

Flood Susceptibility Mapping in Fangshan District Based on Convolutional Neural Networks

Jiankun Xu, Jia Li

Abstract

Under the influence of climate change, extreme precipitation events have become more frequent and intense, leading to an increasing occurrence of urban flooding and waterlogging. The catastrophic "23·7" extreme rainfall event in Fangshan District, Beijing, incurred substantial losses, highlighting the critical urgency of urban flood governance. For predicting the spatial distribution of waterlogging, traditional physically-based hydrological and hydrodynamic models rely heavily on high-resolution local data for parameter calibration, while generally suffering from prohibitive computational costs and limited generalization across diverse scenarios. Recently, deep learning approaches have emerged as promising alternatives for urban waterlogging simulation.

This study develops a flood susceptibility assessment framework leveraging a LeNet-style Convolutional Neural Network (CNN). Using Fangshan District as the study area, the framework uses vehicle insurance claims as ground-truth flood observations from the "23·7" storm into reliable ground-truth labels. By systematically integrating 11 conditioning factors and utilizing high-resolution Digital Elevation Model (DEM) data to extract fine-grained spatial neighborhood features, the framework is developed to generate high-precision risk maps.

Quantitative results indicate that the 11-factor LeNet-CNN model achieves strong predictive performance, yielding an Area Under the Curve (AUC) of 0.876. Systematic ablation studies confirm the prominent role of fine-grained topographic factors (DEM) in waterlogging susceptibility prediction, while sensitivity analyses validate the robustness of the network architecture under real-world data imbalances. This study validates the adaptability of the LeNet-CNN model under diverse hydro-geographical constraints, providing robust technical support for urban emergency response planning and differentiated disaster insurance services.

Keywords: Flood Susceptibility; Convolutional Neural Network; Deep Learning; Insurance Data; Spatial Modeling; Fangshan; Beijing.

1 Introduction

Flooding is one of the most destructive natural hazards worldwide, causing devastating impacts on human societies, infrastructure systems, and economic stability (Hirabayashi et al., 2013; Yin et al., 2015; United Nations Office for Disaster Risk Reduction, 2020; Dhiman et al., 2019). According to recent data, flood-related catastrophes account for a substantial proportion of global economic losses, with annual damages exceeding tens of billions of dollars (Munich Re, 2022). This situation is exacerbated by global climate change, which has intensified the frequency of extreme precipitation events (Masson-Delmotte et al., 2021; Kundu et al., 2017). Meanwhile, rapid urbanization has substantially altered underlying surface conditions.

The expansion of impervious surfaces (e.g., roads and rooftops) drastically reduces natural infiltration and shortens the time of concentration for runoff (Du et al., 2012; Shuster et al., 2005). The combined effects of climate change and urbanization renders transitional geographical zones—such as Fangshan District in Beijing—exceptionally vulnerable to flash floods and urban waterlogging (Liu et al., 2021; Feng et al., 2015).

To deal with these risks, Flood Susceptibility Mapping (FSM) is essential for land-use planning and emergency response. Historically, physically-based hydrodynamic models were widely used but are often constrained by prohibitive computational costs and data scarcity (Tehrany et al., 2014; Teng et al., 2017). Consequently, data-driven Machine Learning (ML) approaches have gained prominence. Algorithms such as Logistic Regression (LR) (Pradhan, 2010), Support Vector Machines (SVM) (Tehrany et al., 2014), Artificial Neural Networks (ANN) (Kia et al., 2012), and Random Forests (RF) (Wang et al., 2018) have been widely applied. However, these conventional algorithms generally rely on the assumption of independent and identically distributed samples. They typically treat study units as isolated pixels, using only local attributes while ignoring the spatial topology and complex physical-social interactions of the infrastructure and surroundings (Rahman et al., 2021; Mosavi et al., 2018; Dong et al., 2020).

To address these limitations, Deep Learning (DL), specifically Convolutional Neural Networks (CNNs), has emerged as a promising solution (LeCun et al., 2015; Sit et al., 2020). First proposed by LeCun et al. (LeCun et al., 1998), CNNs utilize convolutional layers to automatically extract high-level spatial hierarchies from local receptive fields, achieving significant success in computer vision tasks (El-Sawy et al., 2016; Zhang et al., 2019; Li et al., 2019). Inspired by this, Zhao et al. (2020) proposed a pioneering LeNet-style CNN framework that integrates topographic and hydrological factors to map flood susceptibility. Their study, supported by subsequent research (Khosravi et al., 2020; Bentivoglio et al., 2022), demonstrated that CNNs compare favorably with pixel-based baselines by effectively capturing spatial neighborhood information.

However, the efficacy of data-driven models heavily depends on the quality of ground-truth labels. Traditional inventories derived from post-disaster surveys or remote sensing are often incomplete or obstructed by cloud cover (Sanyal and Lu, 2004). While some studies have explored social media or crowdsourced data as alternatives (Fohringer et al., 2015; Restrepo-Estrada et al., 2018), these sources often suffer from location ambiguity and subjective bias. Consequently, there is a critical need for high-precision, objectively verified flood occurrence data to validate deep learning models in complex urban environments.

To address these data challenges, this study focuses on the “23·7” Extreme Rainfall Event in Fangshan District (July 2023) and introduces a novel vehicle insurance dataset as the ground truth. From a financial risk perspective, these claims provide precise geolocations of submerged vehicles, serving as indisputable, objective evidence of waterlogging (Spekkers et al., 2014; Grahn and Wanders, 2020; Zhou et al., 2021). By integrating these high-quality empirical records with the LeNet-based architecture, This study aims to: (1) overcome the limitations of traditional flood inventories; and (2) propose a novel pathway to bridge the gap between technical simulation and financial risk assessment.

2 Study Area and Data Description

2.1 Study Area Description

Located in the southwest of Beijing, Fangshan District ($39^{\circ}29' - 39^{\circ}56' \text{ N}$, $115^{\circ}24' - 116^{\circ}15' \text{ E}$; area: $1,989 \text{ km}^2$) serves as a critical topographical interface between the Taihang Mountains and the North China Plain (Figure 1). The landscape is defined by a dramatic elevation drop, descending from northwestern peaks of over 2000 m to southeastern plains below 30 m . This rugged terrain channels several major river systems, such as the Dashi and Juma Rivers. The region is governed by a semi-arid to semi-humid continental monsoon climate, characterized by concentrated summer precipitation; over 70% of the annual rainfall occurs from June to August, frequently driving high-intensity flood events.

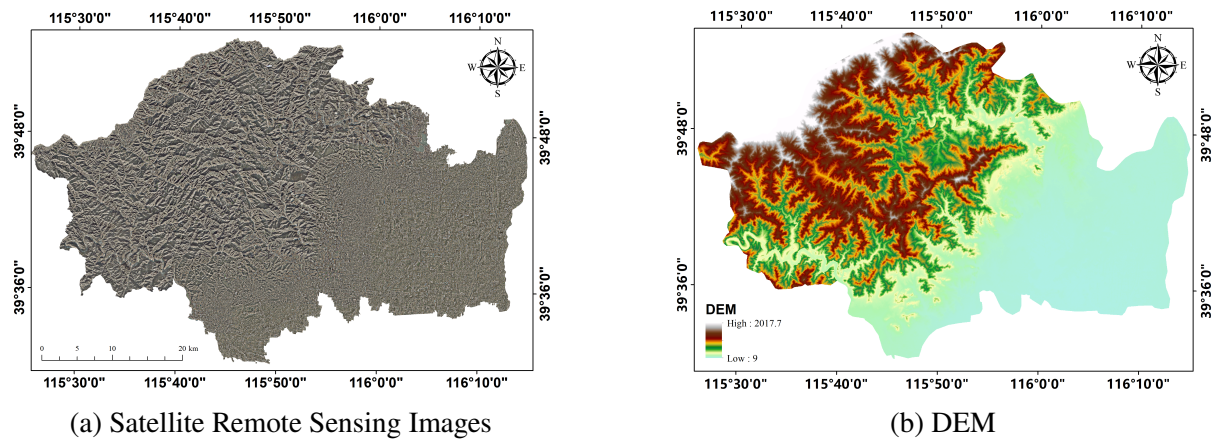


Figure 1: Geographic Location and Digital Elevation Model (DEM) of Fangshan District, Beijing, China.

