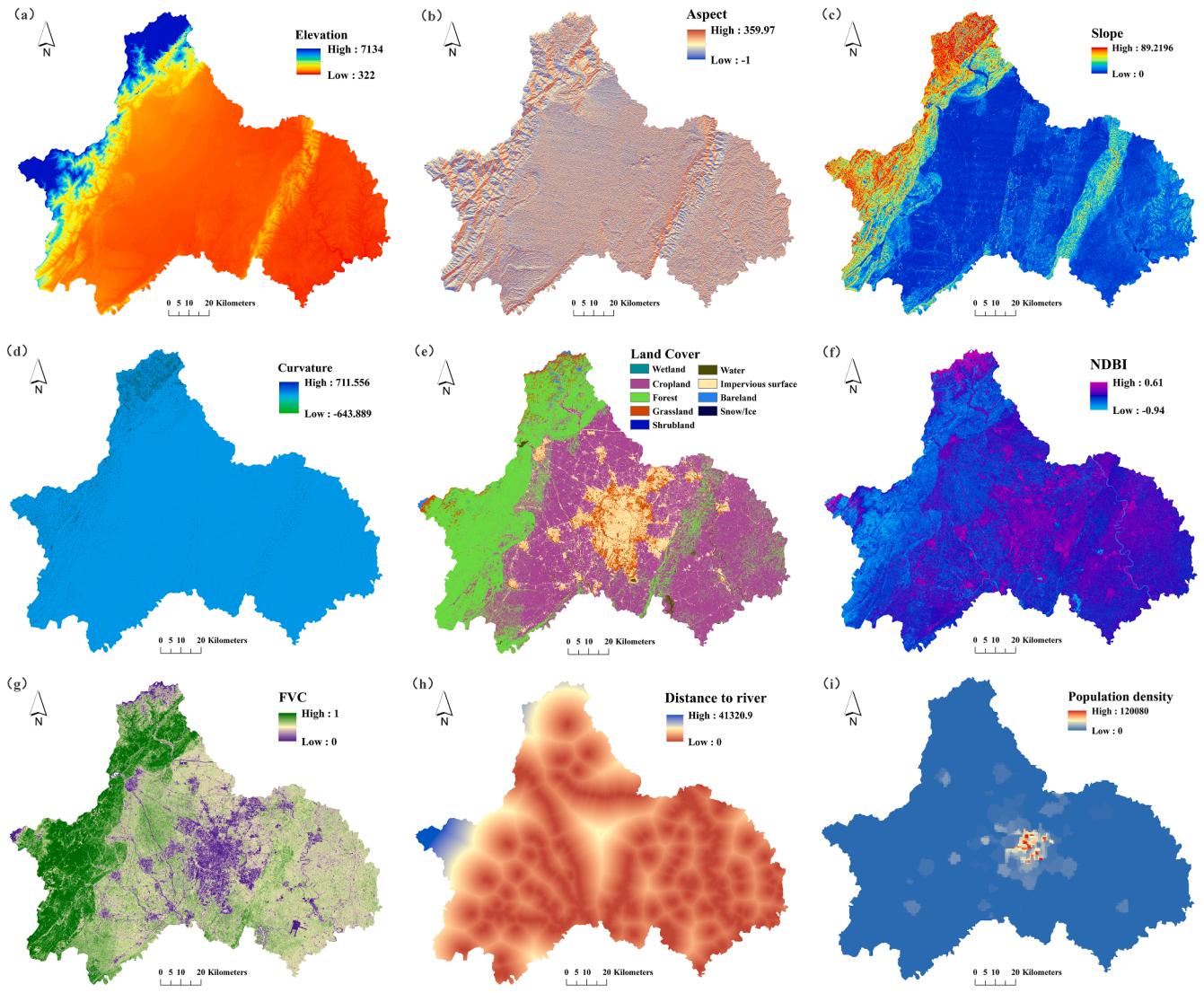



**Fig. 2.** Influencing factors of urban flooding in this study: (a) Elevation, (b) Aspect, (c) Slope, (d) Curvature, (e) Land cover types, (f) Normalized difference built-up index, (g) The fraction of vegetation cover, (h) Distance to river and (i) Population density.

infrared band.

(3) Fraction of Vegetation Cover (FVC): The FVC corresponds to the fraction of ground covered by green vegetation, quantifying the spatial extent of the vegetation (Chu, 2020). The FCV map (Fig. 2g) was obtained from Sentinal-2 satellite images acquired during the flood season in 2018 from the GEE platform, using the following formula:

$$FVC = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \tag{2}$$

where NDVI is the normalized difference vegetation index. The NDVI values were obtained by the red band (R) and the near-infrared band (NIR) using formula (3):

$$NDVI = \frac{NIR - R}{NIR + R} \tag{3}$$

(4) Distance to rivers (DR): Areas close to rivers are more susceptible to flooding than more distant areas (Chapi et al., 2017). Distance to rivers (Fig. 2h) was determined from the DEM using Euclidian Distance module in ArcGIS 10.8.1.

(5) Population density (PD): The impact of human activities on hydrology is huge, therefore we considered the influence of PD on urban flooding. The PD data was from Gridded Population of the World with the resolution of 1 km (https://sedac.ciesin.columbia.edu/data/collection/gpw-v4).

All influencing factors were rasterized and resampled to 30 m spatial resolution for spatial analysis and model development. The raster comprised of 6169 rows by 5018 columns, corresponding to 2,427,151 cells.

## 3. Methods

This study attempted to evaluate urban flood susceptibility based on social media data. To achieve this objective, the study followed the following steps: (1) pre-process the flood related social media data and collect environmental data; (2) analyze the historical urban flood spatiotemporal characteristics and find the AOI in study area; (3) select the main influencing factors by considering the complex urban features; and (4) train and validate the NB susceptibility model and map the urban flood susceptibility.

