

Real-time prediction and ponding process early warning method at urban flood points based on different deep learning methods

Yihong Zhou

Zhengzhou University

Zening Wu

Zhengzhou University

Mengmeng Jiang

Zhengzhou University

Hongshi Xu

Zhengzhou University

Denghua Yan

Zhengzhou University

Huiliang Wang (✉ wanghuiliang@zzu.edu.cn)

Zhengzhou University <https://orcid.org/0000-0003-0057-455X>

Chentao He

Zhengzhou University

Xiangyang Zhang

Zhengzhou University

Research Article

Keywords: Urban flood, Ponding point, Deep learning, Prediction, Early warning

Posted Date: January 24th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1114192/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Real-time prediction and ponding process early warning method at urban flood points based on different deep learning methods

Yihong ZHOU^a, Zening WU^a, Mengmeng JIANG^a, Hongshi XU^a, Denghua YAN^a, Huiliang WANG^{a,*},

Chentao HE^a, Xiangyang ZHANG^a

^a College of Water Conservancy Engineering, Zhengzhou University, Zhengzhou, Henan, 450001, P. R. China

Abstract: Accurate prediction of urban floods is regarded as one of the critical means to prevent urban floods and reduce the losses caused by floods. However, due to the uncertainty and complexity of urban floods and waterlogging, it is very difficult to use simulation models to quickly and accurately predict and warn of urban floods. Therefore, it is necessary to develop new methods to support the rapid and accurate prediction of urban floods and waterlogging. In this study, a refined prediction and early warning method for urban flood and waterlogging processes based on deep learning methods is proposed from three aspects: research feasibility, method applicability, and method system. The spatial autocorrelation of rain and ponding points is analyzed by Moran's I. For each ponding point, the relationship model between the rainfall process and ponding process is constructed based on different deep learning methods, and the results are analyzed and verified by a statistical evaluation method. The results show that the gradient boosting decision tree algorithm has the highest accuracy and efficiency (with a root mean square error of 0.001) for ponding process prediction and is regarded as the most suitable method for ponding process prediction. Finally, the real-time prediction and early warning of urban floods and waterlogging processes driven by rainfall forecast data are realized, and the

*Corresponding author at: College of Water Conservancy Engineering, Zhengzhou University, Zhengzhou, Henan, 450001, P. R. China

Tel: +8618135671702

E-mail address: wanghuiliang@126.com

21 results are verified by the measured data. In addition, the results of the sensitivity analysis show
22 that the rainfall peak and location coefficient have the greatest impact on ponding.

23 **Key words:** Urban flood; Ponding point; Deep learning; Prediction; Early warning

24 **1 Introduction**

25 In recent years, under the background of global warming and urbanization (Masson et al., 2020;
26 Kenney and Janetos, 2020), urban flooding and waterlogging events have occurred frequently
27 worldwide and have become a serious challenge faced by many large cities worldwide
28 (Moghadas et al., 2019; Wang et al., 2020; Thanvisitthpon et al., 2020; Sharifi, 2020). For
29 example, in July 2019, heavy rains caused at least 18 deaths in Mumbai, India, and caused large-
30 scale traffic interruptions (Ke et al., 2020). In July 2021, the “7.20 rainstorms” in Zhengzhou,
31 China, caused 292 deaths and a direct economic loss of 53.2 billion yuan. With the aggravation
32 of climate change, the frequency and intensity of extreme rainfall events may increase in the
33 future, and cities may face a greater flood disaster threat (Schreider et al., 2000; Ntelekos et al.,
34 2010; Xu et al., 2018). Therefore, in the urban flooding and waterlogging process, how to
35 forecast the inundation process in a more accurate way for each ponding point to minimize the
36 life, property and economic losses caused by floods has become a scientific problem to be solved
37 urgently.

38 To prevent the occurrence of floods and implement timely countermeasures to reduce flood
39 losses, city authorities usually need to predict urban flooding and waterlogging (Ke et al., 2020).
40 Urban flood simulation models based on physical mechanisms are the most popular method for
41 predicting the urban flooding process and have attracted great attention from scientists (Mignot
42 et al., 2018). Many scholars have discussed urban floods and waterlogging simulations from
43 different angles (Babaei et al., 2018; Bermudez et al., 2018; Hou et al., 2020). However, due to
44 the lack of calibration and verification data (Schmitt et al., 2004) and the long running time of
45 the model (Suwit and Parinda., 2016; Hou et al., 2020), these physical models are limited to a

46 certain extent in the real-time urban flood prediction and early warning. In contrast, the deep
47 learning technology (or machine learning) can mine knowledge and laws that conventional data
48 processing methods cannot mine and identify highly complex nonlinear relationships between
49 characteristic and predictive variables (Panahi et al., 2020), and it does not need to understand
50 the potential physical process (Mosavi et al., 2019). Therefore, deep learning technology
51 provides a new idea for solving the problem of rapid and accurate urban floods prediction, which
52 is particularly useful in large- and medium-sized cities with sufficient data. In recent years, many
53 deep learning methods have been used in urban flood research. Several types of deep learning
54 methods, gradient boosting decision trees (GBDTs) (Wu et al., 2020; Rahebeh et al. 2020),
55 support vector machines (SVMs) (Nayak and Ghosh, 2013; Wang and Song, 2019) and neural
56 networks (Singh et al., 2010; Lei et al., 2021), have been successfully applied to urban flood
57 prediction. However, in urban flood forecasting research, not all in-depth learning methods are
58 suitable for urban flood research, and the prediction performance of each method may vary
59 greatly due to algorithms and specific prediction requirements. Therefore, it is necessary to
60 develop a relatively complete set of methods using deep learning methods to predict the urban
61 flood ponding process to improve the urban flood prediction method and guide urban flood
62 control.

63 Based on the aforementioned literature, this study analyzed the feasibility of the research
64 method, compared the applicability of different deep learning technologies in urban flood
65 prediction, and applied a deep learning method to propose a set method system of urban flood
66 ponding process prediction. The specific objectives of the study were to i) analyze the feasibility
67 of the urban flood ponding process prediction for each ponding point based on the spatial

68 autocorrelation method, ii) use GBDT, SVM and backpropagation neural network (BPNN) to
69 construct the relationship model between the rainfall process and the ponding process and use
70 statistical evaluation methods to analyze the accuracy of different deep learning methods, which
71 aims to propose a method suitable for predicting the urban flood ponding process, and iii)
72 construct the urban flood ponding process prediction and early warning method driven by rainfall
73 forecast data, combined with measured data to test the accuracy of the early warning results. The
74 novelty of this work lies in proposing a rapid and accurate prediction method system for the
75 urban flood ponding process based on the method feasibility, applicability of the method, and
76 technical system. The research results can provide technical references for urban flood early
77 warning and urban flood control.

78 **2 Study area and data processing**

79 **2.1 Study area**

80 Zhengzhou is the capital city of Henan Province and an important transportation hub in central
81 China. As shown in Fig. 1. It is located in northern central Henan Province, with high terrain in
82 the west and low terrain in the east. It has a temperate continental monsoon climate, with an
83 average annual precipitation of 639.5 mm. However, the precipitation distribution in Zhengzhou
84 is very uneven. The precipitation in the flood season (June to September) accounts for
85 approximately 60% of the annual precipitation. In response to urban floods and waterlogging,
86 Zhengzhou City has taken various measures in recent years, such as dredging rivers and cleaning
87 up sewers. However, the flooding and waterlogging phenomenon still occurs frequently, which
88 seriously threatens the safety of people's lives and property and the normal operation of the city.

