



Urban flood prediction using ensemble artificial neural network: an investigation on improving model uncertainty

Weijun Dai^{1,2} · Yanni Tang¹ · Ning Liao¹ · Shujie Zou³ · Zhiming Cai³

Received: 12 January 2023 / Accepted: 16 April 2024 / Published online: 28 May 2024
© The Author(s) 2024

Abstract

Reducing the impact of artificial neural networks (ANN) affected by sources of uncertainty is crucial to improving the reliability of the flood prediction model. This study proposes an ensemble artificial neural network (EANN) model to predict the degree of flooding in coastal cities. Combined methods are used to reduce the model's uncertainty, heuristic neural pathway strength feature selection is used to select inputs, the coupling method is used to optimize network architecture and parameters, and the integration method which paralleling three ANN models with different predicted lead periods ensemble together is used to capture output uncertainty. The EANN model has successfully predicted flooding in the Chinese coastal city of Macao during a typhoon, with convincing accuracy. The study also analyzed the impacts of both long and short training datasets with appropriate time intervals on ANN modeling performance. It was found that the performance of short training datasets, with appropriate time intervals, was similar to or better than models with long training datasets.

Keywords Uncertainty analysis · EANN · HNPSFS · Floods prediction · Typhoon event

Introduction

Cities in low-lying coastal areas are vulnerable to flooding (Patel, et al. 2019). The increased frequency of flood events has brought serious economic losses and social impacts.

A reliable flood forecast is important to mitigate damage from flooding. Data-driven models based on the training and calibration of hydrological data are used to predict flooding. Different from the traditional numerical models, building such models requires low cognitive requirements, simple structure, and fast computationally, yielding more accurate results (Solomantine and A., 2008). Various machine learning methods are used to develop data-driven models, among which, artificial neural networks are the most widely used.

Neural network models trained from data are affected by uncertainty. Sources of uncertainty and their coping methods include (Abrahart and Anctil 2012; Herzog 2020): (i) The uncertainty of the input data. Models trained on larger data sets are historically more representative than those trained on smaller datasets, though choosing the correct data to train the model is the best way to improve its competitiveness. (ii) The uncertainty of model structure. The most optimal structures are important in fitting a model that can adequately account for all parameters to describe changes in flooding. (iii) The uncertainty of parameters. Models' parameters need to be optimized to get the most exact output. Uncertainty affects the prediction results of the ANN model, thus reducing the potential applications of the model (Kasiviswanathan and Sudheer 2013). Reducing model uncertainty is crucial for both research and practical applications. Numerous

✉ Zhiming Cai
zhiming_cai@sina.cn

Weijun Dai
weijun_dai@sina.cn

Yanni Tang
tangyannier@126.com

Ning Liao
bluesky1122@qq.com

Shujie Zou
2464825635@qq.com

¹ Faculty of Electronic Information Engineering, Heyuan Polytechnic College, No. 1 Donghuan Road, Heyuan 517000, Guangdong, China

² College of Artificial Intelligence, Guangdong Open University, No. 1 Xiatang West Road, Guangzhou 510091, Guangdong, China

³ Faculty of Digital Science and Technology, Macao Millennium College, No. 255 Songyusheng Square, Macao 999078, China

approaches that discuss dealing with uncertainty in neural networks have been put forth (Bowden et al. 2005a, 2005b). The importance of the ANN model for suitable input selection in water resources applications has been discussed with two different input determination techniques: The model-free method and the self-organization mapping methods are proposed (Gomez et al. 2019). The impact of the uncertainties associated with a weather forecast, tidal, and storm surge prediction on the accuracy of the flood prediction model was studied (Tiwari, et al. 2010). An artificial neural network based on the bootstrap program was designed to quantify the uncertainty of the model parameters in flood prediction (Zhang and Shin 2021). A new uncertainty propagation scheme based on the Gaussian Mixture Model for neural networks was proposed in Kabir (2021). Uncertainty bounds were used to quantify the uncertainty in the neural network output (Berkhahn et al. 2019; Jhong et al. 2017). Therefore, integration methods are used to capture the impacts of output uncertainty on flood prediction.

The above work all processes the uncertainty of neural networks from a single source. This study uses the EANN model to predict the degree of flooding in urban areas by combining three methods according to reduce the overall uncertainty of neural networks and improve prediction accuracy. The first method is a heuristic neural path intensity feature selection method, which is used to select the best input variables from the candidate datasets. The second method is a coupling method, which is used to determine the number of hidden layer neurons and the number of training sample sets, reducing the network complexity, and determining the best network architecture. The third method is an integration method, we quote an ensemble by paralleling three neural networks with different predicted lead periods together and to finish the prediction task. A simple linear average ensemble, weighted linear average ensemble, and nonlinear neural ensemble, to conduct integrated output on the prediction result. The best result is chosen as the final output, thus capturing the uncertainty of the network output. Moreover, an attempt is also made to compare the performance of the ANN models on the training datasets of different lengths.