### 3.1. Spatiotemporal analysis of social media data

Since the distribution of social media data is heterogeneous, the spatial distribution of the data needs to be explored to find the highly

concentrated area as AOI. The standard deviation ellipse (SDE) developed by Lefever (1926), can capture approximately 68 % of the data (Xiang et al., 2021). It can help to find AOI that overcomes difficulties with conducting analysis in low social media density areas (Richter et al., 2021). It is a geospatial model that can analyze the spatial distribution characteristics of point data, central tendency, dispersion, and directional trends (Cheng et al., 2022). Assume the total number of social media data is $n$, then the location of a social media point has coordinates $(a_i, b_i)$ $(i = 1, 2, \ldots, n)$. There are mainly three steps to obtain the SDE:

(1) Determine the spatial mean center point of the ellipse $(a_0, b_0)$ as:

$$a_0 = \frac{\sum_{i=1}^{n} a_i}{n} \tag{4}$$

$$b_0 = \frac{\sum_{i=1}^{n} b_i}{n} \tag{5}$$

(2) Determine the rotation angle $\alpha$:

$$\tan\alpha = \frac{\sum_{i=1}^{n} \widetilde{a}_i^{\,2} - \sum_{i=1}^{n} \widetilde{b}_i^{\,2} + \sqrt{\left(\sum_{i=1}^{n} \widetilde{a}_i^{\,2} - \sum_{i=1}^{n} \widetilde{b}_i^{\,2}\right)^2 + 4\left(\sum_{i=1}^{n} \widetilde{a}_i \widetilde{b}_i\right)^2}}{2\sum_{i=1}^{n} \widetilde{x}_i \widetilde{b}_i} \tag{6}$$

where $\alpha$ refers to the azimuth of the ellipse, $\widetilde{a}_i = a_i - a_0, \widetilde{b}_i = b_i - b_0$.

(3) Determine the lengths of minor axe ($A$) and major axe ($B$) of the ellipse (Li et al., 2021). $A$ and $B$ are obtained as below:

$$A = \sqrt{\frac{\sum_{i=1}^{n} (\widetilde{a}_i \sin\alpha + \widetilde{b}_i \cos\alpha)^2}{n}} \tag{7}$$

$$B = \sqrt{\frac{\sum_{i=1}^{n} (\widetilde{a}_i \cos\alpha + \widetilde{b}_i \sin\alpha)^2}{n}} \tag{8}$$

where $\alpha$ refers to the azimuth of the ellipse, $\widetilde{a}_i = a_i - a_0, \widetilde{b}_i = b_i - b_0$.

### 3.2. Influencing factors selection process

#### 3.2.1. Multi-collinearity analysis of influencing factors

The basic assumption of Naïve Bayes modeling is that all influencing factors are independent (Rice, 2014). Therefore, it is necessary to check for multicollinearity of the urban flood influencing factors. The variance inflation factor (VIF), or its reciprocal T = 1/VIF, were used to assess multicollinearity:

$$VIF_i = \frac{1}{1 - R_i^2} \tag{9}$$

where $R_i^2$ is coefficient of determination, obtained by regressing the factor $i$ on all other factors in the analysis (Miles, 2014). Following O'brien (2007), a VIF of more than 10 and/or a T of less than 0.1 implies serious multicollinearity.

#### 3.2.2. Factor selection using recursive feature elimination algorithms

Factor selection selects key urban flood influencing variables, thus avoiding redundancy in factors during modeling. Recursive feature elimination algorithms (RFE) aim to improve generalization performance by removing the least important factors whose deletion will have the least effect on training errors (Chen & Jeong, 2007). In the absence of universal guidelines for urban flood factor selection, RFE was employed to select the key features.

### 3.3. Naïve Bayes and model training

The Naïve Bayes (NB) classifier is a classification algorithms based upon probability theory of the classical machine learning algorithms (Pérez et al., 2009). It relies on Bayes' theorem and assumes that the factor features are conditionally independent given a class (Pham et al., 2017). Therefore, classification converts to a maximum finding problem based upon maximum a posteriori estimation by comparing the values of each factor at every point with whether flooding would be expected or not expected.

#### 3.3.1. NB susceptibility model

The NB susceptibility model is constructed following four steps: (1) collecting flood and non-flood samples and the influencing factors, (2) estimating the prior probabilities of flood class and non-flood class, (3) determining each class label and compute covariance matrices for each class, and (4) obtaining the inverse and determinant to ultimately construct the discriminant function for each class.

In this study, let the vector $Y$ correspond to the predicted class variables $Y = (y_1, y_2)$ ($y_1$represents flood, and $y_2$ represents non-flood). Given a specific sample to be predicted, the explanatory variables are expressed by a vector $X = (x_1, .., x_m)$ representing the $m$ factors (independent variables). Using Bayes' theorem, the posterior probability of sample $X$ predicted as flood ($y_1$) equals:

$$P(Y = y_1|X) = \frac{P(Y = y_1)\prod_{j=1}^{m} P(x_j|Y = y_1)}{P(X)} \tag{10}$$

where $P(Y = y_1)$ is the prior probability of flooding, $P(x_j|Y = y_1)$ is the conditional probability of the $x_j$ given that flooding occurs and the denominator, $P(X)$ is the probability of $X$. Note that $P(X)$ is constant and is independent of the observations as the values of the factors $x_j$ are given.

The Naïve Bayes model is based upon maximum a posterior probability. Hence, the NB susceptibility model was formulated by the following equation:

$$f_{NB} = argmax\ P(Y = y_1) \prod_{j=1}^{m} P(x_j|Y = y_1) \tag{11}$$

where, $P(Y = y_1)$ was estimated from the training set and expressed as: $P(Y = y_1) = \frac{N_{y_1}}{N_{all}}$, in which $N_{y_1}$ means the number of flooding samples and $N_{all}$ means the total number of all training samples. The conditional probability $P(x_j|Y = y_1)$ is derived using maximum likelihood estimation, while the factor data format is of two types:

(1) For LC, being a categorical variable as input, the conditional probability equals

$$P(x_j|Y = y_1) = \frac{N_{x_j, y_1}}{N_{y_1}} \quad (j = 1, 2, 3, \ldots, m) \tag{12}$$

where $N_{x_j, y_1}$ is the number of samples whose factor $x_j$ belongs to flooding samples.

(2) For continuous variables as inputs, the Gaussian density function is used:

$$P(x_j|Y = y_1) = \frac{1}{\sigma_{j1}\sqrt{2\pi}} e^{\frac{-(x_j - \mu_{j1})^2}{2\sigma_{j1}^2}} \tag{13}$$

where $\sigma_{j1}$ and $\mu_{j1}$ are respectively the standard deviation and the mean of the factor $x_j$ in the flooding samples.

The NB susceptibility model was implemented in R Studio 4.0.3 using the 'klaR' package (Weihs et al., 2005).

#### 3.3.2. Model training and validation

The flooding and non-flooding records were assigned with the values 1 and 0, respectively. Then, the records were converted into pixels (30 × 30 m) and unified the coordinate system with the corresponding factors to form the final training and testing datasets. During the running process of the models, 10-fold cross-validation method was used to reduce variability of the model results (Sylvain & Alain, 2010).