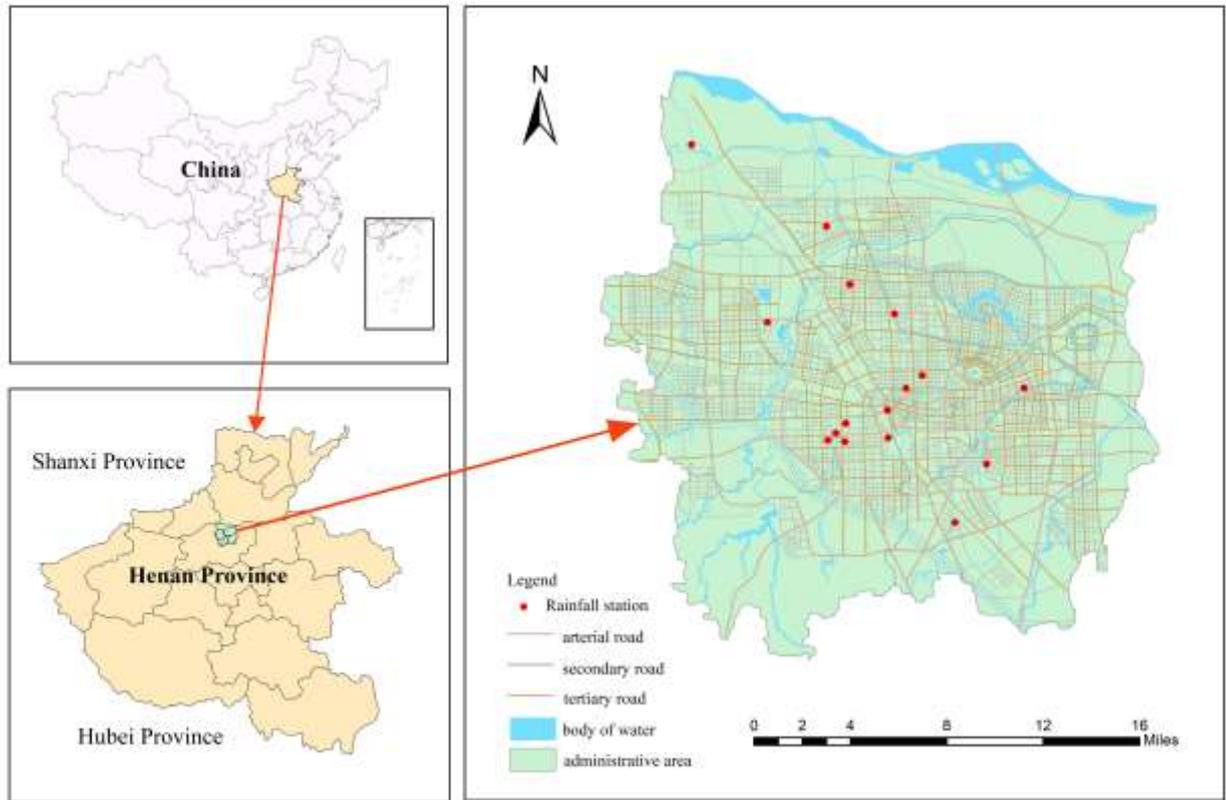


Fig. 1. The Location of study area

2.2 Data

2.2.1 Rainfall observation data

Rainfall observation data refer to the time-history distribution data of rainfall observed and recorded by 16 rainfall stations, which come from the Zhengzhou City Meteorological Department (Fig. 1). The time resolution of the data is 10 minutes. In this study, 21 historical rainfall data points from 2016 to 2018 were selected as the sample data of the model: June 4, 2016, June 23, 2016, July 19, 2016, August 4, 2016, August 6, 2016, August 25, 2016, September 12, 2016, May 22, 2017, June 5, 2017, June 22, 2017, July 18, 2017, August 12, 2017, August 25, 2017, August 30, 2017, May 15, 2018, July 4, 2018, July 13, 2018, July 27, 2018, August 10, 2018, September 15, 2018, and September 25, 2018. To obtain the rainfall spatial distribution data, the kriging interpolation method was used to perform interpolation analysis on

102 the rainfall time-history distribution data of each rainfall station to obtain the rainfall time-history
103 distribution data of each ponding point.

104 2.2.2 Ponding observation data

105 Ponding observation data refer to the ponding depth of ponding points in historical ponding
106 events. In this study, the ponding depth data of 48 typical ponding points were collected, which
107 were obtained from Zhengzhou Municipal Department. The time resolution of the data was 1
108 minute. The ponding depth data of these ponding points was from the ponding detection
109 equipment at each road intersection.

110 2.2.3 Rainfall forecast data

111 The rainfall forecast data in the next two hours were obtained by calling the API of Caiyun
112 technology. The rainfall forecast data have a temporal resolution of 1 minute and a spatial
113 resolution of 1 km. It should be noted that since the update period of the observational rainfall
114 data of the rainfall station is 10 minutes, the update period of rainfall forecast data is also taken
115 as 10 min to ensure the data comparison consistency. Therefore, in this study, a total of 6 updated
116 rainfall forecast data points on August 1, 2019, with forecast periods of 10 minutes, 20 minutes...,
117 60 minutes, were used as the sample forecast and early warning data, and each update of the
118 sample data included the rainfall forecast data of the 48 ponding points.

119 **2.3 Training dataset**

120 The training and testing process of sample data is the basis of constructing the prediction model
121 of the inundation process. To input the process data of these variables into the model, the data
122 processing method of equal distance splitting and reorganizing was used to process the sample
123 data, which is referred to in previous research (Wu et al., 2020). On this basis, calculate the

124 sensitivity index (rainfall, rainfall duration, rainfall peak, position coefficient, rain intensity
125 variance and peak multiplier) of each rainfall process after splitting and reorganizing. The
126 location coefficient is the ratio of the length of the time period from the beginning of rainfall
127 duration to the occurrence of the rainfall peak of the rainfall duration. Rainfall intensity variance
128 is the variance in rainfall intensity, which is a statistical index to describe the rainfall pattern. The
129 peak multiplier is the ratio of the rainfall during the peak period to the total rainfall.

130 The historical rainfall process data, historical ponding process data and rainfall forecast data
131 were processed by using the above sample data procedural processing method, and 35,424
132 sample data points were obtained, where the test samples are the rainfall process and ponding
133 process data of three randomly selected rainfall events (with different rainfall intensities and
134 durations), with a total of 2,592 test samples, the training samples are the rainfall process and
135 ponding process data of the remaining 18 rainfall events, with a total of 31,104 training samples,
136 and the rainfall forecast sample data were rainfall events on August 1, 2019, with a total of 1,728
137 sample data points.

138 **3 Methodology**

139 **3.1 Spatial autocorrelation analysis**

140 Spatial autocorrelation is proposed based on the first law of geography. The first law of
141 geography notes that things distributed in space are interrelated, and the closer the distance
142 between things, the closer the relationship between them (Tobler, 1970). Spatial autocorrelation
143 refers to the relationship between a certain attribute of geographical things distributed in a region
144 and the same attribute of other things (Gatrell, 1989). For example, if an attribute of adjacent
145 objects in space has a similar trend and value, the adjacent objects have a positive spatial

146 correlation; in contrast, if an attribute of an adjacent object in space has the opposite trend and
147 value, the adjacent object has a negative spatial correlation.

148 Rainfall is a phenomenon in which the water vapor in the atmosphere falls to the surface in
149 the form of liquid water after condensation. Therefore, the spatial distribution of rainfall usually
150 has strong spatial autocorrelation. The spatial autocorrelation of the water accumulation point
151 refers to whether the ponding process of each ponding point is independent; if the spatial
152 autocorrelation is high, the ponding process of other ponding points will affect the ponding
153 process of the adjacent water accumulation point. The reason may be that there is a strong
154 hydraulic connection between different water accumulation points. In contrast, the low spatial
155 autocorrelation of the ponding points indicates that the ponding point ponding process is
156 independent, and the ponding process of each ponding point is less affected by the ponding
157 process of the surrounding ponding points. In this paper, the spatial autocorrelation between
158 rainfall and ponding points is analyzed by a rainfall ponding process.

159 Spatial autocorrelation describes the spatial autocorrelation of statistical objects by
160 analyzing the position and attributes of objects (Overmars et al., 2003). Moran's I index and
161 gear's C ratio are commonly used statistical analysis indices, and Moran's I index is the most
162 intuitive and commonly used spatial autocorrelation analysis method. Therefore, this paper uses
163 Moran's I statistics to analyze the spatial autocorrelation of rainfall and ponding points.

164 Moran's I index characterizes the spatial distribution relationship of a certain attribute of
165 adjacent objects, with a value of $-1\sim 1$. A positive value indicates that the spatial distribution of a
166 certain attribute of the adjacent object has a positive spatial correlation, and a negative value
167 indicates that the spatial distribution of a certain attribute of adjacent objects has a negative

168 spatial correlation. The closer the value is to 0, the smaller the spatial autocorrelation of a certain
169 attribute between adjacent objects. A value of zero indicates that there is no spatial correlation.

170 The calculation formula of Moran's I index is as follows:

$$171 \quad I = \frac{n_a \sum_{i=1}^{n_a} \sum_{j=1}^{n_a} w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^{n_a} \sum_{j=1}^{n_a} w_{ij}) \sum_{i=1}^{n_a} (y_i - \bar{y})^2} \quad (1)$$

172 where n_a is the sample point or the number of grids, y_i or y_j represents the attribute value of the i
173 or j point area, and w_{ij} is the weight matrix that measures the relationship between the spatial
174 objects i and j .

175 Moran's I index assumes that the distribution of spatial objects is random and then tests
176 whether the hypothesis is tenable by Z score. It is generally believed that when the score of its
177 normal statistics (Z score) is greater than 1.96, the original hypothesis is rejected; that is, the
178 spatial objects have a significant positive correlation under 95% probability. In contrast, when
179 the Z value is less than 1.96, the original hypothesis is accepted; that is, the spatial objects have
180 no significant positive correlation at 95% probability, and the distribution of spatial objects is
181 random.

182 **3.2 Model construction of ponding process at ponding point**

183 In this study, the GBDT, SVM and BPNN algorithms are used to build the rainfall process and
184 ponding process relationship model. It should be noted that the prediction model of the ponding
185 process is built for each ponding point. For each ponding point, a ponding process prediction
186 model constructed by the GBDT, SVM and BPNN algorithms is required.

187 **3.2.1 GBDT**

188 GBDT is an integrated learning algorithm that combines decision trees and gradient boosting

189 algorithms (Friedman, 2001). The core idea of GBDT is iterative learning based on the residuals
190 predicted by the decision tree, and finally, the weak learners of each iteration are accumulated
191 and output. CART was selected as the base learner in the process of GBDT training because the
192 CART decision tree structure is simple, easy to understand, and robust. That is, in each iteration,
193 the gradient boosting algorithm is used to make the latter decision tree train the previous tree
194 residual along the direction of the maximum descending gradient, and finally, the classification
195 results of all trees are accumulated and output (Deng et al., 2019).