These unique geographical and climatic characteristics make Fangshan District exceptionally susceptible to flood disasters. The sharp elevation drop facilitates the rapid concentration of surface runoff, allowing rainwater from the expansive mountainous catchments to swiftly converge into the low-lying plains. During periods of intense precipitation, the natural drainage capacity of the river networks is often overwhelmed, while the rapid urbanization in the plains has further altered the surface properties, increasing runoff coefficients and shortening the time of concentration. Consequently, the combination of mountainous terrain, concentrated rainfall, and expanding impervious surfaces creates a highly sensitive environment for both flash floods and urban waterlogging.

The vulnerability of this region has been starkly demonstrated by several catastrophic events in recent history. On July 21, 2012, Fangshan was the epicenter of a historic "7.21" extreme rainstorm, which caused severe flooding and significant socio-economic losses. In late July 2023, the "23.7" extraordinary heavy rainfall event triggered by Typhoon Doksuri led to unprecedented flood levels in the Dashi and Juma river basins. These events, characterized by their high intensity and rapid onset, underscore the persistent threat of flooding in Fangshan and the critical importance of accurate susceptibility assessment for this transition zone.

2.2 Data Collection

In This study, flooded locations were identified based on existing insurance data. In the collection of car accidents happened from June 2021 to August 2023, 1917 insurance claims

were identified, including some that happened at the same location. We obtained the WGS84 geographic coordinates (latitude and longitude) for each incident location. Part of the insurance data is shown in Table 1.

Table 1: Sample of Insurance Data

Warranty Number	Type	Date	Place of Incident
XXXX0000692588	Business	July 31, 2023	Beiluyuan Residential Community, Xilu Sub-district, Fangshan District, Beijing
XXXX0000122958	Business	July 31, 2023	No. 77, Sucun 4th District, Fangshan District, Beijing
XXXX0000973354	Business	July 31, 2023	Beiluxin Residential Community, Fangshan District, Beijing
XXXX0001329070	Business	July 31, 2023	Jiazhou Shuijun Community, Changyang Town, Fangshan District, Beijing

Specifically, we focused on the incidents that occurred on July 31st, which correspond to the "23.7" extraordinary heavy rainfall event, and removed the redundant points. After field verification (Figure 2), we made a further selection of those flooded points. The final dataset contains 880 flooded points.



Figure 2: Field Observation of the Flood-prone Areas

2.3 Conditioning factors

A convincing and prospective flood-predicting model requires considering influential factors that modulate the probability of floods. In this model, eleven conditioning factors classified into topographic, hydrological and drainage, anthropogenic land-surface, and soil-related categories are collected and described. Such a classification enhances physical interpretability and facilitates the understanding of how different processes contribute to flood susceptibility within the deep learning framework.

Ten conditioning factors (excluding DEM, which is shown before) are shown in Figure 3.

2.3.1 Topographic Factors

Topographic factors fundamentally control surface runoff routing, flow convergence, and local water accumulation, thereby exerting a primary influence on flood initiation and spatial

distribution.

The Digital Elevation Model (DEM) provides a digital representation of terrain elevation and forms the fundamental basis for describing surface morphology. It plays a key role in flood susceptibility analysis by controlling runoff routing, flow convergence, and water accumulation, and also serves as the primary dataset for deriving secondary topographic and hydrological factors. The DEM used in this study was generated from high-resolution satellite stereo imagery through photogrammetric processing, including aerial triangulation, surface modeling, and subsequent filtering and interpolation, resulting in a raster dataset with a spatial resolution of 5 m referenced to the WGS84 vertical datum.

Slope represents the rate of elevation change and directly influences runoff velocity, infiltration capacity, and flood residence time. Slope was derived from the DEM as the maximum rate of elevation change between each cell and its neighbors. In continuous terms, slope (in radians) can be expressed as

$$\beta = \arctan \left(\sqrt{\left(\frac{\partial z}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial y} \right)^2} \right),$$

where z denotes elevation and $\partial z/\partial x$, $\partial z/\partial y$ are the partial derivatives of elevation in the horizontal directions. Slope values are reported in degrees. Gentle slopes favor surface water retention and prolonged inundation, whereas steep slopes primarily contribute to rapid runoff generation and downstream flood amplification.

The Topographic Wetness Index (TWI) integrates upstream contributing area and local slope to quantify the tendency of a location to accumulate water and reach saturation. It serves as a proxy for spatial soil moisture distribution and flow convergence. TWI was calculated using

$$\text{TWI} = \ln \left(\frac{A_s}{\tan \beta} \right),$$

where A_s denotes the specific catchment area and β is the local slope angle. Higher TWI values indicate zones more susceptible to surface saturation and runoff generation, making this index particularly effective for identifying depressions and valley bottoms prone to flooding.

2.3.2 Hydrological and Drainage Factors

Hydrological and drainage-related factors describe the efficiency of surface water conveyance and the structural controls governing flood propagation within a catchment.

Drainage density (DD) quantifies the degree of channel development within a catchment and is defined as the total length of streams per unit area:

$$\text{DD} = \frac{1}{A} \sum_{i=1}^n L_i,$$

where L_i is the length of the i th stream segment within the analysis unit, n is the number of stream segments, and A is the unit area. Low DD often indicates limited natural or artificial drainage capacity, which may promote surface water accumulation and elevate local flood susceptibility; conversely, high DD generally enhances runoff conveyance but can accelerate flood transmission during extreme events.

The river network (RN) represents the spatial structure and connectivity of natural drainage pathways. Its configuration governs runoff routing, floodwater redistribution, and overflow

behavior during high-flow conditions. In urban environments, altered or constrained river networks can further intensify flood risk by limiting channel conveyance capacity.

Distance to river (DRI) is defined as the Euclidean distance from each grid cell to the nearest river (channel) pixel:

$$\text{DRI}(x, y) = \min_{(x_r, y_r) \in \mathcal{R}} \sqrt{(x - x_r)^2 + (y - y_r)^2},$$

where $\mathcal{R}$ denotes the set of river channel pixel coordinates. DRI is computed in meters using the projected coordinate system. Smaller DRI values indicate larger exposure to fluvial hazards, whereas larger values indicate locations where pluvial mechanisms may dominate flood generation.

The Manning's Roughness Coefficient (n) characterizes the frictional resistance of the land surface to overland flow, fundamentally governing the velocity and routing of surface runoff. Unlike purely topographic factors, this coefficient is quantitatively estimated based on high-resolution soil property data. Specifically, the proportions of clay, silt, sand, and organic matter within the soil matrix dictate the local hydraulic roughness. Higher values of n indicate increased surface friction, which dissipates the kinetic energy of runoff and modulates the timing of flow concentration. By incorporating these soil-derived physical properties, the model captures complex rainfall-runoff response characteristics that are not fully represented by elevation or slope alone.

2.3.3 Anthropogenic and Land Surface Factors

Anthropogenic land-surface factors describe human-induced modifications that significantly alter natural hydrological responses, particularly in urbanized regions.

Impervious Surface Area (ISA) is defined as the fraction of surface area that is covered by impermeable materials and is computed as

$$\text{ISA} = \frac{A_{\text{impervious}}}{A_{\text{total}}},$$

where $A_{\text{impervious}}$ is the area of impermeable surfaces within a given analysis unit (e.g., grid cell or neighborhood) and A_{total} is the unit area. ISA was derived from high-resolution land-cover products using binary classification of impervious versus pervious surfaces. Higher ISA values reduce infiltration and increase runoff coefficients, thus elevating flood susceptibility in urbanized zones.

Land Use/Land Cover (LULC) describes the spatial distribution of different surface types, including built-up areas, forests, croplands, wetlands, and grasslands. Each land-cover class exhibits distinct hydrological behaviors in terms of interception, infiltration, and runoff generation.