### 3.4. Model performance evaluation

True positive (TP) and false positive (FP) represent the number of actual flood and non-flood samples that were predicted as flood points. The true negative (TN) and false negative (FN) represents the number of actual non-flood and flood that were predicted as non-flood points. Among these, TP and TN are correctly predicted. We used the accuracy (Eq. (14)) to evaluate the model predictive performance. Accuracy is the correctly predicted ratio of all samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

The receiver-operator characteristic (ROC) curve is the most crucial method applied for verification of the susceptibility models (El-Haddad et al., 2021). The ROC curve that plots the ratio of correctly predicted pixels by the model (True Positive Rate, TPR, also known as sensitivity) against the ratio of incorrectly predicted pixels (False Positive Rate, FPR, also known as 1-Specificity).

$$\text{TPR(Sensitivity)} = \frac{TP}{TP + FN} \tag{15}$$

$$\text{FPR}(1 - \text{Specificity}) = 1 - \frac{TN}{FP + TN} \tag{16}$$

The area under the ROC curve (AUC) is a significance measurement to assess the model performance (Gorsevski et al., 2006). Therefore, AUC is employed to validate the success of the model. It ranges from 0.5 (poor predictive ability) to 1 (the highest accuracy and reliability). Normally, values exceeding 0.8 indicate very good model performance (Chen et al., 2018).

Random Forest (RF) classifier proposed by Breiman in 2001 was selected for model comparison to test the performance of the NB model (Zhao et al., 2018). And the RF classifier was implemented with the "Random Forest" package in R Studio 4.0.3.

## 4. Results

### 4.1. The spatiotemporal analysis of social media flooding data

#### 4.1.1. Temporal analysis of social media data

According to the rainstorm level standard from China Meteorological Administration, there were two extremely heavy rainstorms on 2nd and 11th of July. The temporal distribution of the social media volume during July is shown in Fig. 3. It shows that the number of social media posts has almost the same change trends as precipitation, especially during the rainstorms. There are three high peaks on July 2nd, 11th and 15th corresponding with the highest number of labeled floods. These are the top three social media high peak events in July, 2018.

#### 4.1.2. Spatial analysis and AOI selection

The SDE model was used to visually express the spatial distribution of flood records during the top three social media high peak events and find AOI. The major SDE axis shows the data distribution direction, while the minor SDE axis represents the data distribution range (Li et al., 2021). The size of the ellipse reflects the spatial concentration of social media data, where the area inside the SDE is the concentrated area of spatial points. SDEs for the top three social media high peak events are shown in Fig. 4, with more recent dates in darker shades. At the three high peaks, changes of the ellipse clearly show dynamic spatial characteristics changes in the flooding influencing area. The results suggested that the urban flooding social media data are concentrated in the northeast-central part of Chengdu city, especially in the center of the city. Three SDEs coincided at the city center. Meanwhile, the spatial mean centers (SMCs) of the SDEs capture the changes in the distribution of urban flood events and show the movement over time. Three SMCs are all located in the center city. Therefore, the center of Chengdu city (including Chenghua, Jingjiang, Jinniu, Qingyang, and Wuhou districts) is set as the AOI where the flood records concentrated.

### 4.2. Influencing factors selection result

#### 4.2.1. Multi-collinearity analysis of influencing factors

Table 2 shows the multicollinearity diagnostics. The collinearity statistics results showed the VIF values of the ten influencing factors were far below the critical value (10). Meanwhile, the T values were higher than the critical value (0.1). It indicated that there was no multicollinearity between the 10 factors, therefore, these factors were all selected as model inputs.

#### 4.2.2. Contribution analysis of influencing factors

The NB method ranked the importance of different influencing factors (Fig. 5). The importance results showed that FVC had the highest importance (0.88), while the PD had the lowest importance (0.50). The green vegetation ratio of ground cover plays an important role in Chengdu city. The topographic factors show less importance than the land surface factors. The results showed that FVC, precipitation, and NDBI had high importance during urban flood susceptibility modelling and mapping, while population density had the lowest importance.