196 The number of weak learners, maximum depth and learning rate are the main parameters
197 of GBDT (Wu et al., 2020). The number of weak learners reflects the number of model
198 iterations, which increases with the increase in the number of iterations. Max depth refers to the
199 maximum depth of the decision tree, which is "none" by default. However, due to the large
200 quantity of data in this study, it is necessary to reasonably set the number of trees to prevent
201 overfitting. The learning rate is a parameter between 0 and 1 that represents the shrinkage step
202 in the update process. The above parameters are optimized by the grid search algorithm. A
203 complete mathematical and technical description of GARP model can be found in Friedman.,
204 2001, Wu et al., 2020.

205 3.2.2 SVM

206 The SVM algorithm is a supervised machine learning algorithm proposed by Vapnik and Cortes in
207 1995 (Cortes and Vapnik, 1995). It improves the generalization ability of the learning machine
208 by minimizing the empirical risk and structured risk so that when the number of samples is small,
209 a good statistical law can be obtained (Zhou et al., 2020). The core idea of SVM technology is to
210 use kernel function $K(X_i, X_j)$ to map the data to a high-dimensional or even infinite dimensional

211 feature space so that the data can be easily processed by a linear learning machine in the obtained
 212 feature space. On this basis, the optimal hyperplane can be found to solve the problem that the
 213 sample data cannot be fitted in the low dimensional space.

214 SVM can solve both classification and regression problems. In this paper, the regression
 215 method is used to predict the ponding process. Assume the training sample $D = \{(x_1, y_1), (x_2, y_2),$
 216 $(x_n, y_n)\}$, where y_i is the observation value of the objective function corresponding to x_i , and the
 217 ultimate goal of the regression support vector machine is to find the regression fitting function
 218 $f(x) = \omega\varphi(x) + b$, where $\varphi(x)$ is the mapping function, which is used to map the sample to a linearly
 219 separable high-dimensional space. Set the estimated value of the sample data to not less than ε .
 220 To find the optimal ω and b , it is transformed into the following optimal solution problem:

$$\begin{aligned}
 & \min \frac{1}{2} \omega^T \omega \\
 & \text{s.t.} \begin{cases} \|y_i - \omega x_i - b\| \leq \varepsilon \\ i = 1, 2, \dots, n \end{cases}
 \end{aligned} \tag{2}$$

222 SVM provides a variety of kernel functions, such as the linear kernel function, polynomial
 223 kernel function, radial basis kernel function and sigmoid kernel function. Among them, the radial
 224 basis kernel function is widely used and has higher efficiency for nonlinear data mapping (Xiao
 225 et al., 2019). Therefore, the radial basis function was used as the kernel function of the support
 226 vector machine in this study. The penalty parameter (C) and kernel function (K) are the key
 227 parameters to be optimized for SVM. In this study, a grid search algorithm was used to optimize
 228 C and K. A complete mathematical and technical description of SVM model can be found in
 229 Nayak and Ghosh, 2013, Wang and Song, 2019.

230 3.2.3 BPNN

231 A BPNN is a backpropagation neural network connected by multiple neurons, which can be

232 divided into an input layer, hidden layer and output layer. BPNN adopts a full interconnection
233 mode between layers, and there is no connection between neurons in the same layer. The BPNN
234 transmission is divided into the forward propagation stage and the backpropagation stage. In the
235 forward propagation stage, the signal starts to propagate from the input layer, and the deviation is
236 calculated in the output layer. If the deviation meets the requirements, the program is terminated.
237 If the deviation does not meet the requirements, the program enters the backpropagation stage. In
238 the backpropagation stage, the weight of each layer is modified by calculating the local gradient
239 of the network and then the forward propagation stage is entered again after the network is
240 reassigned. The program is terminated when the deviation meets the requirements. The grid
241 search algorithm was also used to optimize the main parameters of BPNN (learning rate, number
242 of hidden layers, number of nodes in hidden layers). A complete description of BPNN model can
243 be found in Jiang and Hong, 2013, Li et al., 2019.

244 **3.3 Model accuracy analysis**

245 The mean absolute error (MAE), root mean square error (RMSE), Nash efficiency coefficient
246 (NSE) and correlation coefficient (CC) are effective methods for evaluating the predictive ability
247 of each model attribute and the overall model predictive performance. MAE refers to the average
248 value of the absolute deviation between the predicted value and the actual value (Equation 3),
249 which reflects the average level of model prediction results. RMSE is a standard metric that
250 measures the deviation between the predicted value and the actual value (Equation 4). In addition,
251 RMSE is more sensitive to the large model prediction result error, which reflects the model
252 consistency level to a certain extent. CC refers to the correlation degree (Equation 5) between the
253 predicted results and the measured results, with a value of 0~1. It is generally believed that the

254 closer the CC is to 1, the better the prediction result consistency. NSE is a commonly used
 255 parameter to test the simulation quality of the hydrological model (Equation 6). The closer NSE
 256 (the value is negative infinity to 1) is to 1, the better the model quality is and the higher the
 257 model credibility. Therefore, MAE, RMSE, CC, and NSE were used to evaluate and compare the
 258 performance of the models in this study.

$$259 \quad MAE = \frac{\sum_{i=1}^n (x_i - y_i)}{n} \quad (3)$$

$$260 \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (4)$$

$$261 \quad CC = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

$$262 \quad NSE = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

263 where x_i is the predicted value of the sample, y_i is the measured value, \bar{x} is the mean value of
 264 the predicted sample, \bar{y} is the mean value of the measured sample, and n is the total number of
 265 samples.

266 Precision and recall were used to evaluate the performance of early warning results.
 267 Precision refers to the proportion of true positive samples in prediction samples (Faceli et al.,
 268 2011), which reflects how many prediction results are true (Equation 7). Recall (Equation 8)
 269 refers to the proportion of true positive samples in all positive samples (Faceli et al., 2011).

$$270 \quad \text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$271 \quad \text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

272 where TP is the number of samples correctly classified as positive (Faceli et al., 2011). TN is the

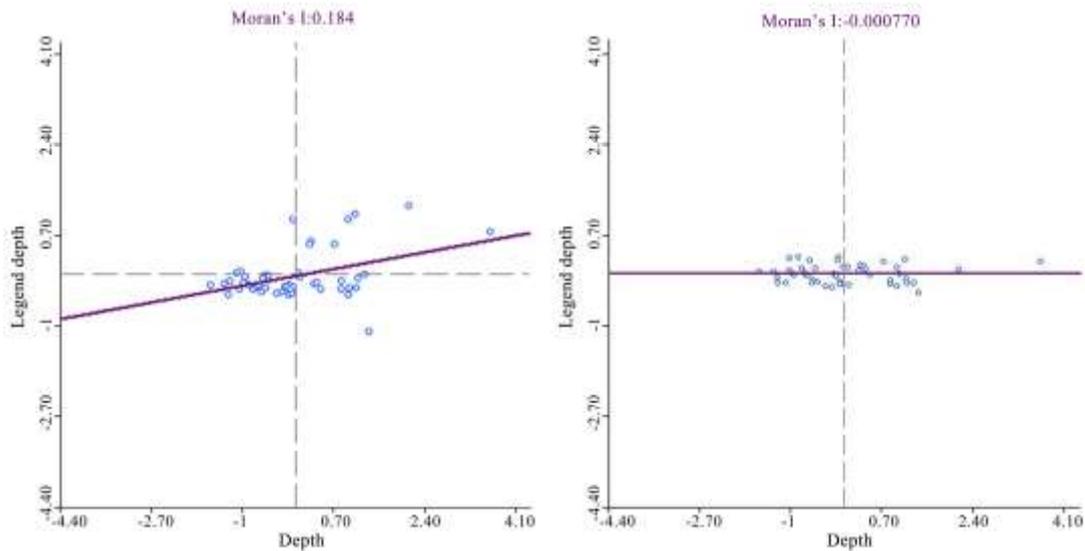
273 number of samples correctly classified as negative (Faceli et al., 2011). FP is the number of
274 samples incorrectly classified as positive because the right category is negative (Faceli et al.,
275 2011). FN is the number of samples incorrectly classified as negative because the right class is
276 positive (Faceli et al., 2011).

277 **4 Results**

278 **4.1 Spatial autocorrelation results**

279 The spatial statistical analysis software GeoDa developed by the University of Chicago Spatial
280 Data Center was used to analyze the spatial autocorrelation of rainfall and ponding points in this
281 paper. The spatial autocorrelation results of rainfall are shown in Fig. 2(a). The value of Moran's
282 I index is 0.184, and its normal statistic Z value is 4.23, which is greater than 1.96. Therefore, it
283 is believed that there is a significant positive spatial correlation of rainfall under 95% probability,
284 which verifies the spatial autocorrelation of rainfall.

285 Moran's scatter diagram of ponding is shown in Fig. 2 (b), the value of Moran's I is -
286 0.00077, which is close to 0, indicating that the spatial distribution of ponding has a low
287 correlation. In addition, the test score (Z value) of the normal statistic is 0.8648, which is less
288 than 1.96, indicating that the spatial distribution of ponding is irrelevant under 95% probability.
289 This result shows that there is no obvious hydraulic connection among the ponding points and
290 that they are independent of each other in space. Therefore, it is theoretically feasible to construct
291 a rainfall and ponding process relationship model for each ponding point.