Distance to road was computed as the Euclidean distance from each cell to the nearest road segment:

$$\text{DRO}(x, y) = \min_{(x_d, y_d) \in \mathcal{D}} \sqrt{(x - x_d)^2 + (y - y_d)^2},$$

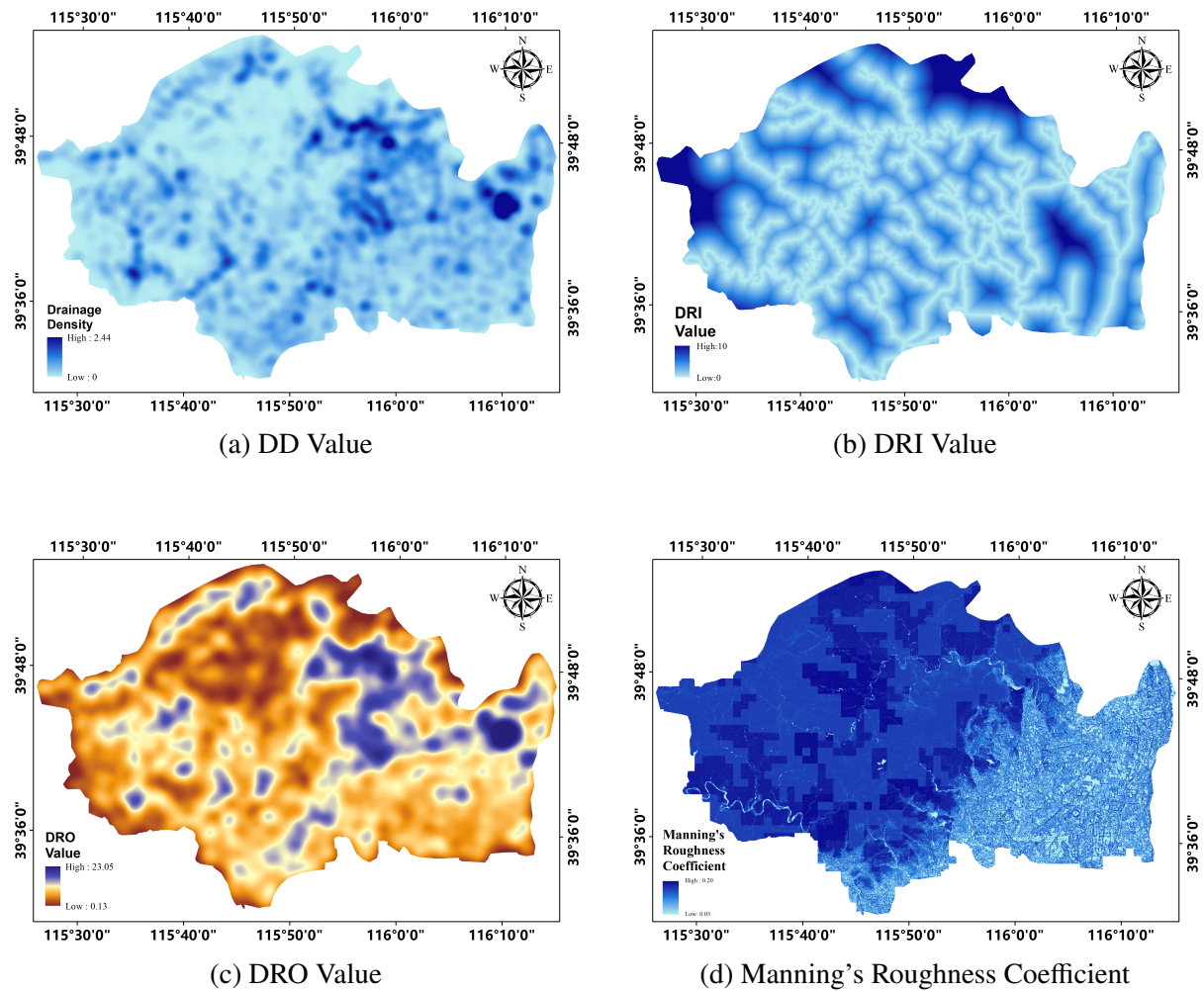
where $\mathcal{D}$ denotes the set of road pixel coordinates. Roads can act as preferential flow paths or barriers depending on their design, and proximity to roads may alter local runoff routing and accumulation patterns.

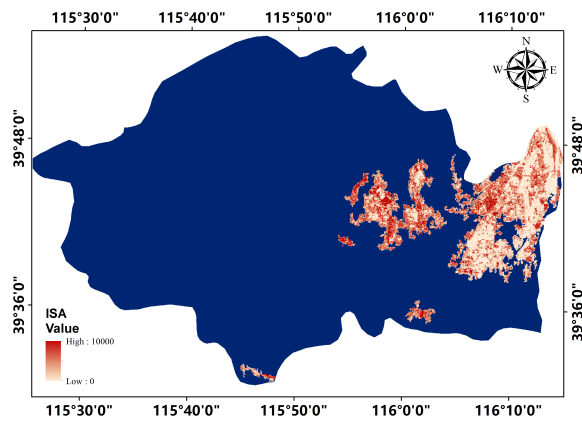
2.3.4 Soil-Related Factors

Soil-related factors regulate infiltration capacity, permeability, and subsurface water storage, thereby influencing both surface runoff generation and flood persistence.

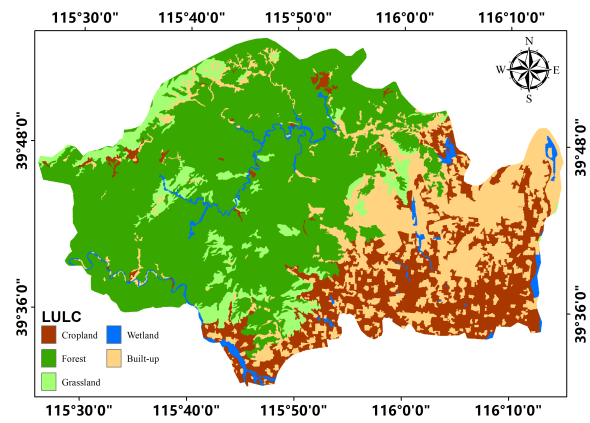
The soil distribution types used in this study are derived from the “1:1,000,000 Soil Map of the People’s Republic of China” digital product, provided by the Resource and Environment Science and Data Center (Resource and Environment Science and Data Center, 2023). This national dataset employs the traditional genetic soil classification system with the suborder as the basic mapping unit. It delineates twelve soil orders, sixty-one soil types, and 227 soil subtypes that reflect differences in soil-forming processes, parent materials, and physiographic conditions.

Each soil type represents a spatial unit with distinct physical and hydraulic characteristics (e.g., texture, porosity, and saturated hydraulic conductivity) that directly affect infiltration efficiency and surface runoff generation. Soils dominated by fine particles (e.g., clay-rich classes) typically exhibit lower permeability and smaller saturated hydraulic conductivity, which tend to increase surface runoff during short-duration intense rainfall and thus raise local flood susceptibility; conversely, coarse-textured soils generally favor infiltration and subsurface storage, thereby reducing the likelihood of surface water accumulation.