Materials and methods

Data acquisition

This paper utilizes the flood data set from the Chinese city of Macao for the case study. It is presented in another paper by the authors (Dai et al. 2021), which contains flood-related data related to seven distinct typhoon events. The prediction model output is the flood depth value at future intervals. Potential inputs to the model include geographic parameters and flood depth of the flooded areas, optimal typhoon path,

urban meteorology, and tides; which can be categorized by invariant and time-lag features.

Input variable selection

The first step in artificial neural network development is to determine the important input variables or characteristics to account for. The selection of high-quality and representative data is integral to creating an accurate model. A large number of potential input variables need to be identified, including all variables concerning the output variables being modeled, to minimize any potential loss of information. Additionally, the model must be able to eliminate variables with little information or noise to ensure that the variables are all relevant and an appropriate input set is reached. The flood prediction studied in this paper is a time series problem, and appropriate lags must also be chosen to maximize reliability. The output of the prediction model can be obtained from the following formula:

$$D_{(t+T)} = f(z, u_{(t-0*T)}, u_{(t-1*T)}, \dots, u_{(t-m*T)}) \quad (1)$$

In the formula, z is the invariant feature of the input, u is the input's time-lag feature, T is the lead period, and m is the lagged value. For example, when $m = 2$, the model outputs the flood depth value $D_{(t+30\text{min})}$ after 30 minutes of T time. The lagged feature is the flood-related time series features of the $u_{(t-0*30\text{min})}$, $u_{(t-1*30\text{min})}$ before 30 min and $u_{(t-2*30\text{min})}$ before 60 min at current T time. Determining how many lagged values are included in each input time series is a difficulty (Bowden et al. 2005c). In typhoon events, flooding can be rapid and transient. For example, in the typhoon "1713 HATO", the flooding that caused the disaster in Macao occurred within 30 min. Considering the limited length of the flood-related time series in the datasets, it was determined that the maximum of the lagged value m is set to 2 and the maximum of the lagged time is $2T$. Table 1 describes the potential input variables used in the model selection.

The HNPSFS method is used to select the input features in this study. The forward selection was firstly performed to construct an initial network by selecting a set of input features. During each subsequent step, a set of time-lagged features was added to the input and the model was trained until it achieved the desired results. Backward selection is then performed to introduce a neural pathway strength feature selection method (Usman et al. 2017). Uncorrelated variables and their subsequent output variables are removed to retain the appropriate input features. Formula 2 calculates the strength of a specific path of neural networks from a given input to an output:

$$W_{IO} = W_{IH} * W_{HO} \quad (2)$$

Table 1 A summary of potential input variables

Categories	Symbol	Definition	Numbers
Invariant features (z)	H, L_o, L_a, T_{yI}, L_d	Geographic elevation, longitude, latitude of submerged area Typhoon intensity, distance from the city at time t	5
Lagged features ($u_{(t-m:T)}$)	$T_{yLa}, T_{yLo}, T_{yLa}(t-1), T_{yLo}(t-1), T_{yLa}(t-2), T_{yLo}(t-2), T_{yP}, T_{yW}, T_{yP}(t-1), T_{yW}(t-1), T_{yP}(t-2), T_{yW}(t-2), R, W_s, W_d, R(t-1), W_s(t-1), W_d(t-1), R(t-2), W_s(t-2), W_d(t-2), T_{iM}, T_{iZ}, T_{iM}(t-1), T_{iZ}(t-1), T_{iM}(t-2), T_{iZ}(t-2), D(t), D(t-1), D(t-2)$	Typhoon motion longitude, latitude at time $t, t-1, t-2$ Typhoon center pressure and wind speed at time $t-1, t-2$ Urban rainfall, wind speed, wind direction at time $t, t-1, t-2$ Urban tide at time $t, t-1, t-2$ Flood depth of submerged area at time $t-1, t-2$	3*10

In Formula 2, W_{IH} is the weight between the input layer and the hidden layer. W_{HO} is the weight between the hidden and output layers. The larger the W_{IO} is, the more important the feature is, leading to a retained input variable. When the value of W_{IO} is 0 and below, it indicates that the input feature suppresses the output and should be removed.