(a) Rain (b) Ponding points

Fig. 2. Moran scatter diagram of rainfall and ponding

4.2 Analysis and comparison of prediction model results

In this study, Python 3.7 developed by Google was used to train and verify the model. In the model training process, the 18 randomly selected sample datasets are read into the model in CSV format by using the Python data import module, completing the training of the GBDT, SVM and BPNN models. In the verification phase, the sample datasets of the three randomly selected rainfall events are input into the model as verification data, and MAE, RMSE, CC, and NSE are used to evaluate and compare the performance of the model. As shown in Table 1, the mean absolute error of GBDT, SVM and BPNN for the ponding depth prediction are not greater than 0.03 m, and the CC between the prediction results and the measured results is greater than 0.97, which shows the effectiveness of these three algorithms in the ponding depth prediction to a certain extent. However, in comparison, the RMSE of the GBDT prediction model is significantly lower than that of SVM and BPNN, indicating that the GBDT model is more stable and more robust for predicting ponding depth. The stability of the model prediction results is

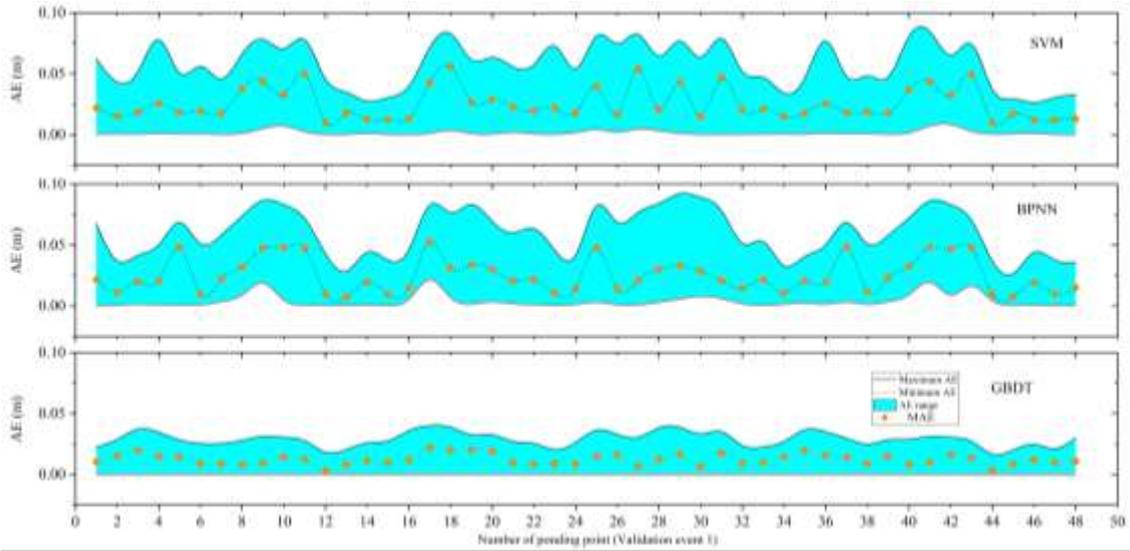
308 often very important for the prediction of urban flood depth. Therefore, from the perspective of
 309 the above statistical evaluation indicators, the GBDT prediction model is more suitable for urban
 310 flood depth prediction.

311 Table 1 The prediction performance of GBDT, SVM and BPNN

Category	MAE (m)	RMSE (m)	CC	NSE
SVM	0.0220	0.0427	0.9878	0.9538
BPNN	0.0204	0.0313	0.9937	0.9776
GBDT	0.0107	0.0207	0.9953	0.9880

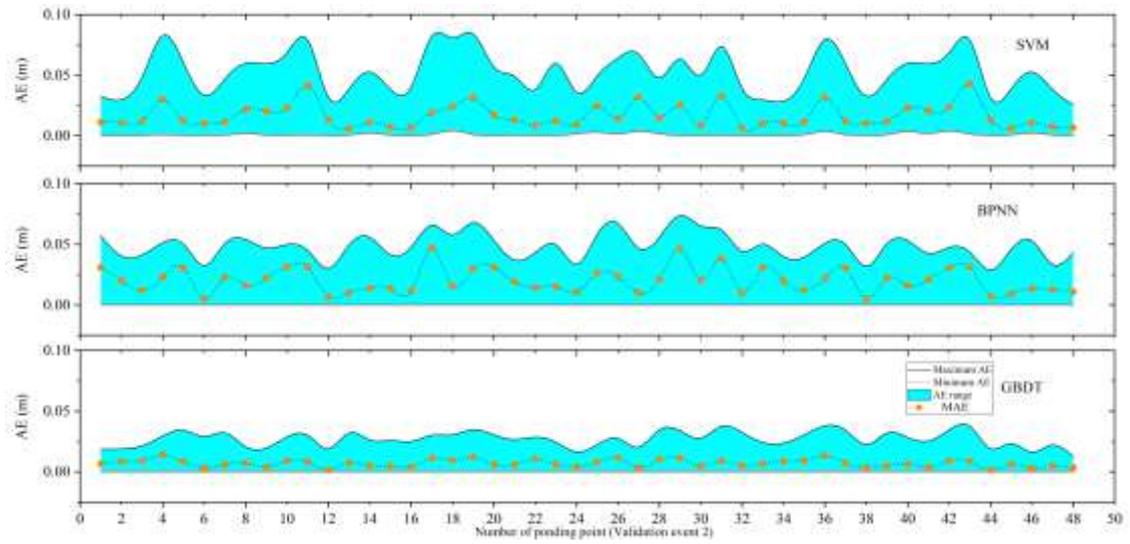
312 In addition, Fig. 3, Fig. 4 and Fig. 5 show the prediction performance of the SVM, BPNN
 313 and GBDT prediction models in different validation events and different ponding points. In the
 314 first verification event, the absolute error (AE) range and MAE of the GBDT (0~0.0405 mm and
 315 0.0122 mm) prediction model are significantly lower than those of the SVM (with
 316 0.0001~0.0922 mm AE and 0.025 mm MAE) and BPNN (with 0.0014~0.0973 mm AE and 0.024
 317 mm MAE). Similarly, in the second verification event, the AE range and MAE of the GBDT
 318 prediction model are 0~0.0450 mm and 0.0090 mm, respectively, which are also lower than those
 319 of SVM (with 0.0001~0.0992 mm AE and 0.0166 mm MAE) and BPNN (with 0~0.0823 mm AE
 320 and 0.0203 mm MAE). Similarly, the AE range and MAE of the GBDT prediction model
 321 (0~0.4667 mm and 0.0110 mm) are lower than those of SVM (with 0.0001~0.0988 mm AE and
 322 0.0248 mm MAE) and BPNN (with 0.0001~0.0912 mm AE and 0.0218 mm MAE). Therefore, it
 323 is not difficult to find that the performance of the GBDT prediction model is superior to that of
 324 SVM and BPNN in different validation events. At different ponding points, as shown in Fig. 3,
 325 Fig. 4 and Fig. 5, the AE range and MAE of the prediction results are quite different. This
 326 phenomenon is a common feature of SVM, BPNN and GBDT models. However, the MAE of the

327 GBDT prediction model at different ponding points displays a lower fluctuation degree than
328 SVM and BPNN, which demonstrates that the prediction performance of the GBDT model is
329 obviously more stable.