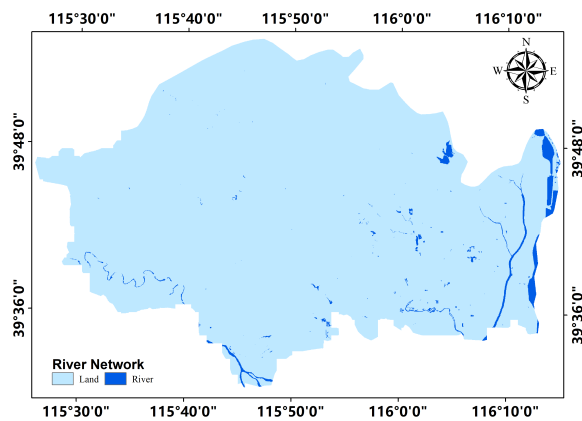




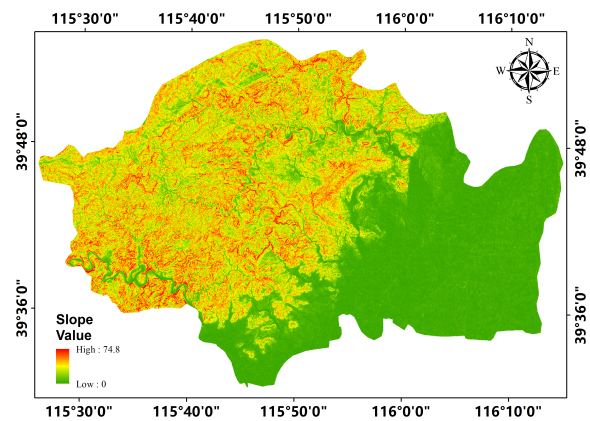
(e) ISA Value



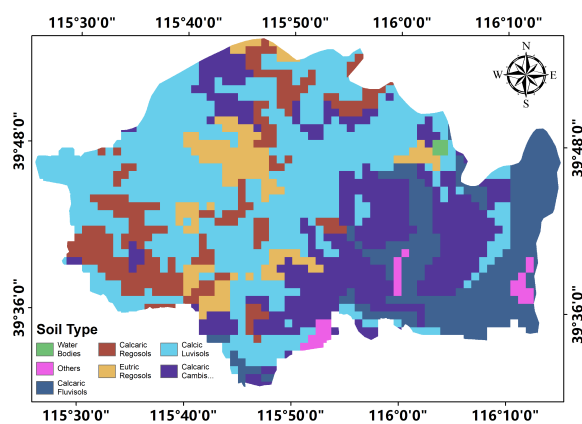
(f) LULC Type



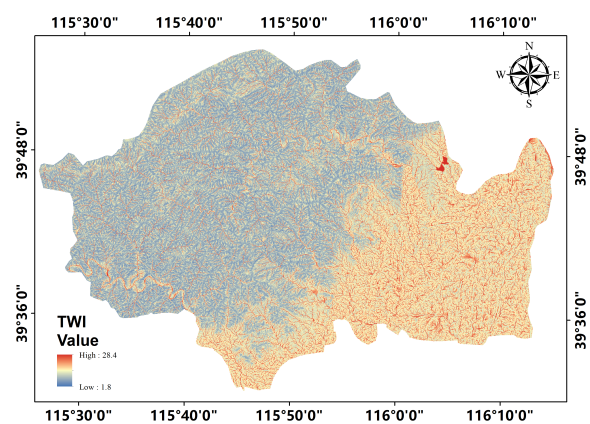
(g) River Network



(h) Slope Value



(i) Soil Type



(j) TWI Value

Figure 3: Ten Conditioning Factors Used for Flood Susceptibility Modeling in the Study Area

3 Model Description

This section discusses the training model and related processes.

3.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNN) are a class of deep, feed-forward artificial neural networks that have proven highly effective in analyzing spatially correlated data. Unlike traditional machine learning models that treat environmental factors as independent pixels, CNNs leverage a hierarchical structure to automatically extract spatial features and capture complex nonlinear relationships between flood conditioning factors and flood occurrences.

The architecture typically consists of three primary types of layers:

- **Convolutional Layers:** These layers apply a set of learnable filters to the input images to produce feature maps. This process allows the model to identify local patterns such as topographical gradients or drainage textures.
- **Pooling Layers:** These layers perform down-sampling operations to reduce the spatial dimensions of the feature maps, and thereby decrease computational complexity and providing a degree of translational invariance.
- **Fully Connected Layers:** Located at the end of the network, these layers integrate the high-dimension features extracted by the previous layers to perform the final classification into flood or non-flood categories.

3.2 LeNet-Style Model

In practice, there are different CNN architectures, while no universally optimal CNN architecture exists. As a result, several training models have been developed and demonstrated for their efficiency and convenience, including LeNet (LeCun et al., 1998), AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), and ResNet (He et al., 2016).

The LeNet-5 structure was first proposed by LeCun et al. in 1998 (LeCun et al., 1998); it was then developed for handwriting recognition (El-Sawy et al., 2016) and spread into other fields (Zhang et al., 2019; Li et al., 2019). This study adopted a LeNet-style CNN architecture for the flood susceptibility assessment, as shown in Figure 4. The training model in the research includes one inputted layer, two convolutional layers which are alternatively followed by one pooling layer, one fully connected layer, and one output layer.

3.3 Input Image Construction

Compared to traditional pointwise supervised training models like Supported Vector Machine (SVM) and Random Forest (RF), which operate in a point-wise supervised learning manner, CNNs like LeNet operate with input containing information of the training sites as well as the surroundings. The input samples are combined images with dimensions $iw \times iw \times N$, where iw represents the width of each image, and N represents the number of factors that are combined in the input. In this study, the input samples are tensors of shape (23, 23, 11). Each sample reflects the geological information of the flooded site as well as that of the surroundings.

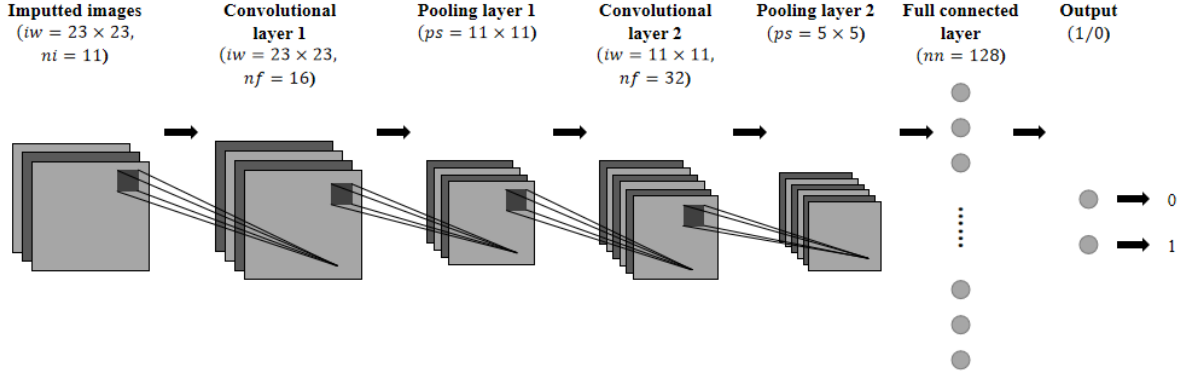


Figure 4: Architecture of the Proposed LeNet-style CNN Model (iw : image width; wf : width of filters; nf : number of filters; ps : pool size; nn : number of neurons)

3.4 Other Model Parameters

In this study, the model training was conducted using the Adam optimizer with a learning rate of 0.001, which provides an adaptive mechanism for efficient weight updates. The training process was executed for 20 epochs with a mini-batch size of 128. The activation function is ReLU. The convolutional kernel size is 3×3 , and the pooling method is max pooling (2×2). A gradient clipping strategy with a maximum norm of 1.0 was applied during backpropagation. The Cross-Entropy loss function was utilized to minimize the error between the predicted susceptibility and the actual flood occurrence. Due to the inherent imbalance between flood and non-flood samples, a random sampling strategy was implemented to ensure data balance and model objectivity. Specifically, a random function was used to select a set of negative samples (non-flood points) within the study area equal in number to the flooded sample points. This approach prevents the model from developing a bias toward the majority class and ensures that the learned features are equally representative of both flooded and non-flooded conditions.

3.5 Performance Evaluation

To evaluate the predictive performance of the CNN model, the dataset is divided into training and testing sets. Based on the binary classification results, where flood instances are designated as positive and non-flood instances as negative, a confusion matrix is constructed. This matrix includes True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN):

$$A = \begin{bmatrix} TP & TN \\ FP & FN \end{bmatrix}. \quad (1)$$

The following quantitative metrics are utilized to assess the model:

1. **Precision (P):** Measures the accuracy of the positive predictions, representing the ratio of correctly identified flood points to the total number of points predicted as floods.

$$P = \frac{TP}{TP + FP} \quad (2)$$

2. **Recall (R):** Also known as sensitivity, it measures the proportion of actual flood points

that were correctly identified by the model.

$$R = \frac{TP}{TP + FN} \quad (3)$$

3. **F_1 -score (F_1):** The harmonic mean of Precision and Recall, providing a single score that balances both metrics, especially useful in cases of class imbalance.

$$F_1 = \frac{2PR}{P + R} \quad (4)$$

4. **Average Accuracy (AA):** Represents the overall proportion of correct classifications (both flood and non-flood) among the total number of samples.

$$AA = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

5. **Kappa Coefficient (κ):** This metric assesses the agreement between the model predictions and the ground truth, accounting for the possibility of agreement occurring by chance.

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (6)$$

where P_o is the observed proportionate agreement (AA), and P_e is the probability of random agreement.