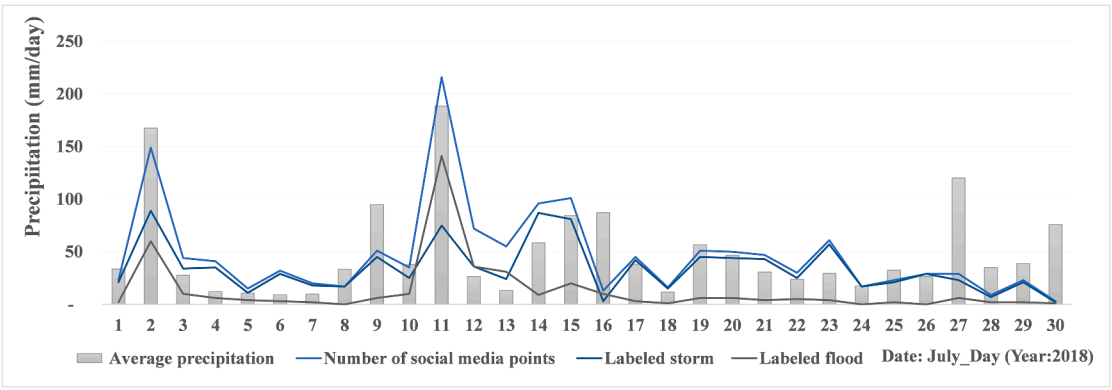



**Fig. 3.** Temporal trends of Weibo numbers and daily precipitation during July, 2018.

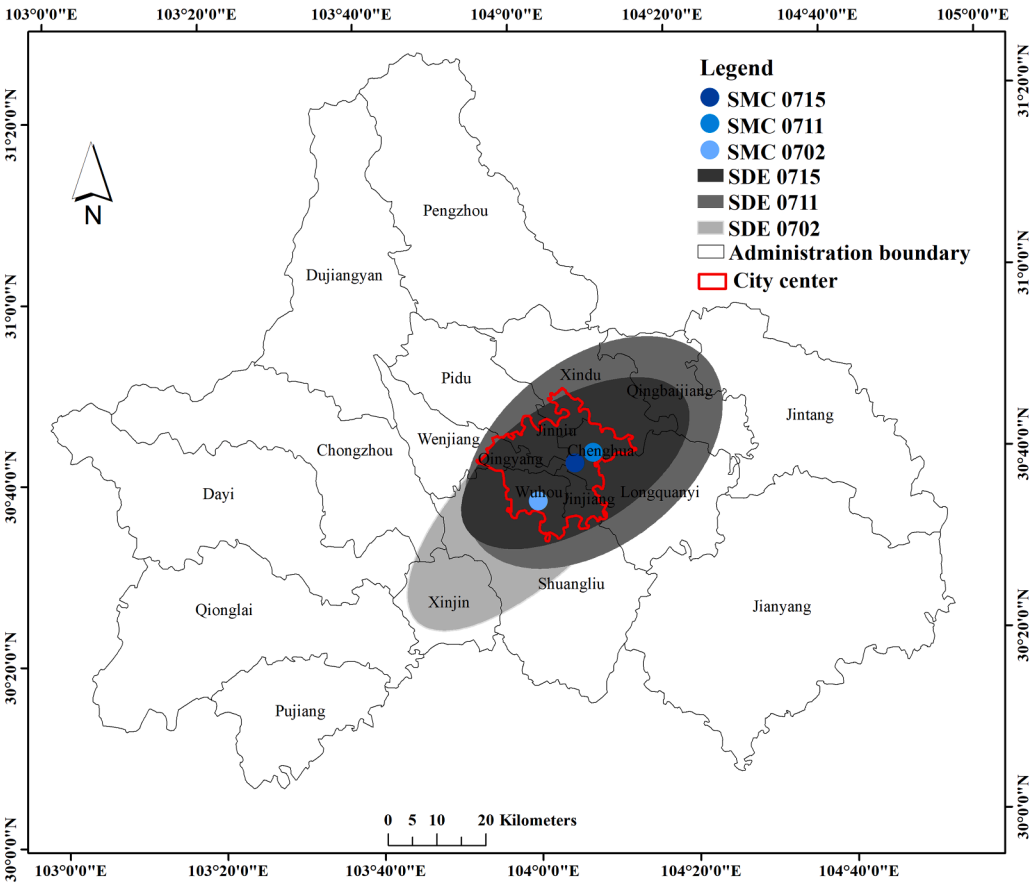



**Fig. 4.** The Standard Deviational Ellipse changes of three high peak events.

**Table 2**
The multi-collinearity diagnostics results.

| Influencing factors | | E | As | S | C | P | LC | NDBI | FVC | DR | PD |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Collinearity Statistics | T | 0.36 | 0.82 | 0.38 | 0.80 | 0.83 | 0.41 | 0.47 | 0.23 | 0.69 | 0.92 |
| | VIF | 2.81 | 1.22 | 2.65 | 1.25 | 1.20 | 2.46 | 2.13 | 4.27 | 1.46 | 1.08 |

**Note:** The initially ten influencing factors are: elevation (E), aspect (As), slope (S), curvature (C), average precipitation/Day (P), land cover types (LC), normalized difference built-up index (NDBI), the fraction of vegetation cover (FVC), distance to river (DR) and population density (PD).

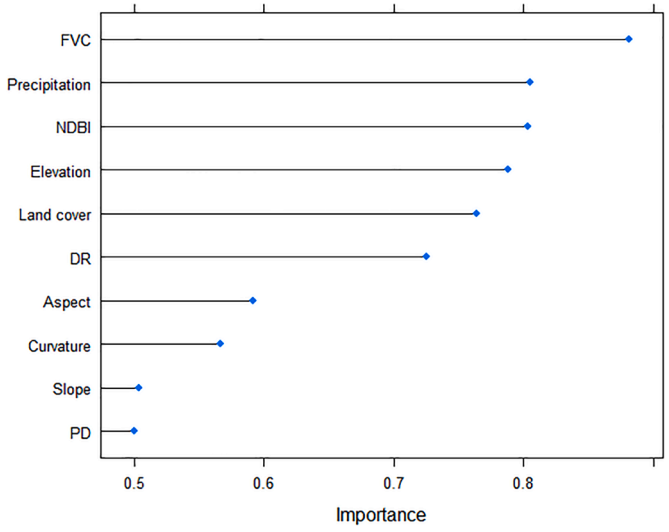



**Fig. 5.** The importance rank of the influencing factors.

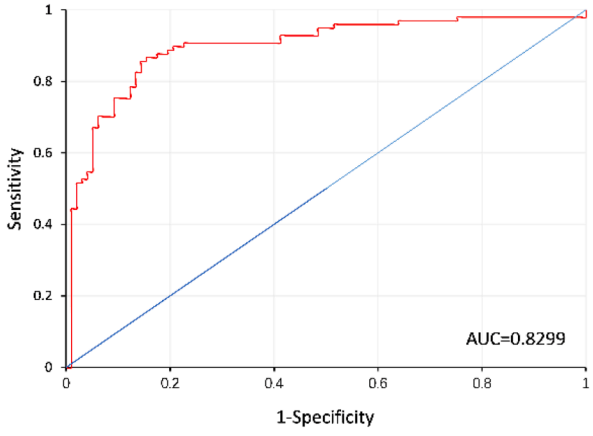



**Fig. 6.** The ROC curve of testing dataset.

### 4.3. Model performance results

The model achieved satisfactory accuracy performance with Accuracy=0.95. The ROC curve and AUC value (0.8299) of the testing dataset are shown in Fig. 6. Therefore, we think the model predicted performance is sufficient.

### 4.4. Flood susceptibility mapping results

After validating the NB model, the urban flood susceptibility of center area in Chengdu city is mapped (Fig. 7). According to the susceptibility map, the high urban flood susceptibility areas with red color were mainly concentrated in densely developed inner-city areas where mainly high-density building areas and roads. The area in blue color shows the low urban flood susceptibility areas is mostly distributed in mountain areas or parks with lush vegetation in the northeast, northwest, and southeast parts of the inner-city area. We selected two classic example areas in rectangular labeled "A" and "B" as shown in Fig. 7. Part A, Chengdu Research Base of Giant Panda Breeding is a classic example of low flood susceptibility area, its lush vegetation could resist urban flood. There are eight people posted here during extreme rainfall and only tagged storm without flood/inundation tag. Part B, a classic example of high-density urban commercial/ residential area, shows high urban flood susceptibility. The high-density impervious surface and flat terrain in Part B make it easy for rainfall to accumulate and to generate surface runoff, resulting high urban flood susceptibility.