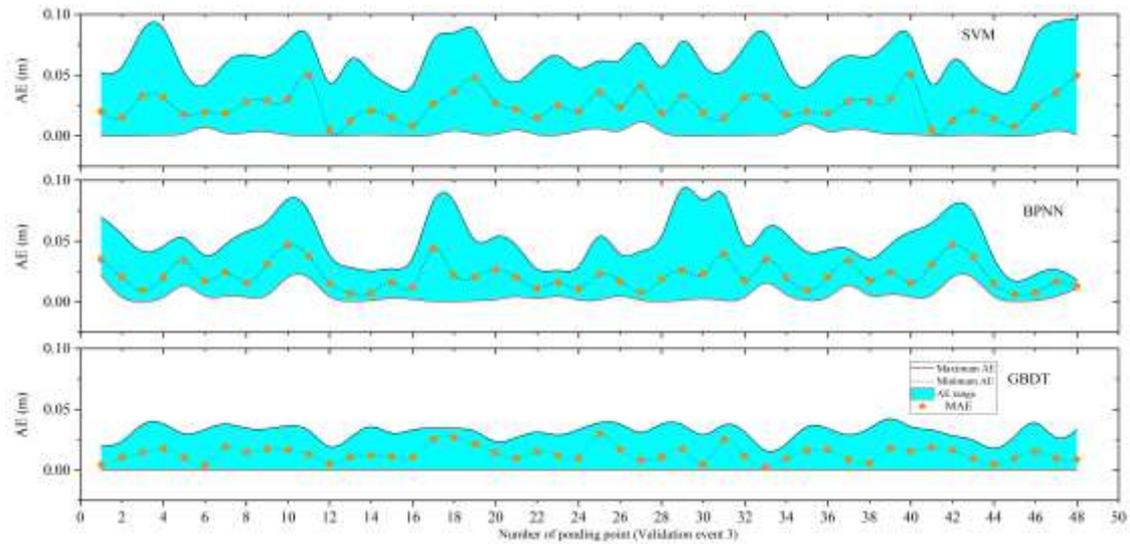
330

331 Fig. 3. AE of prediction results of different ponding points in the first verification event



332

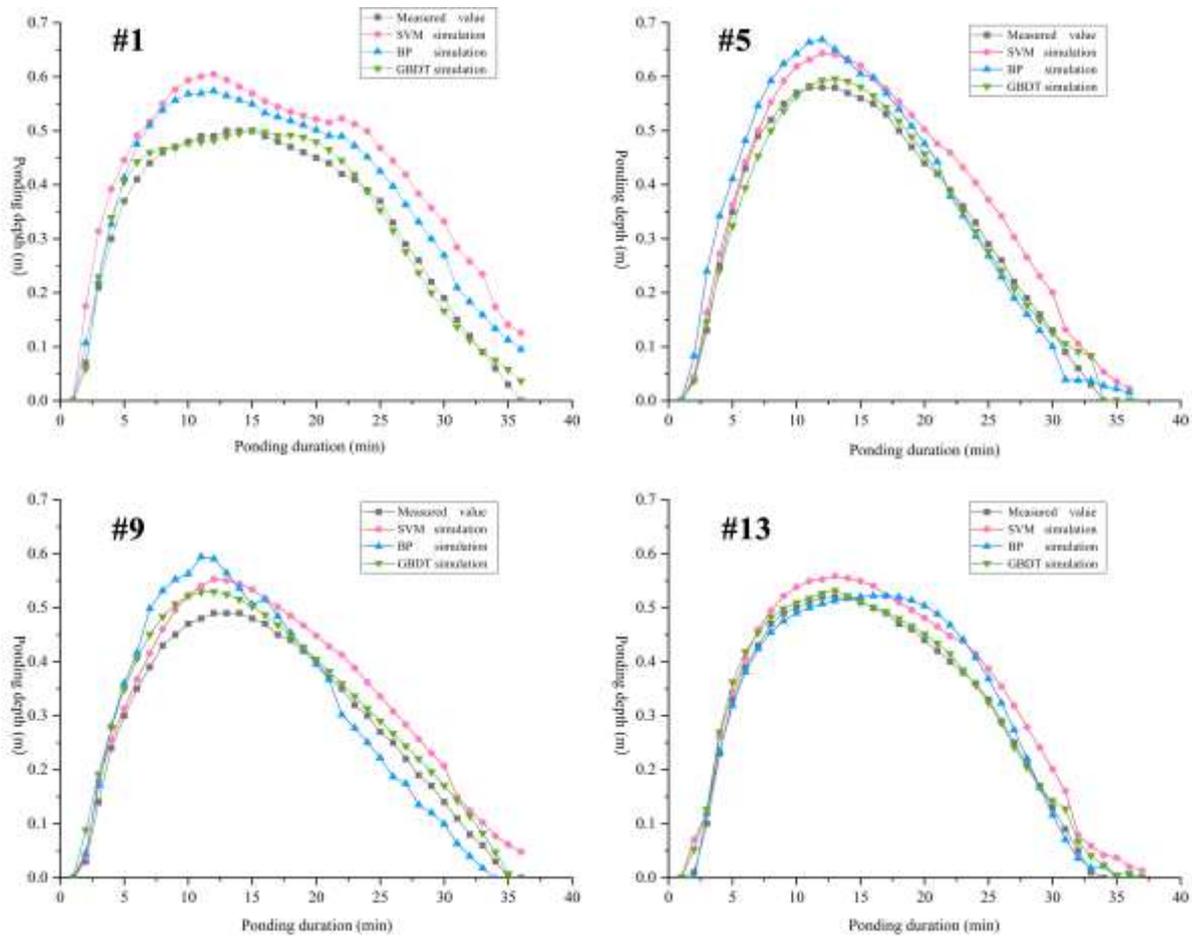
333 Fig. 4. AE of prediction results of different ponding points in the second verification event



334
 335 Fig. 5. AE of prediction results of different ponding points in the third verification event

336 To more clearly compare the difference between GBDT, SVM and random forest in the
 337 prediction of ponding depth, the prediction results of the ponding depth of four ponding points (#
 338 1, # 5, # 9, # 13) in a rainfall event were extracted by a random sampling method. As shown in
 339 Fig. 6, the ponding hydrograph at ponding point No. 1 was relatively flat, and the ponding depth
 340 fluctuated slightly from 8 to 20 minutes, which was obviously different from the other three
 341 ponding points. The reason for this phenomenon may be that the catchment area of ponding point
 342 No. 1 is larger than that of the other three ponding points. After rainwater reaches the surface,
 343 rainwater near the ponding point quickly collects at the ponding point to form ponds. However,
 344 the rainwater far away from the ponding point takes longer to converge to the ponding point after
 345 reaching the surface, which causes the ponding point to still have considerable rainwater
 346 converging to the ponding point after the ponding reaches the peak value, extending the duration
 347 of the ponding peak value. The other three ponding points have similar ponding formation and
 348 dissipation processes, and the ponding subsides soon after reaching the maximum depth. It is
 349 worth noting that the prediction results of the GBDT model for the ponding peak are very close
 350 (with 0.009 MAE) to the measured results, and the prediction accuracy is significantly better

351 than that of the SVM and BPNN models. This result shows that GBDT has obvious advantages
352 in predicting the ponding peak.



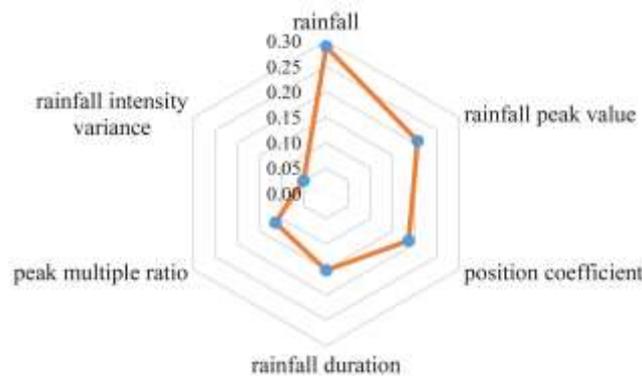
353
354 Fig. 6. Prediction results of ponding process (part)

355 4.3 Feature importance analysis

356 The information gain ratio (IGR) is an effective feature importance selection method. The larger
357 the IGR of the condition factor is, the higher the influence of the condition factor on the
358 prediction ability of the model. Therefore, in this study, the feature importance of each feature
359 variable (rainfall, rainfall duration, rainfall peak, position coefficient, peak multiplier ratio, and
360 rain intensity variance) on the influence of ponding depth was calculated based on the IGR.

361 As shown in Fig. 7, the result of feature selection shows that rainfall (0.2893) is the most
362 important factor affecting the ponding depth, and their cumulative contribution is more than 80%.

363 The reason is that rainfall is the driving factor of ponding, and only sufficient rainfall will
 364 produce ponding. In addition, the rainfall peak value (0.2070) and position coefficient (0.1864)
 365 are also important factors affecting the ponding depth. The reason is that the peak rainfall
 366 determines the size of the rainfall events and directly affects ponding severity. The position
 367 coefficient determines the rainfall pattern characteristics, and different types of rainfall patterns
 368 also have different degrees of impact on ponding depth. In contrast, the peak multiple ratio
 369 (0.1148) and rainfall intensity variance (0.0511) have little influence on the ponding prediction,
 370 indicating that the small rainfall pattern fluctuation has little or even negligible influence on
 371 ponding. These results indicate that the ponding depth is more sensitive to rainfall, rainfall peaks
 372 and position coefficients. In urban flood forecasting and early warning, we should focus on
 373 heavy rainfall events.