6. **Area Under the ROC Curve (AUC):** A critical metric for evaluating the performance of classification models at various threshold settings. It represents the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR). The AUC value ranges from 0 to 1, where a value of 1.0 indicates a perfect model and 0.5 suggests a model with no predictive capability. TPR and FPR are defined as:

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN}. \quad (7)$$

The AUC can be calculated using the trapezoidal rule to integrate the area under the curve:

$$AUC = \int_0^1 TPR d(FPR) \quad (8)$$

Alternatively, it can be expressed based on the ranks of all positive and negative samples:

$$AUC = \frac{\sum_{i \in \text{positive}} \text{rank}_i - \frac{n_p(n_p+1)}{2}}{n_p \times n_n} \quad (9)$$

where n_p and n_n are the number of positive and negative samples, respectively, and rank_i is the rank of the i -th positive sample when all samples are sorted by their predicted probabilities.

4 Result Analysis

4.1 Model Result and Accuracy

Using the model described above, a flood susceptibility map is developed, as shown in Figure 5.

The susceptibility map indicates that the areas with the highest flood susceptibility are mainly concentrated in four regions, all with a high density of buildings. Three out of four regions are located in the eastern urban centers.

- A1 region covers the corridors along S320 (108 Fuxian) and S328 (Liangsan Road) within Fangshan District. Distributed along the Dashi River and winding through the mountainous terrain, these two roads are significantly affected by the local topography. This configuration causes localized precipitation to rapidly convert into surface runoff, which subsequently accumulates in the vicinity of the roadbeds.
- A2 region mainly covers the Chengguan Neighborhood. The Chengguan Neighborhood is to the west of Western Hills. After the rain, the surface runoff from the surrounding mountains will flow down the slopes and converge into the Chengguan area, increasing the probability of flooding and waterlogging.
- A3 region covers the triangular region formed by Yancun Town, Wangzuo Town, and Liangxiang Town in Fangshan District. Due to the exceptionally high building and vehicle densities, this urban sector is highly susceptible to the accumulation of surface runoff and exhibits high flood exposure.
- A4 region covers the nearby area of Doudian Town. Due to its concentrated high-density buildings, the region is also an area where surface runoff is prone to surface water accumulation.

In contrast, regions with relatively lower flood susceptibility are mainly distributed in areas with low built-up density or in the mountains. These areas, due to their steep slopes and high terrain, are not prone to becoming accumulation zones for runoff.

4.2 Model Validation and Sensitivity Analysis

To evaluate the robustness and robustness of the proposed CNN framework, we conducted a comprehensive validation process from two perspectives: (1) an ablation study to verify the contribution of key conditioning factor, DEM, and (2) a sensitivity analysis of the sampling ratio to assess model performance under imbalanced data scenarios.

4.2.1 Effect of Topographic Factors

To examine the necessity of topographic factors, we compared the standard model (trained with all 11 factors) against a reduced model trained with only 9 factors (excluding DEM and slope). The visual comparison of the resulting susceptibility maps is shown in Figure 6 and Figure 7.

Qualitatively, the map produced by the 9-factor model exhibits a distinct phenomenon of excessive spatial smoothing and loss of spatial heterogeneity. Without the micro-topographic constraints of elevation and slope, the network struggles to express detailed information. It

over-generalizes the waterlogging risk across the low-lying plain areas, merging regions A1 to A4 into an expansive, contiguous zone of high or highest susceptibility. For example, the 9-factor model incorrectly classifies some areas with low susceptibility as with highest susceptibility. The lack of spatial heterogeneity fails to capture the fine-grained variations necessary for precision street-level emergency planning.

Quantitatively, the 9-factor model correspondingly exhibits an abnormal inflation across several statistical indicators, with Recall rising to 0.908 and the ROC-AUC reaching 0.920, as shown in Table 2. We infer that this inflation occurs because the ground-truth vehicle insurance claims are heavily clustered within the dense urban built-up areas of the eastern plain with massive vehicles. Without detailed topographical data, the 9-factor model homogenizes big areas in the east consistently while simplifies the scattered, small areas in the west, thereby boosting its statistical recall. However, the overestimation comes at a critical cost: the model’s Precision suffers a major drop from 0.863 to 0.735, representing a severe surge in false alarms across safe urban pixels. It indicates that omitting DEM and Slope degrades the model’s localized discriminative power, and stresses that topographic factors are indispensable for maintaining spatial heterogeneity and reducing costly false positives.

This performance gap is visualized in the ROC analysis shown in Figure 8. By comparing the 11-Factor (1:1) curve (Blue) with the 9-Factor (1:1) curve (Orange), we observe that the ablated curve exhibits an artificially higher position with an AUC of 0.920 due to this smoothing effect.

4.2.2 Sensitivity Analysis

In real-world scenarios, flood events are rare compared to non-flood occurrences. To evaluate the robustness of our model under such conditions, we conducted a sensitivity analysis by altering the ratio of positive (flood) to negative (non-flood) samples from the balanced 1:1 ratio to an imbalanced 1:3 ratio.

The results are presented in the third column of Table 2 and in Figure 8.

Remarkably, the Recall remained constant even when the negative samples were increased from 1:1 to 1:3. This indicates that the model’s ability to detect true flood events is robust and does not deteriorate with increased background noise. As illustrated in Figure 8, the ROC curve for the 1:3 scenario closely follows the standard 1:1 curve, with the AUC even slightly increasing to 0.891. This suggests that the model maintains excellent separability between classes regardless of the data distribution. A decrease in Precision from 0.863 to 0.709 was observed, which is a expected owing to the increased prevalence of the negative class outcome when the prevalence of the negative class increases. However, the F1-score (0.697) and Kappa (0.556) remain at satisfactory levels.

Table 2: Model Performance Evaluation Metrics Comparison across Different Settings

Metric	11 Factors (1:1)	9 Factors (1:1)	11 Factors (1:3)
Precision (P)	0.863	0.735	0.709
Recall (R)	0.686	0.908	0.686
F1-score (F_1)	0.765	0.812	0.697
Average Accuracy (AA)	0.749	0.790	0.807
Kappa Coefficient (κ)	0.503	0.581	0.556
AUC	0.876	0.920	0.891

We now compare our results with existing standards.

First, according to the classification standards in natural hazard assessments (Hosmer and Lemeshow, 2000; Tehrany et al., 2014), an AUC value between 0.8 and 0.9 is considered "excellent" predictive capability. Our model's AUC of 0.876 places it in the competitive predictive performance, comparable to or exceeding recent deep learning benchmarks in similar complex terrains (e.g., Zhao et al. (2020), who reported AUCs ranging from 0.82 to 0.88 for urban flooding). This confirms that the proposed CNN architecture effectively captures the nonlinear spatial dependencies that traditional models often miss.

Second, given the significant spatial heterogeneity of flood risks in Fangshan District—where flood-prone areas are geographically concentrated rather than uniformly distributed—the F1-score provides a more robust measure than AA by harmonizing Precision and Recall. Our model achieved an F1-score of 0.765. In the domain of data-driven flood mapping, F1-scores typically range between 0.65 and 0.80 depending on data quality (Wang et al., 2018). Our result lies at the higher end of this spectrum, indicating that the model successfully minimizes false alarms (High Precision: 0.863) without compromising the detection of actual flood events.

Finally, the Kappa coefficient (κ) of 0.503 falls within the range of moderate to substantial agreement ($0.41 < \kappa < 0.60$) according to the Landis and Koch (1977) scale. Unlike perfectly delineated hydrological simulations, insurance data contains inherent stochastic noise (e.g., varying claim reporting locations). A Kappa exceeding 0.50 in this context is acceptable, and this demonstrates that the model has learned robust physical patterns rather than overfitting to the noise in the training data. These findings suggest that shows promising potential for practical large-scale screening..

5 Conclusion

In this study, we developed a LeNet-style Convolutional Neural Network (CNN) framework to assess flood susceptibility in Fangshan District, Beijing, a region characterized by complex topographical transitions. By integrating 11 conditioning factors, including DEM, DRO, DRI, etc., the model effectively captured the relationships between environmental features and historical flood events. Unlike traditional pixel-based machine learning methods such as SVM or RF, the proposed CNN architecture effectively incorporated spatial neighborhood information, thereby improving the spatial consistency and accuracy of the susceptibility mapping.