## 5. Discussion

Urban flood susceptibility mapping is essential for urban planning and management to promote the sustainable development. However, the lack of hydrological monitoring data and the complex hydrological model limit the understanding process of urban flood. This study used social media data to carry out urban flood spatial analyses and map urban flood susceptibility in Chengdu city. As a new flood monitoring data source, social media data can provide refined spatiotemporal information about disasters with multidimensional information like timestamp, pictures, text, and location tag (Dou et al., 2021). More and more researchers began to address the importance of social media information for disaster assessment (Yao et al., 2021; Zhang et al., 2021). From the temporal analysis, this study found that the temporal changes of the social media data has a similar trend with the average precipitation, especially the extremely heavy rainstorm period. Zeng et al. (2020) also found the public was very much active on the web during the flooding event. This study has shown that social media data as a new data source has potential in assisting urban flood monitoring. From the spatial perspectives, the SDE model can find the flood density area as the AOI and measure the distribution of social media data.

Due to the complex urban flood process, the factors that influence urban flood are complicated (Chen et al., 2019). In summary, these influencing factors can be categorized into four categories: topographical factors (Botzen et al., 2013), precipitation factors (Ke et al., 2020), land surface characteristics (Zhang et al., 2020), and urban drainage capacity (Zhao et al., 2018). To improve the susceptibility model efficiency and performances, it is necessary to find the optimal influencing factors and remove the factors without contributions on the models. In our study, after multi-collinearity test, 10 influencing factors were finally selected. Among these factors, FVC is the most important influencing factor (0.88) based on the RFE algorithm. Similar factor importance rankings have been detected in previous studies (Chen et al., 2019; Hosseini et al., 2020). However, the influencing factors selection process changes from one study area to another due to the different local characteristics. For instance, Khosravi et al. (2019) found that altitude (0.99) and distance to the river (0.81) had the greatest effects on flood occurrence during flood susceptibility modeling in the Ningdu catchment. There, the flood types are mainly river floods which are influenced by topographic factors and the proximity to river. The appropriate influencing factors, therefore, need to be selected according to the

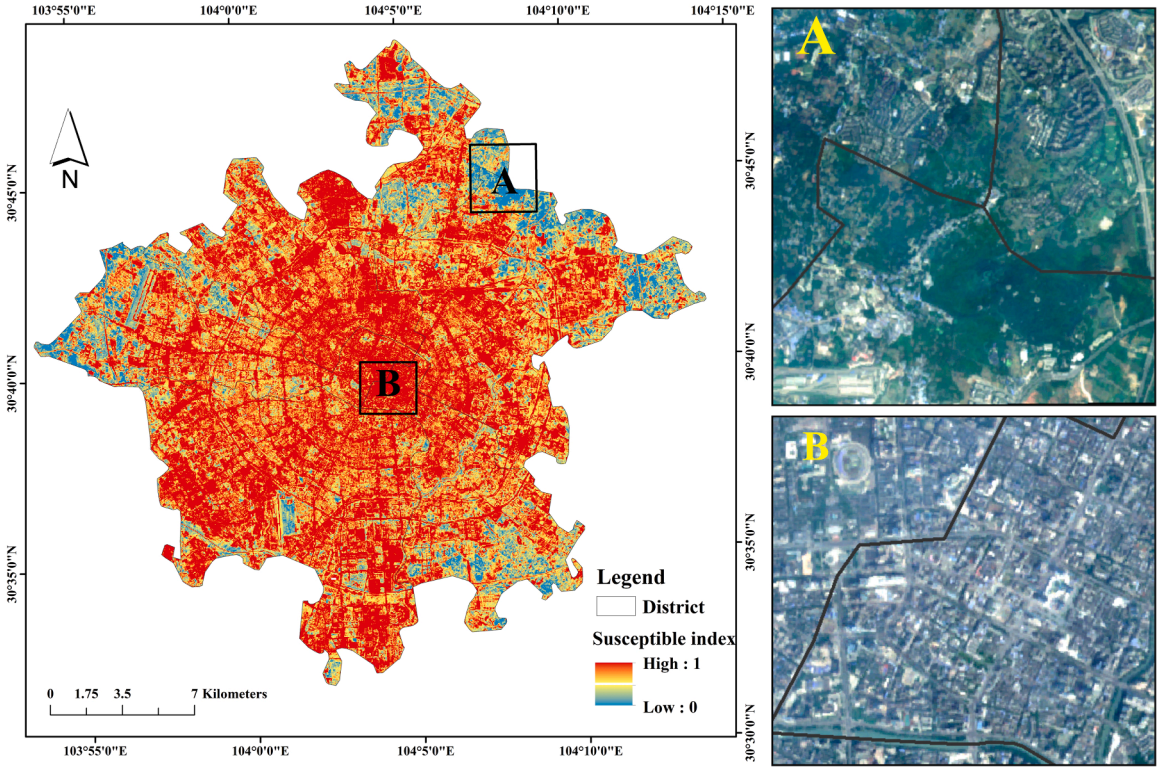



**Fig. 7.** The flood susceptibility map of center area in Chengdu. A is part of the Chengdu Research Base of Giant Panda Breeding and B is highly developed inner-city areas.

particular situation.

Comparing the validation results with a previous study (Zhao et al., 2019), we found that the NB model in this study returns accurate and stable results (AUC=0.830) even with limited data. According to model validation with only two class, flood and non-flood, the RF model (AUC=0.852) has a slightly better performance than the NB model. These results are in line with other researches (Khosravi et al., 2019; Pham et al., 2017). For instance, Pham et al. (2017) compared the landslide susceptibility assessment capability of NB, multilayer perceptron neural networks (MPNN), and functional trees (FT) methods in the Uttarakhand area. The results showed that the NB model (AUC = 0.838) had a slightly lower predictive capability as compared to the FT model (AUC = 0.849) and the MPNN model (AUC = 0.850). However, The NB model is based upon probability theory, providing posterior probabilities. It is thus much easier to understand than the other two models. Meanwhile, the NB model can deal with both discrete and continuous properties quickly and efficiently. From the model comparison of Khosravi et al. (2019), NB outperformed other models. In summary, NB is a simple, effective, and interpretable model that can successfully predict the urban flood by learning from the training dataset.

The urban flood susceptibility map can provide significant information for urban plan and flood hazard management, as it can help to mitigate the social-economic losses of urban spaces. The urban flood susceptibility results in Chengdu central area suggested that densely developed inner-city areas has high flood susceptibility and the low flood susceptibility areas are mainly concentrated in the suburbs with lush vegetation. This also explained why the FVC is the most important factor. This emphasizes the importance of urban green spaces in flood mitigation and adaptation. Zimmermann et al. (2016) also highlighted the importance of green area for the urban flood risk reduction. Dense vegetation can help to absorb, hold the water, and decrease the runoff. In this regard, city planners might develop green spaces and increase the vegetation density with a focus on the high urban flood risk area throughout the city . For the places which has no option to provide urban green space, Afriyanie et al. (2020) provided some suggestions for urban flood protection, such as the improvement of water retention, drainage, and sewage in high-risk areas, the application of building regulations that can reduce risk.