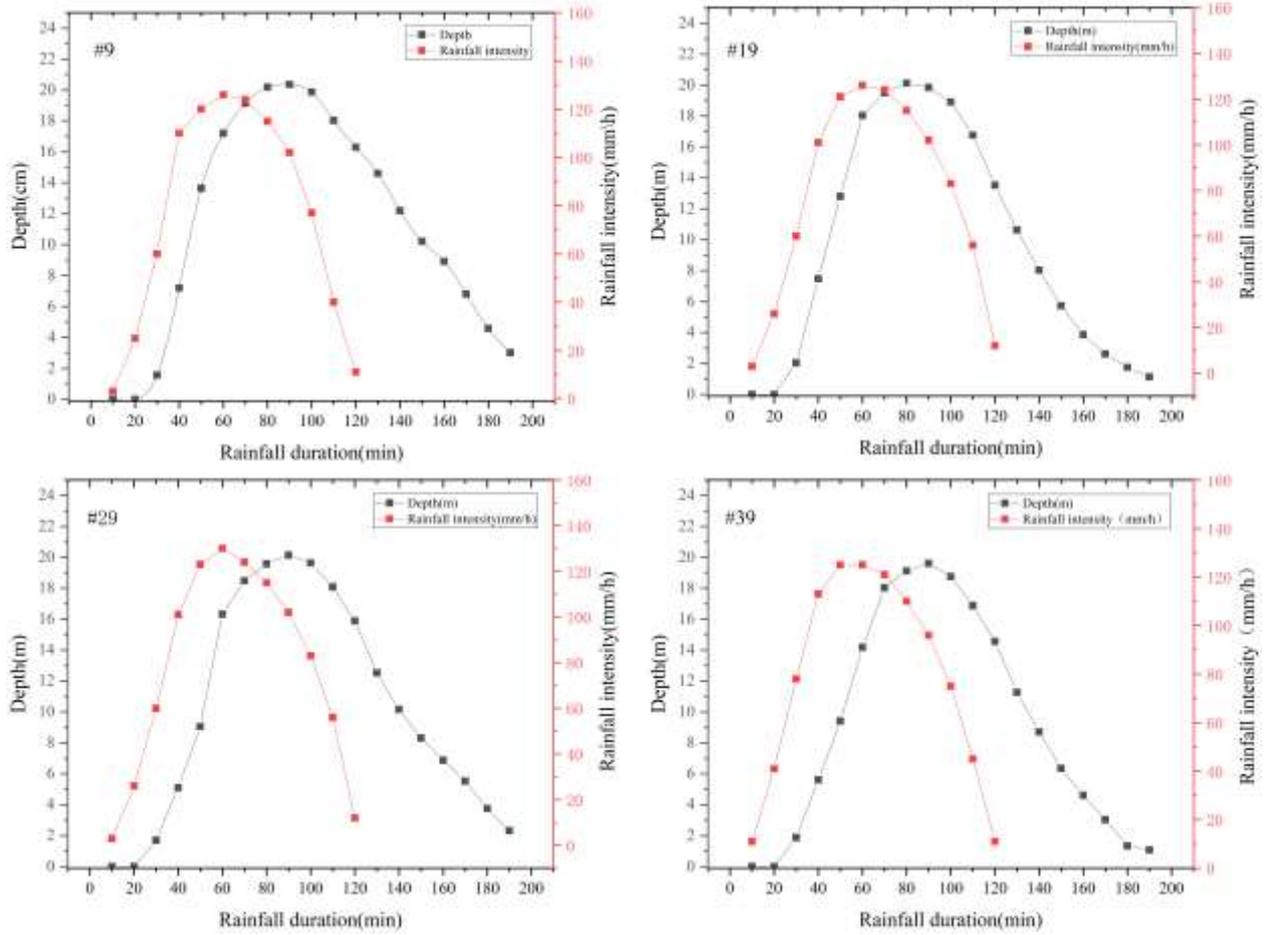
374
 375 Fig.7. Feature importance analysis results

376 **4.4 Prediction and early warning of the ponding process driven by rainfall forecast data**

377 It can be seen in 4.2 that the prediction effect accuracy and stability of the GBDT prediction
 378 model are better than SVM and BPNN. Therefore, the GBDT prediction model was used to
 379 predict ponding depth in this study. The rainfall forecast data are the basis for ponding depth
 380 prediction. In this study, the rainfall event in Zhengzhou city on August 1, 2019, was taken as the

381 forecast event. By calling the API of Caiyun technology, the rainfall forecast data of the 48
382 ponding points in this event were obtained 60 minutes in advance. The ponding process data of
383 48 ponding points were obtained by inputting the sensitivity index of rainfall prediction data into
384 the GBDT prediction model. On this basis, four ponding points (# 9, # 19, # 29, # 39) were
385 selected by the equidistant sampling method to draw the ponding process curve. As shown in Fig.
386 8, there is a slight hysteresis between the ponding process curve and the rainfall process curve of
387 these ponding points. The reason is that when the rainfall reaches the surface, it flows into the
388 pipe network. In the beginning, the rainwater does not exceed the pipeline drainage capacity, so
389 it will not form ponds. However, when the rainwater collection speed exceeds the pipe network
390 drainage capacity, the drainage pipe network will fill with rainwater, which will form
391 accumulated water near the rainwater collection center. Therefore, the ponding hydrograph at the
392 ponding point has a slight lag compared with the rainfall hydrograph.

393 To more intuitively display the ponding depth prediction results of ponding points, the
394 ponding depth and ponding duration early warning classification method (Table 2) is used to
395 classify the ponding depth of ponding points. The reason is that the early warning classification
396 method based on ponding depth and ponding duration can not only directly reflect the ponding
397 depth but also consider the continuous impact of long-term ponding.



398
 399 Fig. 8. Prediction results of ponding process at ponding points under the prediction period of
 400 60min in advance (part)

401 Table 2 Classification standard of ponding early warning level

Early warning level	Ponding depth (cm)	Ponding duration (min)	The degree of ponding
1	0~3	/	No ponding
2	3~10	>10	Mild ponding
3	10~25	>3	Moderate ponding
4	>25	/	Severe ponding

402 By inputting the rainfall forecast data of 60 min, 50 min, 40 min..., 10 min before rainfall
 403 into the model, the ponding process data of 48 ponding points at 60 min, 50 min, 40 min..., 10
 404 min before rainfall are obtained. Based on this, the maximum depth and ponding duration are
 405 selected as early warning indicators, and the spatial processing and display function of GIS are

406 used as early warning technology to draw the real-time correction of ponding point early warning
407 results. As shown in Fig. 9, the areas with serious ponding are mainly concentrated in the central
408 and eastern parts of Zhengzhou city, especially in Longhai Road, Hanghai Road and the old
409 urban area in the central area. The reason is that there are many underpass tunnels on Haihang
410 Road and Longhai Road. These underpass tunnels often have the characteristics of low elevation,
411 large terrain slope and wide catchment area. After rainwater reaches the surface, it quickly
412 collects and flows to form ponds at the bottom of the tunnel. Therefore, ponding points with
413 serious ponding mostly appear near these underpass tunnels. In contrast, the degree of ponding in
414 the northern part of Zhengzhou city is obviously lower. The reason may be that the northern part
415 of Zhengzhou is adjacent to the Yellow River wetland, which has a relatively flat terrain and a
416 large proportion of the surface permeable area. Therefore, the amount of infiltration and
417 interception of rainwater after reaching the ground is large, resulting in low confluence and
418 confluence velocities.

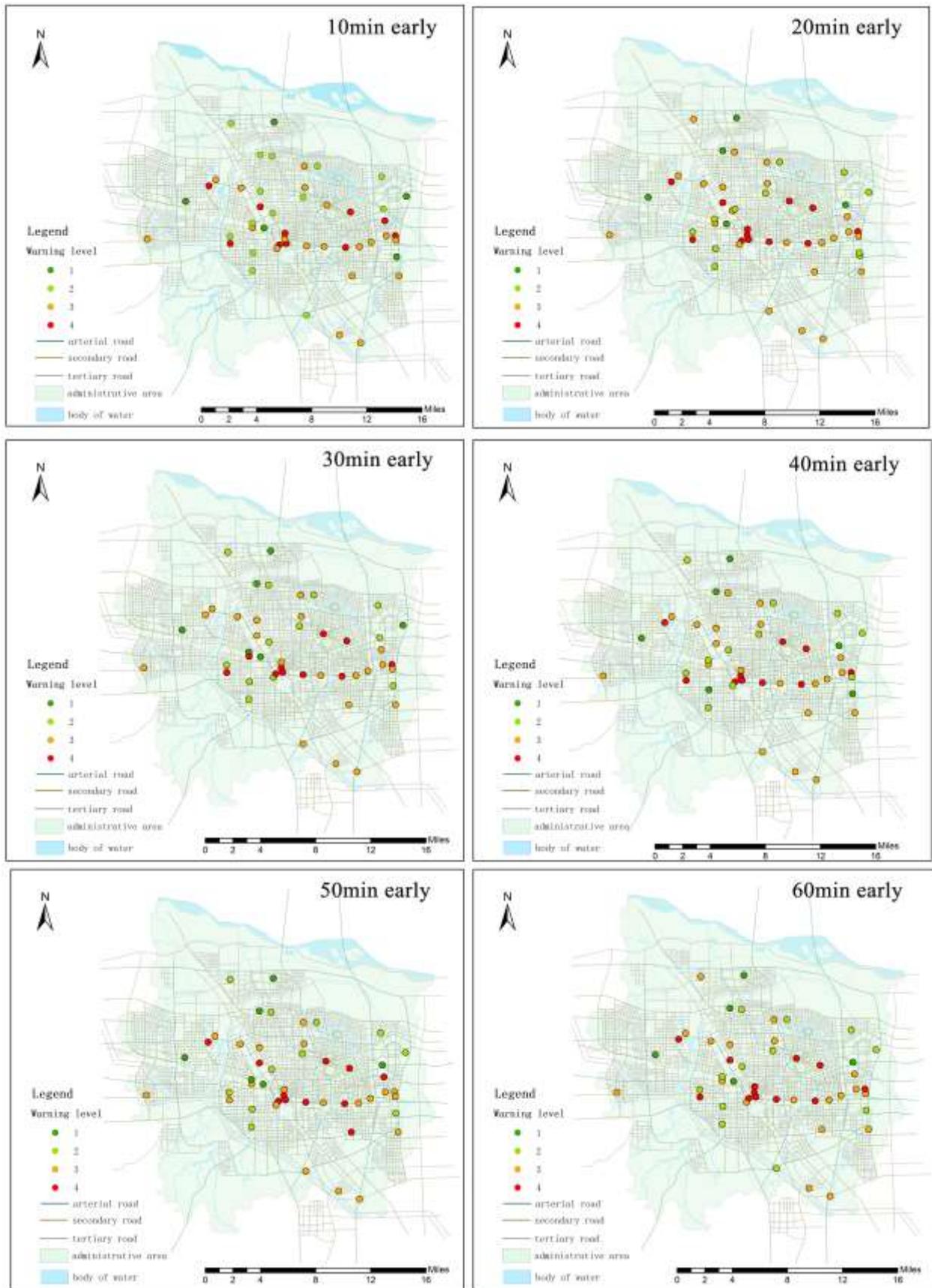
419 To further verify the effectiveness of real-time early warning results of ponding points using
420 rainfall forecast data, the ponding depth of each rainfall ponding point on August 1, 2019, was
421 obtained by means of ponding monitoring equipment, electronic water gauges and actual
422 measurements, and the early warning results and measured results of ponding points were
423 analyzed by the accuracy evaluation index of precision and recall. As shown in Table 3, the
424 overall precision and recall of real-time ponding point early accuracy results is more than 80%,
425 indicating that the overall early warning result precision can meet the requirements of urban
426 flood forecasting and early warning. Moreover, the prediction precision shows an overall trend of
427 improvement, although the prediction precision fluctuates with the shortening of the forecast