Quantitative validation confirms the efficacy and robustness of the proposed framework. The optimal model achieved an AUC of 0.876 and an F1-score of 0.765, demonstrating high capability in distinguishing flood-prone areas. The ablation study suggested that topographic factors are important for capturing local spatial variations. Although the 9-factor model showed higher statistical recall, this may be partly related to the spatial clustering of sample points in the plain area. The integration of all 11 factors helps maintain a more reasonable spatial distribution of risks, keeps spatial heterogeneity, and reduces potential false positives. Furthermore, the model exhibited promising stability under an imbalanced sampling ratio (1:3), maintaining a consistent Recall of 0.686 and achieving an AUC of 0.891, thereby verifying its robustness for practical risk screening in realistic, data-imbalanced scenarios.

Furthermore, we analyzed and proposed the flood susceptibility map in Fangshan District. We demonstrated four main regions, A1 to A4, with significantly high flood susceptibility. These areas feature low elevation, flat terrain, and high urbanization, which facilitate surface runoff accumulation and impede drainage. Conversely, the northwestern mountainous regions are generally classified as low susceptibility zones, except in narrow valleys and steep slopes

prone to flash floods.

Although the proposed CNN model demonstrated satisfactory performance in flood susceptibility mapping, further optimization of the proposed model remains possible. In this study, the validation process primarily focused on the selection of factors and the comparison of a limited subset of model hyperparameters. An exhaustive investigation into the model's architecture or a comprehensive grid search for global parameter optimization was not conducted due to the scope of this study. Future research could focus on a more systematic sensitivity analysis and extensive hyperparameter tuning to further unlock the potential of deep learning in this domain. In addition, further investigations are warranted to explore how the inherent spatial bias and clustering of empirical source data, such as vehicle insurance data, might skew spatial statistical evaluation, which may help establish more adaptive validation paradigms for urban flood modeling.

Acknowledgement

The DEM data set is provided by the Geographic Remote Sensing Ecological Network Platform (www.gisrs.cn).

References

- Bentivoglio, R., Isufi, E., Jonkman, S.N., Taormina, R., 2022. Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrology and Earth System Sciences* 26, 4345–4378. doi:10.5194/hess-26-4345-2022.
- Dhiman, R., VishnuRadhan, R., Eldho, T.I., Inamdar, A., 2019. Flood risk and adaptation in Indian coastal cities: recent scenarios. *Applied Water Science* 9, 5.
- Dong, S., Esmalian, A., Farahmand, H., Mostafavi, A., 2020. An integrated physical-social analysis of disrupted access to critical facilities and community service-loss tolerance in urban flooding. *Computers, Environment and Urban Systems* 80, 101443.
- Du, J., Qian, L., Rui, X., Zuo, T., Zheng, D., Xu, Y., Xu, C.Y., 2012. Assessing the effects of urbanization on annual runoff and flood events using an integrated hydrological modeling system for Qinhuai River basin, China. *Journal of Hydrology* 464, 127–139.
- El-Sawy, A., Loey, M., El-Bakry, H., 2016. CNN for handwritten Arabic digits recognition based on LeNet-5, in: *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*. Springer, pp. 566–575.
- Feng, Q., Liu, J., Gong, J., 2015. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—a case of Yuyao, China. *Water* 7, 1437–1455.
- Fohringer, J., Dransch, D., Kreibich, H., Schröter, K., 2015. Social media as an information source for rapid flood inundation mapping. *Natural Hazards and Earth System Sciences* 15, 2725–2738.
- Grahn, T., Wanders, N., 2020. Utilizing insurance data for flood loss estimation. *Environmental Research Letters* 15, 044002.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
- Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. *Nature Climate Change* 3, 816–821.
- Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*. 2 ed., John Wiley & Sons, New York.
- Khosravi, K., Panahi, M., Bui, D.T., 2020. Convolutional neural network approach for spatial prediction of flood hazard at national scale of Iran. *Journal of Hydrology* 591, 125552.

- Kia, M.B., Pirasteh, S., Pradhan, B., Gonbadi, A.R., Ahmadi, H., 2012. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences* 67, 251–264.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems*, pp. 1097–1105.
- Kundu, S., Olli, V., Vaka, D.S., Lakshmi, A., 2017. Flood depth estimation in the context of the 2015 Chennai flood event using Resourcesat-2 images. *Current Science* 113, 588.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 2278–2324.
- Li, L., Liu, Z., Zhang, H., 2019. Classification of flood-affected areas using deep learning. *Remote Sensing* 11, 1923.
- Liu, Y., Cao, X., Chen, J., 2021. Flood risk assessment under land use and climate change in Wuhan City of the Yangtze River Basin, China. *Sustainability* 13, 2196.
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., et al. (Eds.), 2021. *Climate Change 2021: The Physical Science Basis*. Cambridge University Press, Cambridge, UK. doi:10.1017/9781009157896.
- Mosavi, A., Ozturk, P., Chau, K.w., 2018. Flood prediction using machine learning models: Literature review. *Water* 10, 1536.
- Munich Re, 2022. Hurricanes, cold waves, tornadoes: Weather disasters in 2021 result in the second-highest insured losses ever. <https://www.munichre.com/en/company/media-relations/media-information-and-corporate-news/media-information/2022/natural-disaster-losses-2021.html>. Accessed: 2026-07-07.
- Pradhan, B., 2010. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *Journal of Spatial Hydrology* 9, 1–18.
- Rahman, M., Saha, N., Islam, M., 2021. Flood susceptibility assessment in Bangladesh using machine learning algorithms. *Journal of Environmental Management* 290, 113111.
- Resource and Environment Science and Data Center, 2023. 1:1,000,000 soil map of the people's republic of china (based on soil genetic classification system). <https://www.resdc.cn/data.aspx?DATAID=145>. Chinese Academy of Sciences. Accessed: 2026-07-07.
- Restrepo-Estrada, C., de Andrade, S.C., Abe, N., Fava, M.C., Mendiondo, E.M., Porto, A.G., 2018. Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. *Computers & Geosciences* 111, 148–158.

- Sanyal, J., Lu, X.X., 2004. Application of remote sensing in flood management with special reference to monsoon Asia: a review. *Natural Hazards* 33, 283–301.
- Shuster, W.D., Bonta, J., Thurston, H., Warnemuende, E., Smith, D.R., 2005. Impacts of impervious surface on watershed hydrology: A review. *Urban Water Journal* 2, 263–275.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sit, M.A., Demir, I., Xiang, Z., Ewing, G., Sermet, Y., 2020. A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology* 82, 2635–2670.
- Spekkers, M.H., Kok, H., Clemens, F.H.L.R., ten Veldhuis, J.A.E., 2014. Validating radar-based rainfall products using insurance claim data: a case study for the Netherlands. *Natural Hazards and Earth System Sciences Discussions* 2, 5287–5316.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9.
- Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology* 512, 332–343.
- Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F.W., Dutta, D., Kim, S., 2017. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software* 90, 201–216.
- United Nations Office for Disaster Risk Reduction, 2020. Human cost of disasters: An overview of the last 20 years (2000–2019). <https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019>. Accessed: 2026-07-07.
- Wang, Y., Hong, H., Chen, W., Li, S., Panahi, M., Khosravi, K., 2018. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of the Total Environment* 642, 1032–1049.
- Yin, J., Ye, M., Yin, Z., Xu, S., 2015. A review of advances in urban flood risk analysis over China. *Stochastic Environmental Research and Risk Assessment* 29, 1063–1070.
- Zhang, Y., Yang, P., Zeng, C., Wen, X., 2019. Physical vulnerability of pedestrians under the joint effect of wind and floodwater. *Journal of Hydrology* 578, 124036.
- Zhao, G., Pang, B., Xu, Z., Peng, D., Zuo, D., 2020. Urban flood susceptibility assessment based on convolutional neural networks. *Journal of Hydrology* 590, 125235.
- Zhou, Q., Zhang, H., Wang, L., 2021. Social sensing of urban waterlogging: A case study of the 2021 Henan heavy rain using Weibo data. *International Journal of Disaster Risk Reduction* 65, 102559.