There are, however, some limitations in this study: (1) the social media data can easily be influenced by subjective factors; for example, some users may not share the geolocation considering their privacy. Meanwhile, we should consider the uncertainty of the real flood location and the location from social media data. In this research, every blog was reviewed manually and cross-validated by photos/video, mainstream media, and other flooding reports for the geographical locations. That may make the early acquisition, processing, and cleaning of social media data time-consuming. Therefore, disaster information extraction methods need to be further improved, including micro-blog precise geolocation information extraction and disaster information mining. In addition, areas without social media records may also be influenced by the flood. Thus, the number of urban flood occurrences may be underestimated in sparsely inhabited areas. (2) the data availability may affect the model performance. Due to the complexity and dynamic nature of urban flood, in order to further improve the model accuracy, more urban flood data and related influencing factors like drainage pipe density, the maximum of the total rainfall/day, rainfall duration and intensity should be collected and considered. (3) the model results only provide flood susceptibility mappings with flood susceptibility index. It can simulate the urban flood susceptibility under different influencing factors, but cannot reflect real-time flood situation including flood depth and velocity. In the future, we will try to integrate the hydrological/hydrodynamic models with machine learning algorithm for fast and large-scale urban flood susceptibility mapping and protection.

## 6. Conclusion

This study presents a spatiotemporal analysis and NB machine learning model integration framework to evaluate the urban flood susceptibility using social media data at the central area of Chengdu city, Sichuan, China. The main conclusions are as follows:

(1) The spatiotemporal analysis results of the social media data imply that there are three urban flood related social media high peaks in July, 2018 (on July 2$^{nd}$, 11$^{th}$, and 15$^{th}$) and the historical urban flood events are concentrated in the northeast-central part of Chengdu city, especially the center of city area. SDEs can help to find the AOI. The same trend with extreme storm means social media data can be reliably used for urban flood detection and susceptibility mapping. Furthermore, the social media data and NB susceptibility model of our study framework can also be extended to other regions or natural hazards that are influenced by meteorological and environment conditions, such as wildfires, landslides, avalanches.
(2) There were 10 urban flood related influencing factors including elevation, aspect, slope, curvature, average precipitation/Day, land cover types, NDBI, FVC, distance to river, and population density finally selected for model input. The rationality of these factors was diagnosed by multicollinearity test of tolerance and VIF. Among the factors, the FVC factor had the highest importance during the urban flood susceptibility modelling. This highlights the importance of green spaces and vegetation density in urban flood resilience.
(3) The NB urban flood susceptibility model was validated and evaluated by using 10-fold cross-validation, Accuracy confusion matrix, and the ROC curve. The results showed a very good performance (Accuracy=0.95, AUC=0.83) and the urban flood susceptibility of center area in Chengdu city is mapped. The resultant urban flood susceptibility map estimates reveal where urban floods are likely to occur under certain environmental conditions, which can facilitate understanding the easy-flood area and provide Chengdu government a realistic reference for decision-making.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## References

Afriyanie, D., Julian, M. M., Riqqi, A., Akbar, R., Suroso, D. S. A., & Kustiwan, I. (2020). Re-framing urban green spaces planning for flood protection through socio-ecological resilience in Bandung City, Indonesia. *Cities, 101*, Article 102710. https://doi.org/10.1016/j.cities.2020.102710

Bai, Y. B., Wu, W. Q., Yang, Z. X., Yu, J. Z., Zhao, B., Liu, X., … Koshimura, S. (2021). Enhancement of detecting permanent water and temporary water in flood disasters by fusing sentinel-1 and sentinel-2 imagery using deep learning algorithms: Demonstration of Sen1Floods11 benchmark datasets. *Remote Sensing*, (11), 13. https://doi.org/10.3390/rs13112220

Balz, T., Hammer, H., & Auer, S. (2015). Potentials and limitations of SAR image simulators – A comparative study of three simulation approaches. *ISPRS Journal of Photogrammetry and Remote Sensing, 101*, 102–109. https://doi.org/10.1016/j.isprsjprs.2014.12.008

Botzen, W. J. W., Aerts, J. C. J. H., & Van Den Bergh, J. C. J. M (2013). Individual preferences for reducing flood risk to near zero through elevation. *Mitigation and Adaptation Strategies for Global Change, 18*(2), 229–244. https://doi.org/10.1007/s11027-012-9359-5

Casali, Y., Aydin, N. Y., & Comes, T. (2022). Machine learning for spatial analyses in urban areas: A scoping review. *Sustainable Cities and Society*, 85. https://doi.org/10.1016/j.scs.2022.104050

Chang, H. J., Pallathadka, A., Sauer, J., Grimm, N. B., Zimmerman, R., Cheng, C. W., … Herreros-Cantis, P. (2021). Assessment of urban flood vulnerability using the social-ecological-technological systems framework in six US cities. *Sustainable Cities and Society*, 68. https://doi.org/10.1016/j.scs.2021.102786

Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T., & Khosravi, K. (2017). A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software, 95*, 229–245. https://doi.org/10.1016/j.envsoft.2017.06.012

Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang, X., & Ahmad, B. B. (2019). Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *Journal of Hydrology, 575*, 864–873. https://doi.org/10.1016/j.jhydrol.2019.05.089

Chen, W., Peng, J. B., Hong, H. Y., Shahabi, H., Pradhan, B., Liu, J. Z., … Duan, Z. (2018). Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Science of The Total Environment, 626*, 1121–1135. https://doi.org/10.1016/j.scitotenv.2018.01.124

Chen, X., & Jeong, J. C. (2007). Enhanced recursive feature elimination. *Paper presented at the sixth international conference on machine learning and applications (ICMLA 2007)*.

Cheng, T., Xu, Z., Hong, S., & Song, S. (2017). Flood risk zoning by using 2D hydrodynamic modeling: A case study in Jinan City. *Mathematical Problems in Engineering, 2017*, Article 5659197. https://doi.org/10.1155/2017/5659197

Cheng, T., Zhao, Y. H., & Zhao, C. J. (2022). Exploring the spatio-temporal evolution of economic resilience in Chinese cities during the COVID-19 crisis. *Sustainable Cities and Society*, 84. https://doi.org/10.1016/j.scs.2022.103997

Chu, D. (2020). Fractional vegetation cover. *Remote sensing of land use and land cover in mountain region* (pp. 195–207). Singapore: Springer.