428 period. For example, the prediction precision in the next 10 minutes is 14.7% higher than that in
 429 the next 60 minutes. Moreover, from the perspective of the local accuracy of the early warning
 430 results, the prediction precision of ponding level 4 (i.e., serious ponding) is obviously higher than
 431 that of other levels in the early warning results of different ponding point levels (Table 3), which
 432 indicates that the GBDT algorithm has obvious advantages for predicting more serious ponding.

433 Table 3 Statistics of early warning results of ponding points

Ponding level	1		2		3		4		Precision	Recall
									(%)	(%)
	TP	FP	TP	FP	TP	FP	TP	FP		
Measured	5		13		19		11			
10min early	5	0	12	0	19	0	11	1	97.9	98.1
20min early	5	1	10	1	19	1	10	1	90.0	92.0
30min early	5	1	10	2	17	3	10	0	87.9	89.3
40min early	5	1	10	2	17	3	9	1	85.4	87.1
50min early	4	1	9	1	19	2	10	2	86.0	85.0
60min early	4	1	11	3	16	3	9	1	83.2	82.7
Precision (%)	85.0		87.7		89.9		90.9			

434



435

436

Fig. 9. Early warning level map of ponding point (10~60min early)

437 **5. Discussion**

438 As seen in Fig. 2 and Fig. 9, although the ponding points do not have spatial correlation, the
439 ponding points with the same ponding grade still have certain spatial aggregation. Specifically,
440 ponding in the central region is more serious, which is mainly due to the combined influence of
441 the spatial correlation of rainfall (Wang et al., 2020), the urban heat island effect and the
442 distribution (Min et al., 2018) of urban functional areas. The urban heat island effect makes the
443 rainfall in the center of the city larger. The city center is a densely populated old city with lower
444 pipe network design standards, aging and severely damaged pipe networks and higher
445 impervious areas. After rainwater reaches the surface, there is less infiltration, faster water
446 collection speed, and limited drainage capacity, which easily causes more serious ponding in the
447 central part of the city. In addition, Fig. 7 shows that urban ponding is most sensitive to rainfall,
448 rainfall peaks and location coefficients. Therefore, to reduce losses caused by floods, urban
449 management departments should not only increase the area of permeable ground in the central
450 region of urban areas and transform, repair and dredge the drainage pipe network but also take
451 timely measures such as drainage and cutting off roads when dealing with heavy rainfall.

452 In terms of the early warning results accuracy, as shown in Table 3, with the shortening of
453 the forecast period, the prediction accuracy shows an overall improvement trend. The main
454 reason is that rainfall forecast data accuracy gradually improves, and the accuracy of input
455 variables improves with the shortening of the forecast period, which leads to the improvement of
456 prediction variable accuracy. It is generally believed that there is a contradiction between the
457 forecast period and the prediction accuracy. The shortening of the forecast period will result in
458 higher prediction accuracy, and a longer forecast period will reduce the prediction accuracy.

459 Fortunately, the prediction precision and recall of the model proposed in this study still exceed
460 80% when the forecast period is 60 minutes. It effectively guarantees a certain forecast accuracy
461 while obtaining a longer forecast period, which effectively addresses one of the contradictions
462 between the forecast period and the prediction accuracy. The research results can provide more
463 guiding theoretical and technical references for improving prediction and early warning methods
464 and preventing flood disasters.

465 **6. Conclusion**

466 In this study, constructing ponding process prediction and early warning methods for ponding
467 points is systematically explained in three aspects: method feasibility, method comparison and
468 screening, and practical application of methods. The conclusions are as follows:

469 (1) The feasibility of this study is explained by using the spatial autocorrelation analysis
470 method, that is, each ponding point is spatially independent of each other, and there is no spatial
471 correlation. Therefore, the urban flood inundation process model based on ponding points is
472 theoretically feasible.

473 (2) Based on different deep learning methods (GBDT, SVM, and BPNN), the relationship
474 model between the rainfall process and the ponding process is constructed for each ponding point.
475 A statistical evaluation method was used to analyze the applicability of different deep learning
476 methods in ponding depth prediction. The results show that the GBDT algorithm has the highest
477 accuracy for ponding depth prediction, which indicates that it is the most suitable method for the
478 prediction of ponding depth.

479 (3) Taking the rainfall event in Zhengzhou on August 1, 2019, as an example, a refined real-
480 time prediction model for the urban flood ponding process driven by rainfall forecast data was

481 constructed. The real-time early warning for ponding points is realized by the warning level
482 standard and GIS. The results show that the overall accuracy of early warning results is more
483 than 80%, and the accuracy of early warning results shows an upward trend with the shortening
484 of the prediction period, which can meet the urban flood control requirements.

485 However, due to limitations in rainfall and ponding data, the ponding process prediction
486 model can only be built for ponding points with detailed data in this study. With the gradual
487 enrichment of data and advancement of physical simulation technology, future research can
488 attempt to expand the research scope combined with physical models.

489 **Authors contributions** Yihong Zhou: Methodology, Validation, Writing - original draft,
490 Writing - review & editing. Zening Wu: Writing - review & editing, Funding acquisition.
491 Mengmeng Jiang: Validation. Hongshi Xu: Methodology. Denghua Yan: Supervision. Huiliang
492 Wang: Methodology, Funding acquisition. Chentao He and Xiangyang Zhang: Data curation.

493 **Funding** The work was funded by the Key Project of National Natural Science Foundation of
494 China (No: 51739009), Excellent Youth Fund of Henan Province of China (212300410088),
495 Science and Technology Innovation Talents Project of Henan Education Department of China
496 (1HASTIT011) and Young backbone Teachers Training Fund of Henan Education Department of
497 China (2020GGJS005).

498 **Availability of Data and Materials** Data can be made available through contacting the
499 corresponding author.

500 **Declarations**

501 Ethical Approval Not applicable.

502 Consent to Participate The authors agree to participate in the journal.

503 Consent to Publish The authors providing consent publication of the manuscript to the journal
504 publisher.

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

507 **References**

508 Babaei S, Ghazavi R, Erfanian M (2018) Urban flood simulation and prioritization of critical
509 urban sub-catchments using swmm model and promethee ii approach. *Phys Chem Earth*
510 105:3-11. <https://doi.org/10.1016/j.pce.2018.02.002>

511 Bermudez M, Ntegeka V, Wolfs V, Willens P (2018) Development and comparison of two fast
512 surrogate models for urban pluvial flood simulations. *Water Resour Manag* 32(8):1-15.
513 <https://doi.org/10.1007/s11269-018-1959-8>

514 Carreau J, Guinot V (2020) A pca spatial pattern based artificial neural network downscaling
515 model for urban flood hazard assessment - sciencedirect. *Adv Water Resour* 147:103821.
516 <https://doi.org/10.1016/j.advwatres.2020.103821>

517 Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273-297.
518 <https://doi.org/10.1023/A:1022627411411>

519 Deng SK, Wang CG, Wang MY, Sun Z (2019) A gradient boosting decision tree approach for
520 insider trading identification: An empirical model evaluation of China stock market. *Appl*
521 *Soft Comput* 83:105652. <https://doi.org/10.1016/j.asoc.2019.105652>

522 Faceli K, Lorena AC, Gama JA, Carvalho A et al (2011) *Inteligência Artificial: Uma abordagem*
523 *de aprendizado de máquina*. 2:192

524 Friedman JH (2001) Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat*

525 29:1189-1232. <https://doi.org/10.1214/aos/1013203451>

526 Gatrell AC (1989) Book reviews: Griffith, D.A. 1987: Spatial autocorrelation: a primer.
527 Pennsylvania: Association of American Geographers. iv + 82 pp. US \$6.00. Prog. Hum
528 Geog 13:144-145. <https://doi.org/10.1177/030913258901300115>

529 Hou JM, Wang N, Guo KH, Li DL, Jing HX, Wang T, Hinkelmann R (2020) Effects of the
530 temporal resolution of storm data on numerical simulations of urban flood inundation. J
531 Hydrol 589:125100. <https://doi.org/10.1016/j.jhydrol.2020.125100>

532 IPCC (2014) In: Core Writing Team, Pachauri, R.K., Meyer, L.A. (Eds.), Climate Change 2014:
533 Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment
534 Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland (151
535 pp.)