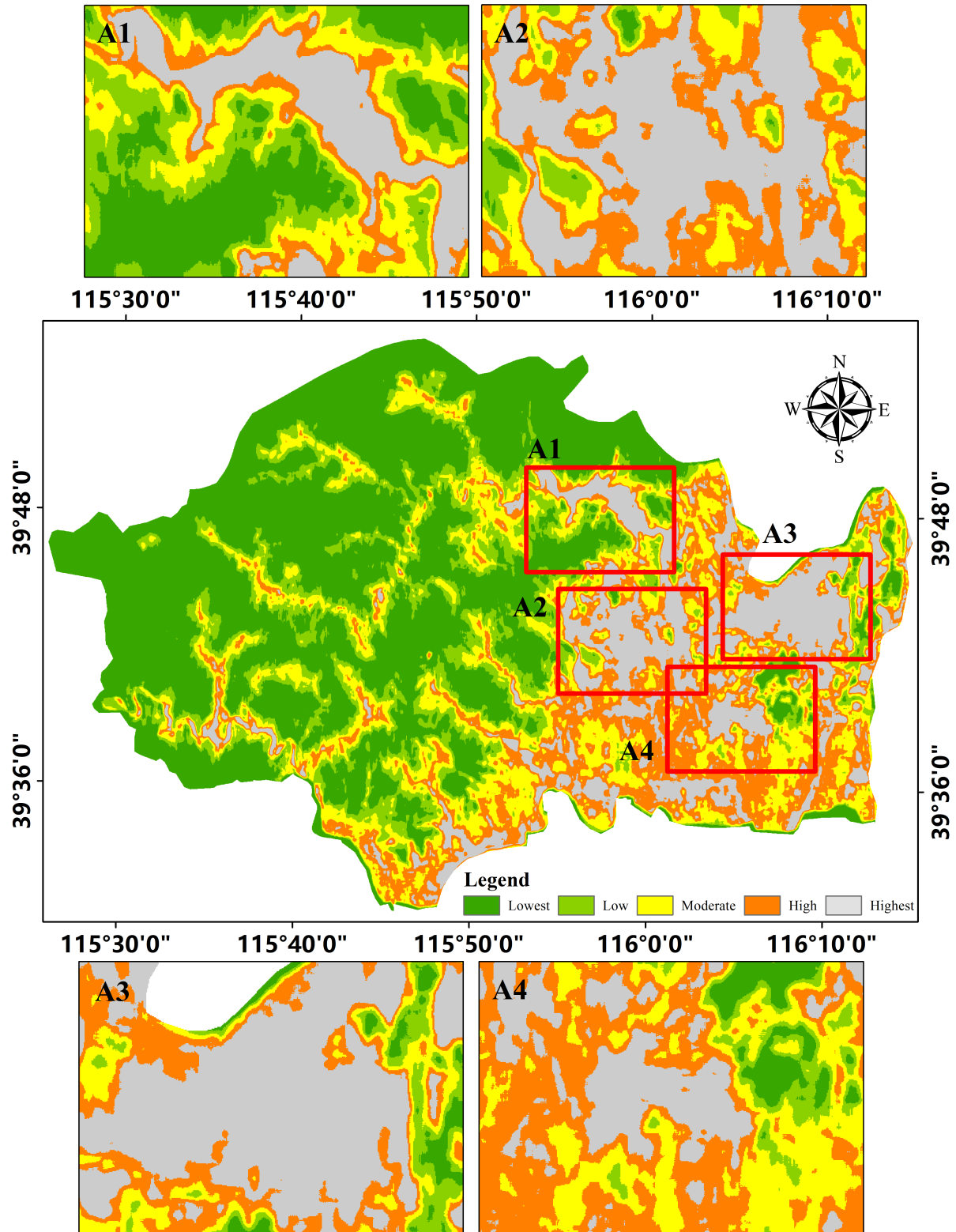


Figure 5: Flood Susceptibility Map Produced by Using the LeNet-Style Model

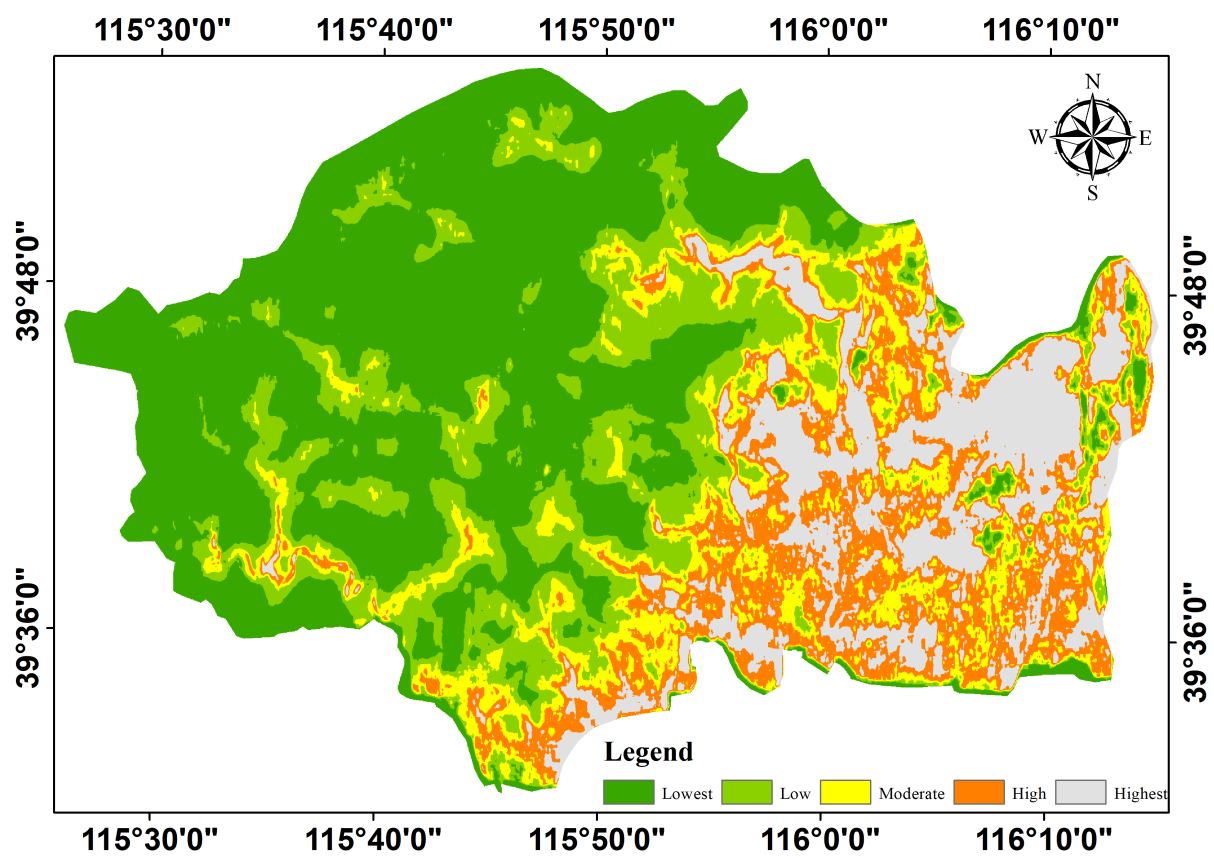
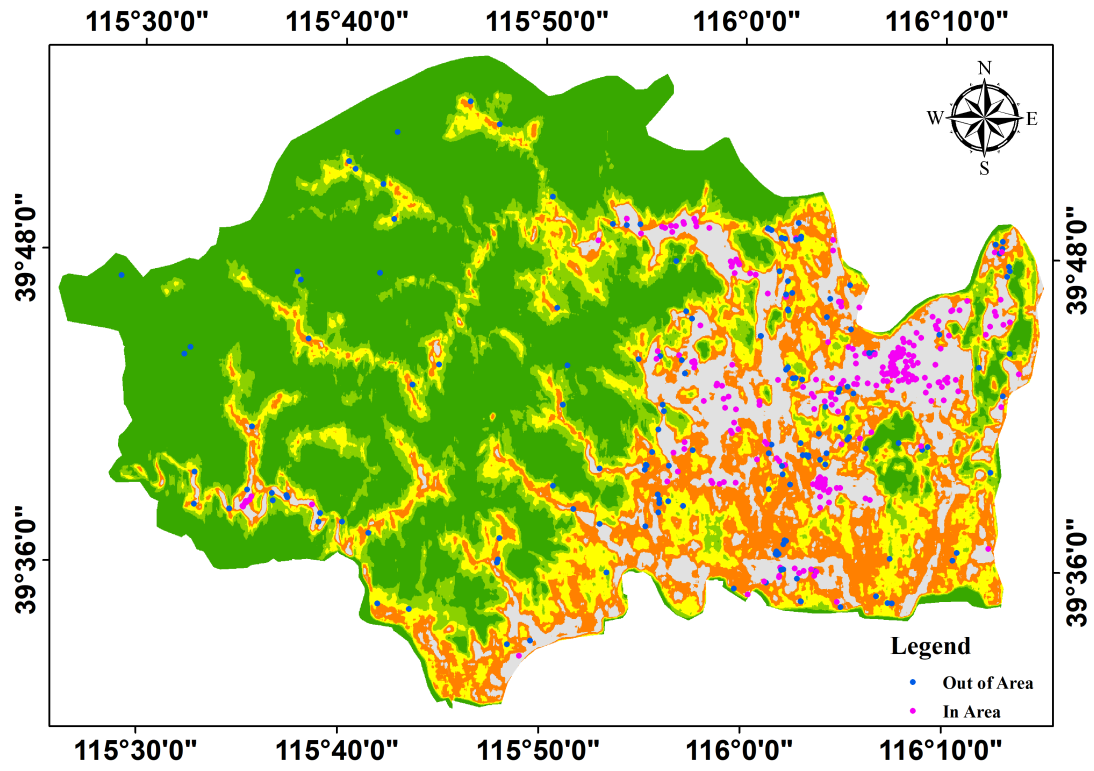
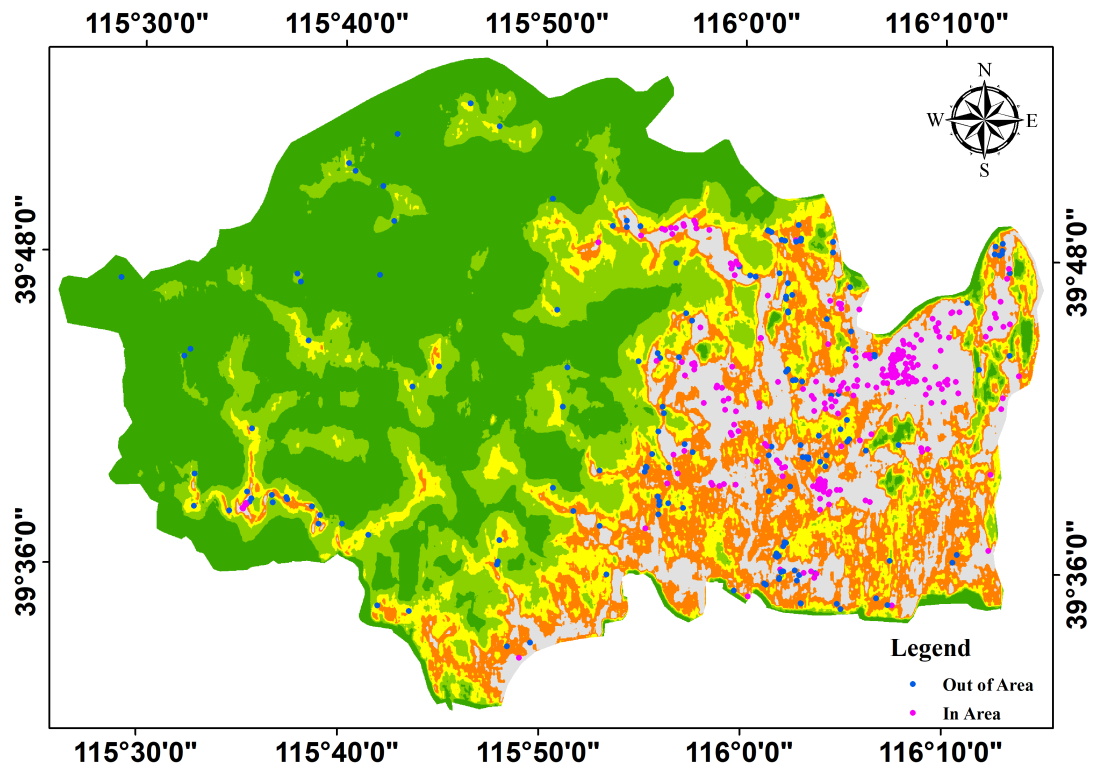