Costache, R. (2019). Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models. *Science of The Total Environment, 659*, 1115–1134. https://doi.org/10.1016/j.scitotenv.2018.12.397

Coulthard, T. J., & Frostick, L. E. (2010). The Hull floods of 2007: Implications for the governance and management of urban drainage systems. *Journal of Flood Risk Management, 3*(3), 223–231. https://doi.org/10.1111/j.1753-318x.2010.01072.x

Darabi, H., Choubin, B., Rahmati, O., Torabi Haghighi, A., Pradhan, B., & Kløve, B. (2019). Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *Journal of Hydrology, 569*, 142–154. https://doi.org/10.1016/j.jhydrol.2018.12.002

Deng, Q., Liu, Y., Zhang, H., Deng, X. L., & Ma, Y. F. (2016). A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan. *Natural Hazards, 84*(2), 1241–1256. https://doi.org/10.1007/s11069-016-2484-9

Dou, M., Wang, Y., Gu, Y., Dong, S., Qiao, M., & Deng, Y. (2021). Disaster damage assessment based on fine-grained topics in social media. *Computers & Geosciences, 156*, Article 104893. https://doi.org/10.1016/j.cageo.2021.104893

El-Haddad, B. A., Youssef, A. M., Pourghasemi, H. R., Pradhan, B., El-Shater, A. H., & El-Khashab, M. H. (2021). Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt. *Natural Hazards, 105*(1), 83–114. https://doi.org/10.1007/s11069-020-04296-y

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., … Zhou, Y. (2020). Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment, 236*, Article 111510. https://doi.org/10.1016/j.rse.2019.111510

Gorsevski, P. V., Gessler, P. E., Foltz, R. B., & Elliot, W. J. (2006). Spatial prediction of landslide hazard using logistic regression and ROC analysis. *Transactions in GIS, 10*(3), 395–415. https://doi.org/10.1111/j.1467-9671.2006.01004.x

Helmrich, A. M., Ruddell, B. L., Bessem, K., Chester, M. V., Chohan, N., Doerry, E., … Zahura, F. T. (2021). Opportunities for crowdsourcing in urban flood monitoring. *Environmental Modelling & Software, 143*, Article 105124. https://doi.org/10.1016/j.envsoft.2021.105124

Hénonin, J., Hongtao, M., Zheng-Yu, Y., Hartnack, J., Havnø, K., Gourbesville, P., & Mark, O. (2015). Citywide multi-grid urban flood modelling: The July 2012 flood in Beijing. *Urban Water Journal, 12*(1), 52–66. https://doi.org/10.1080/1573062X.2013.851710

Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A. X., … Kazakis, N. (2018). Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Science of The Total Environment, 621*, 1124–1141. https://doi.org/10.1016/j.scitotenv.2017.10.114

Hosseini, F. S., Choubin, B., Mosavi, A., Nabipour, N., Shamshirband, S., Darabi, H., & Haghighi, A. T. (2020). Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Science of The Total Environment, 711*, Article 135161. https://doi.org/10.1016/j.scitotenv.2019.135161

Kankanamge, N., Yigitcanlar, T., Goonetilleke, A., & Kamruzzaman, M. (2020). Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets. *International Journal of Disaster Risk Reduction*, 42. https://doi.org/10.1016/j.ijdrr.2019.101360

Kassogué, H., Bernoussi, A., Maâtouk, M., & Amharref, M. (2017). A two scale cellular automaton for flow dynamics modeling (2CAFDYM). *Applied Mathematical Modelling, 43*, 61–77. https://doi.org/10.1016/j.apm.2016.10.034

Ke, Q., Tian, X., Bricker, J., Tian, Z., Guan, G., Cai, H., … Liu, J. (2020). Urban pluvial flooding prediction by machine learning approaches – A case study of Shenzhen city, China. *Advances in Water Resources*, 145. https://doi.org/10.1016/j.advwatres.2020.103719

Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., … Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology, 573*, 311–323. https://doi.org/10.1016/j.jhydrol.2019.03.073

Lefever, D. W. (1926). Measuring geographic concentration by means of the standard deviational ellipse. *American Journal of Sociology, 32*(1), 88–94. https://doi.org/10.1086/214027

Li, Hu, Zheng, Shen, & Fan, & Zhang. (2019). An improved simplified urban storm inundation model based on urban terrain and catchment modification. *Water, 11*(11), 2335. https://doi.org/10.3390/w11112335

Li, D., Chen, H., Shen, W., & Ye, M. (2021). An analysis of the temporal and spatial gathering and dispersion patterns of crowds at the community level after the 2020 M5.1 Tangshan Guye earthquake. *International Journal of Disaster Risk Reduction, 61*, Article 102331. https://doi.org/10.1016/j.ijdrr.2021.102331

Li, Z., Xie, j., Yang, T., & Mou, N (2018). *A multi-source spatiotemporal dataset of floods in Chengdu*.

Liu, J., Shao, W., Xiang, C., Mei, C., & Li, Z. (2020). Uncertainties of urban flood modeling: Influence of parameters for different underlying surfaces. *Environmental Research, 182*, Article 108929. https://doi.org/10.1016/j.envres.2019.108929

Miles, J. (2014). Tolerance and variance inflation factor. *Wiley StatsRef: Statistics reference online*.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Pérez, A., Larrañaga, P., & Inza, I. (2009). Bayesian classifiers based on kernel density estimation: Flexible classifiers. *International Journal of Approximate Reasoning, 50*(2), 341–362. https://doi.org/10.1016/j.ijar.2008.08.008

Pham, B. T., Tien Bui, D., Pourghasemi, H. R., Indra, P., & Dholakia, M. B (2017). Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of Naïve Bayes, multilayer perceptron neural networks, and functional trees methods. *Theoretical and Applied Climatology, 128*(1-2), 255–273. https://doi.org/10.1007/s00704-015-1702-9

Rice, D. M. (2014). Chapter 4 - Causal reasoning. In D. M. Rice (Ed.), *Calculus of thought* (pp. 95–123). Academic Press.