536 Jiang H, Hong L (2013) Application of bp neural network to short-term-ahead generating power
537 forecasting for pv system. Adv Mater Res 609:128-131.
538 <https://doi.org/10.4028/www.scientific.net/AMR.608-609.128>

539 Ke Q, Tian X, Bricker J, Tian Z, Guan GH, Cai HY, Huang XX, Yang HL, Liu JG (2020) Urban
540 pluvial flooding prediction by machine learning approaches – a case study of Shenzhen city,
541 China. Adv Water Resour 145:103719. <https://doi.org/10.1016/j.advwatres.2020.103719>

542 Kenney MA, Janetos AC (2020) National indicators of climate changes, impacts, and
543 vulnerability. Climatic Change 163:1695-1704. <https://doi.org/10.1007/s10584-020-02939-4>

544 Lei XX, Chen W, Panahi M, Falah F, Rahmati, Uuemaa E, Kalantari Z, Ferreira CSS, Rezaie F,
545 Tiefenbacher JP, Lee S, Bian HY (2021) Urban flood modeling using deep-learning
546 approaches in Seoul, South Korea. J Hydrol 601:126684.

547 <https://doi.org/10.1016/j.jhydrol.2021.126684>

548 Leong WC, Kelani RO, Ahmad Z (2019) Prediction of air pollution index (API) using support
549 vector machine (SVM). *J Environ Eng* 8:103208.
550 <https://doi.org/10.1016/j.jece.2019.103208>

551 Li X, Cheng X, Wu W, Wang Q, Li Y (2019) Forecasting of bioaerosol concentration by a back
552 propagation neural network model. *Sci Total Environ* 698:134315.
553 <https://doi.org/10.1016/j.scitotenv.2019.134315>

554 Ma J, Zhang YF, Li YM, Zhou L, Qin LY, Zeng YW, Wang P, Lei Y (2021) Deep dual-side
555 learning ensemble model for parkinson speech recognition. *Biomed Signal Proces*
556 69:102849. <https://doi.org/10.1016/j.bspc.2021.102849>

557 Masson V, Lemonsu A, Hidalgo J, Voogt JA (2020) Urban climates and climate change. *Annu.*
558 *Rev Env Resour* 45:411-444. <https://doi.org/10.1146/annurev-environ-012320-083623>

559 Mignot E, Li XF, Dewals B (2018) Experimental modelling of urban flooding: a review. *J Hydrol*
560 568:334-342. <https://doi.org/10.1016/j.jhydrol.2018.11.001>

561 Min M, Hongbo Z, Changhong M (2018) Spatio-temporal evolution analysis of the urban heat
562 island: a case study of Zhengzhou city, China. *Sustainability* 10(6):1992.
563 <https://doi.org/10.3390/su10061992>

564 Moghadas M, Asadzadeh A, Vafeidis A, Fekete A, Ktter T (2019) A multi-criteria approach for
565 assessing urban flood resilience in tehran, iran. *Int J Disast Risk Re* 35:101069.
566 <https://doi.org/10.1016/j.ijdr.2019.101069>

567 Mosavi A, Ozturk P, Chau KW (2019) Flood prediction using machine learning models:
568 literature review. *Statistics* 10:1536. <https://doi.org/10.3390/w10111536>

569 Nayak MA, Ghosh S (2013) Prediction of extreme rainfall event using weather pattern
570 recognition and support vector machine classifier. *Theor Appl Climatol* 114:583-603.
571 <https://doi.org/10.1007/s00704-013-0867-3>

572 Ntelekos AA, Oppenheimer M, Smith JA, Miller AJ (2010) Urbanization, climate change and
573 flood policy in the United States. *Climatic Change* 103:597-616.
574 <https://doi.org/10.1007/s10584-009-9789-6>

575 Overmars KP, Koning GHJD, Veldkamp A (2003) Spatial autocorrelation in multi-scale land use
576 models. *Ecol Model* 164: 257-270. [https://doi.org/10.1016/S0304-3800\(03\)00070-X](https://doi.org/10.1016/S0304-3800(03)00070-X)

577 Panahi M, Jaafari A, Shirzadi A, Shahabi H, Rahmati O, Omidvar E, Lee S, Bui DT (2020) Deep
578 learning neural networks for spatially explicit prediction of flash flood probability. *Geosci.*
579 *Front* 12:376-389. <https://doi.org/10.1016/j.gsf.2020.09.007>

580 Rahebeh A, Romulus C, Hossein SM, Quoc Bao P (2020) Flash-flood susceptibility mapping
581 based on XGBoost, Random Forest and Boosted Regression Trees. *Geocarto Int* 1-18.
582 <https://doi.org/10.1080/10106049.2021.1920636>

583 Schmitt TG, Thomas M, Ettrich N (2004) Analysis and modeling of flooding in urban drainage
584 systems. *J Hydrol* 299:3-4. <https://doi.org/10.1016/j.jhydrol.2004.08.012>

585 Schreider SYu, Smith DI, Jakeman AJ (2000) Climate change impacts on urban flooding.
586 *Climatic Change* 47:91-115. <https://doi.org/10.1023/A:1005621523177>

587 Sharifi A (2020) Trade-offs and conflicts between urban climate change mitigation and
588 adaptation measures: a literature review. *J Clean Prod* 276:122813.
589 <https://doi.org/10.1016/j.jclepro.2020.122813>

590 Singh KK, Pal M, Singh VP (2010) Estimation of mean annual flood in indian catchments using

591 backpropagation neural network and m5 model tree. *Water Resour Manag* 24(10):2007-
592 2019. <https://doi.org/10.1007/s11269-009-9535-x>

593 Sun JY, Shen H, Qu Q, Sun W, Kong XQ (2021) The application of deep learning in
594 electrocardiogram: where we came from and where we should go? *Int J Cardiol* 337:71-78.
595 <https://doi.org/10.1016/j.ijcard.2021.05.017>

596 Suwit O, Parinda P (2016) Urban flood mitigation and prevention using the MIKE 21 model: a
597 case study of Nakhon Ratchasima province, Thailand. *Suranaree J Sci Technol* 23:461-479.
598 <https://doi.org/10.1080/10106049.2021.1920636>

599 Thanvisitthpon N, Shrestha S, Pal I, Ninsawat S, Chaowiwat W (2020) Assessment of flood
600 adaptive capacity of urban areas in thailand. *Environ Impact Asses* 81:106363.
601 <https://doi.org/10.1016/j.eiar.2019.106363>

602 Tobler WR (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ*
603 *Geogr* 46:234-240. <https://doi.org/10.2307/143141>

604 Wang B, Loo B, Zhen F, Xi G (2020) urban resilience from the lens of social media data:
605 responses to urban flooding in nanjing, china. *Cities* 106:102884.
606 <https://doi.org/10.1016/j.cities.2020.102884>

607 Wang H, Song L (2019) Water level prediction of rainwater pipe network using an svm-based
608 machine learning method. *Int J Pattern Recogn* 34:2551002.
609 <https://doi.org/10.1142/S0218001420510027>

610 Wang J, Hu CH, Ma B, Mu X (2020) Rapid urbanization impact on the hydrological processes in
611 Zhengzhou, China. *Water* 12(7):1870. <https://doi.org/10.3390/w12071870>

612 Wu ZN, Zhou YH, Wang HY (2020) Real-time prediction of the water accumulation process of

613 urban stormy accumulation points based on deep learning. IEEE Access 8:151938-151951.
614 <https://doi.org/10.1109/ACCESS.2020.3017277>

615 Xiao JK, Zhu XH, Huang CX, Yang XG, Wen FH, Zhong MR (2019) A new approach for stock
616 price analysis and prediction based on ssa and svm. Int J Inf Tech Decis 18:87-310.
617 <https://doi.org/10.1142/S021962201841002X>

618 Xu H, Ma C, Lian J, Xu K, Evance C (2018) Urban flooding risk assessment based on an
619 integrated K-Means cluster algorithm and improved entropy weight method in the region of
620 Haikou, China. J Hydrol 563:975–986. <https://doi.org/10.1016/j.jhydrol.2018.06.060>

621 Zhou WW, Chen MY, Yang ZL, Song XB (2020) Real estate risk measurement and early warning
622 based on PSO-SVM. Socio-Econ Plan Sci 77:101001.
623 <https://doi.org/10.1016/j.seps.2020.101001>

624 **Highlights**

- 625 • Proposes a prediction method for urban flood process based on different deep learning methods.
- 626 • It is feasible to construct prediction model for each water accumulation point.
- 627 • Gradient Boosting Decision Tree is the most suitable method for ponding process prediction.
- 628 • Rainfall, rainfall peak and location coefficient have the greatest impact on ponding.
- 629 • Realize the real-time prediction of urban flood driven by rainfall forecast data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Highlights.doc](#)