Figure 6: Flood Susceptibility Map Produced by the 9-factor Model (without DEM and Slope)



(a) Validation with 11 Factors



(b) Validation with 9 Factors

Figure 7: Comparison of flood susceptibility maps produced by 11 Factors and 9 Factors

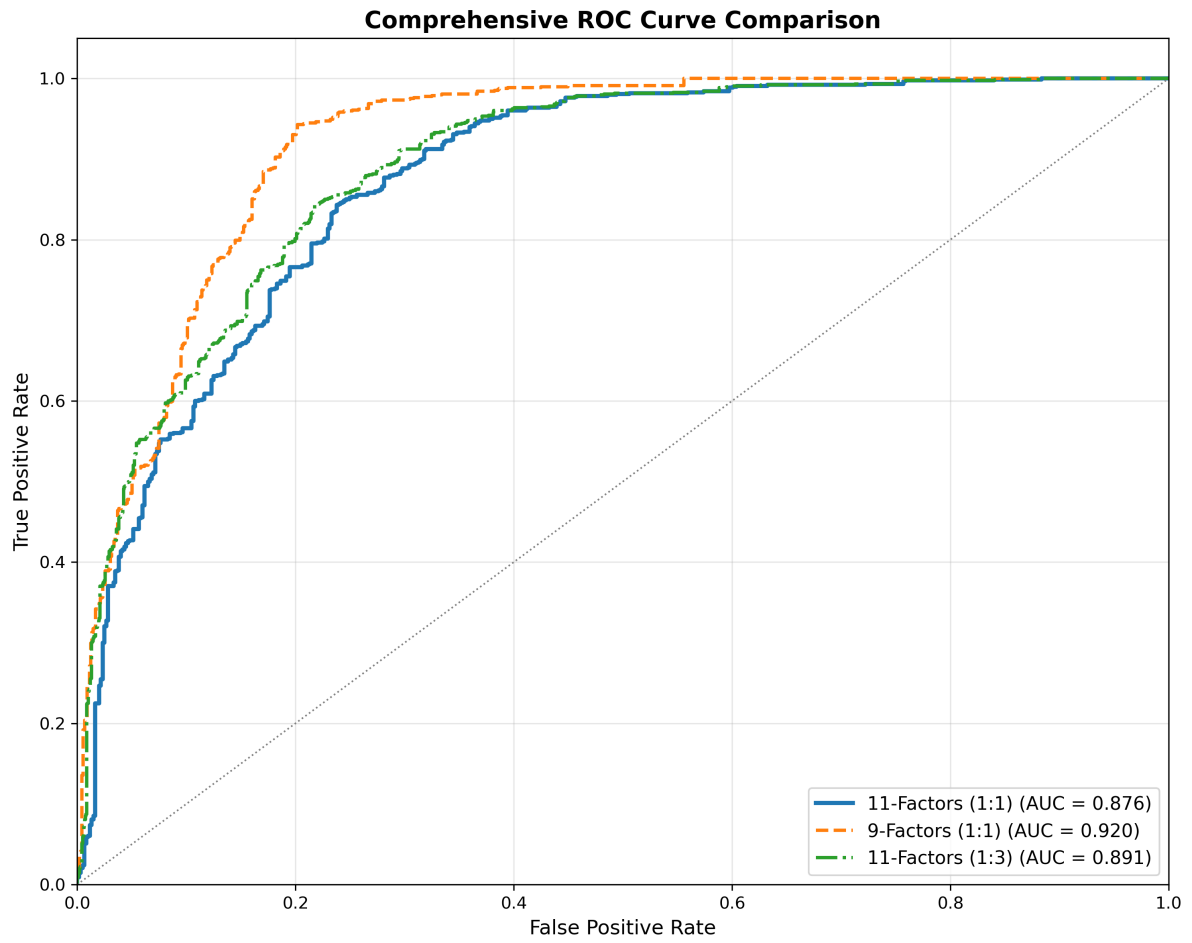


Figure 8: ROC Curves of the Three Experimental Settings. The plot illustrates the performance of the 11-factor model (Blue), the reduced 9-factor model (Orange), and the 11-factor model under 1:3 sampling (Green).