Richter, A., Ng, K. T. W., Karimi, N., & Chang, W. J. (2021). Developing a novel proximity analysis approach for assessment of waste management cost efficiency in low population density regions. *Sustainable Cities and Society*, 65. https://doi.org/10.1016/j.scs.2020.102583

Rodriguez, E. L., Berrocal, F. L., & Rodriguez, A. B. M. (2021). Flood hazards in Extremadura and associated impacts. *Investigaciones Geograficas-Spain*, (75), 121–137. https://doi.org/10.14198/ingeo.16990

Rosenzweig, B. R., Herreros Cantis, P., Kim, Y., Cohn, A., Grove, K., Brock, J., … Chang, H. (2021). The value of urban flood modeling. *Earth's Future, 9*(1), Article e2020EF001739. https://doi.org/10.1029/2020EF001739

Shahabi, H., Shirzadi, A., Ronoud, S., Asadi, S., Pham, B. T., Mansouripour, F., … Bui, D. T. (2021). Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. *Geoscience Frontiers, 12*(3), Article 101100. https://doi.org/10.1016/j.gsf.2020.10.007

Songchon, C., Wright, G., & Beevers, L. (2021). Quality assessment of crowdsourced social media data for urban flood management. *Computers, Environment and Urban Systems, 90*, Article 101690. https://doi.org/10.1016/j.compenvurbsys.2021.101690

Sun, D., Li, S., Zheng, W., Croitoru, A., Stefanidis, A., & Goldberg, M. (2016). Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery. *International Journal of Digital Earth, 9*(5), 427–441. https://doi.org/10.1080/17538947.2015.1040474

Sylvain, A., & Alain, C. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*(none), 40–79. https://doi.org/10.1214/09-SS054

Tang, X., Li, J., Liu, M., Liu, W., & Hong, H. (2020). Flood susceptibility assessment based on a novel random Naïve Bayes method: A comparison between different factor discretization methods. *Catena, 190*, Article 104536. https://doi.org/10.1016/j.catena.2020.104536

Tien Bui, D., Hoang, N.-D., Martínez-Álvarez, F., Ngo, P. T. T., Hoa, P. V., Pham, T. D., … Costache, R. (2020). A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Science of The Total Environment, 701*, Article 134413. https://doi.org/10.1016/j.scitotenv.2019.134413

Wang, Z., Lam, N. S. N., Obradovich, N., & Ye, X. (2019). Are vulnerable communities digitally left behind in social responses to natural disasters? An evidence from Hurricane Sandy with Twitter data. *Applied Geography, 108*, 1–8. https://doi.org/10.1016/j.apgeog.2019.05.001

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). klaR Analyzing German business cycles. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data analysis and decision support* (pp. 335–343). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wu, X., Wang, Z., Guo, S., Liao, W., Zeng, Z., & Chen, X. (2017). Scenario-based projections of future urban inundation within a coupled hydrodynamic model framework: A case study in Dongguan City, China. *Journal of Hydrology, 547*, 428–442. https://doi.org/10.1016/j.jhydrol.2017.02.020

Wu, Z., Zhou, Y., Wang, H., & Jiang, Z. (2020). Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse. *Science of the Total Environment, 716*, Article 137077. https://doi.org/10.1016/j.scitotenv.2020.137077

Xiang, Y., Huang, C. B., Huang, X., Zhou, Z. X., & Wang, X. S. (2021). Seasonal variations of the dominant factors for spatial heterogeneity and time inconsistency of land surface temperature in an urban agglomeration of central China. *Sustainable Cities and Society*, 75. https://doi.org/10.1016/j.scs.2021.103285

Yang, K. X., Hou, H., Li, Y., Chen, Y., Wang, L. Y., Wang, P., & Hu, T. A. (2022). Future urban waterlogging simulation based on LULC forecast model: A case study in

Haining City, China. *Sustainable Cities and Society*, 87. https://doi.org/10.1016/j.scs.2022.104167

Yao, K., Yang, S., & Tang, J. (2021). Rapid assessment of seismic intensity based on Sina Weibo — A case study of the changning earthquake in Sichuan Province, China. *International Journal of Disaster Risk Reduction, 58*, Article 102217. https://doi.org/10.1016/j.ijdrr.2021.102217

Zeng, Z., Lan, J., Hamidi, A. R., & Zou, S. (2020). Integrating Internet media into urban flooding susceptibility assessment: A case study in China. *Cities, 101*, Article 102697. https://doi.org/10.1016/j.cities.2020.102697

Zhang, H., Zhang, J. P., Fang, H. Y., & Yang, F. (2022). Urban flooding response to rainstorm scenarios under different return period types. *Sustainable Cities and Society*, 87. https://doi.org/10.1016/j.scs.2022.104184

Zhang, Q., Wu, Z., Zhang, H., Dalla Fontana, G., & Tarolli, P. (2020). Identifying dominant factors of waterlogging events in metropolitan coastal cities: The case study of Guangzhou, China. *Journal of Environmental Management, 271*, Article 110951. https://doi.org/10.1016/j.jenvman.2020.110951

Zhang, Y., Chen, Z., Zheng, X., Chen, N., & Wang, Y. (2021). Extracting the location of flooding events in urban systems and analyzing the semantic risk using social sensing data. *Journal of Hydrology, 603*, Article 127053. https://doi.org/10.1016/j.jhydrol.2021.127053

Zhao, G., Pang, B., Xu, Z., Peng, D., & Xu, L. (2019). Assessment of urban flood susceptibility using semi-supervised machine learning model. *Science of the Total Environment, 659*, 940–949. https://doi.org/10.1016/j.scitotenv.2018.12.217

Zhao, G., Pang, B., Xu, Z., Yue, J., & Tu, T. (2018). Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of The Total Environment, 615*, 1133–1142. https://doi.org/10.1016/j.scitotenv.2017.10.037

Zheng, Z., Gao, J., Ma, Z., Wang, Z., Yang, X., Luo, X., … Fu, G. (2016). Urban flooding in China: Main causes and policy recommendations. *Hydrological Processes, 30*(7), 1149–1152. https://doi.org/10.1002/hyp.10717

Zimmermann, E., Bracalenti, L., Piacentini, R., & Inostroza, L. (2016). urban flood risk reduction by increasing green areas for adaptation to climate change. *Procedia Engineering, 161*, 2241–2246. https://doi.org/10.1016/j.proeng.2016.08.822

Zou, L., Lam, N. S. N., Cai, H., & Qiang, Y. (2018). Mining Twitter data for improved understanding of disaster resilience. *Annals of the American Association of Geographers, 108*(5), 1422–1441. https://doi.org/10.1080/24694452.2017.1421897