Network architecture selection

Artificial neural networks need to quantify the model coefficients and adapt and modify the network architecture to become more accurate and facilitate data learning. Network training is used to determine the weights and thresholds, like found in the Levenberg–Marquardt algorithm, between the input and hidden layers, the hidden layers and the output layers, and the appropriate transfer functions, generally Sigmoid and linear functions, are selected. These methods are consistent. Two other important parameters for selecting the network architecture, which are the number of hidden layer neurons and the number of datasets used for training, currently have no consistent method, which is usually dealt with based on a typical trial-and-error approach (Abrahart and Anctil 2012). This study adopts a coupling method to select these two parameters. First, the initial value of the number of neurons in a hidden layer is set to derive the following formula:

$$n_H = \sqrt{m + l} + cc \in [110] \tag{3}$$

In Formula 3, n_H is the number of neurons in the hidden layer, m is the number of neurons in the input layer and l is the number of neurons in the output layer. Then, starting from the calculated initial value of n_H , the value is gradually

reduced during each network training. Mean Squared Error (MSE) and training epochs are used as indicators to evaluate the network performance in each iteration. When the indicator value is the minimum value, the corresponding n_H is selected as the nodes of neurons in the hidden layer.

Similarly, on the basis that the n_H value has been selected properly, whole-body training data are sampled at different time intervals to derive datasets varying from large to small. Network training is performed using each subset, when the MSE and training epochs are minimal, the corresponding subset is selected as the determined training dataset.

Ensemble artificial neural network

Integration methods can capture the effects of the ANN output uncertainty, reduce output variance, and obtain more accurate results than individual models. Three ensemble techniques are adopted to improve the model’s performance, simple average ensemble (SAE), weighted average ensemble (WAE), and nonlinear neural ensemble (NNE) (Nourani et al. 2018).

The prediction model must choose the appropriate lead period. The longer the predicted lead period, the more potential the model has, but a longer lead period also leads to greater degrees of uncertainty (Jhong et al. 2017). As said previously, urban flooding usually has a short lead period. This study has determined a forecast lead period of 30 min based on the needs of the relevant urban management departments to issue flood warnings. With this information, three network models with a prediction lead period of 30 min, 60 min, and 120 min were constructed. They are all trained using the error backpropagation algorithm and are named BPNN30, BPNN60, and BPNN120.

The error backpropagation algorithm stands as one of the most widely used optimization algorithms in neural network training, playing a pivotal role in enhancing the network's performance. By adjusting the weights and biases within the network, this algorithm aims to minimize the error function, thereby enabling the neural network to better adapt to the training data. This advantage also serves as a springboard for us to delve into another pertinent issue, namely, the impact of varying lengths of training datasets on the performance of ANN models. The output of the three networks is aligned with the basis of time. The integrated output of the flood forecast value with an advance time of 30 min was conducted with the SAE, WAE, and NNE technologies. The performance of the three integrated techniques was then compared and the best result among them was selected as the final result of flood prediction. The structure of the ensemble neural network is shown in Fig. 1.

Different from general integration methods, this study does not integrate multiple different types of models, instead integrating the results of neural network models with different prediction lead periods. This is an innovative attempt to fit a model with the appropriate lead period with high accuracy.

Model evaluation

This study aimed to predict the depth of continuous flooding and employs rigorous model evaluation techniques to

identify the most suitable regression model for accurate predictions. Root of Mean Square Error (RMSE) and Coefficient of Determination (R^2) are measures used to evaluate the network model (Dai and Cai 2021). The calculation formula can be expressed as:

$$RMSE(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x_i) - y_i)^2} \tag{4}$$

$$R^2(X, h) = 1 - \frac{\sum_{i=1}^m (h(x_i) - y_i)^2}{\sum_{i=1}^m (h(x_i) - \bar{y})^2} \tag{5}$$

In the formula, m is the samples number, $h(x_i)$ is the predicted value of the i -th sample, y_i is the observed value of the i -th sample, and \bar{y} is the average value of the observation sample.

Results and discussion

Input selection

The HNPSFS is used to select different input features and form multiple training subsets to train the BPNN30 model. Figure 2 shows the distribution of the training results in a scatterplot. In the forward selection phase, graph a in Fig. 2

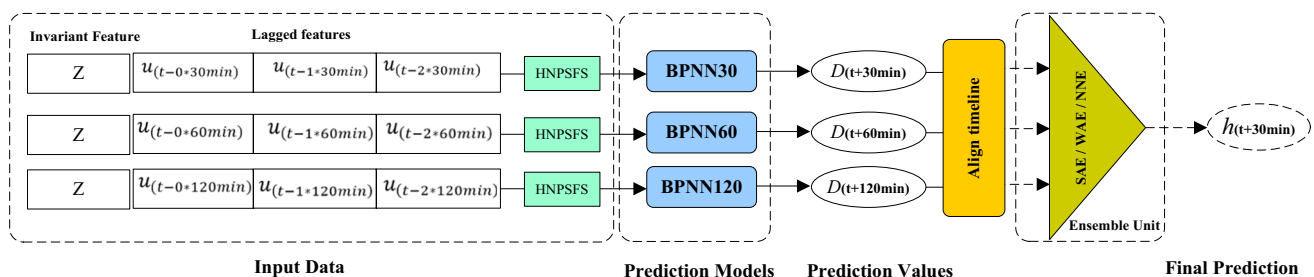


Fig. 1 Structure diagram of the EANN

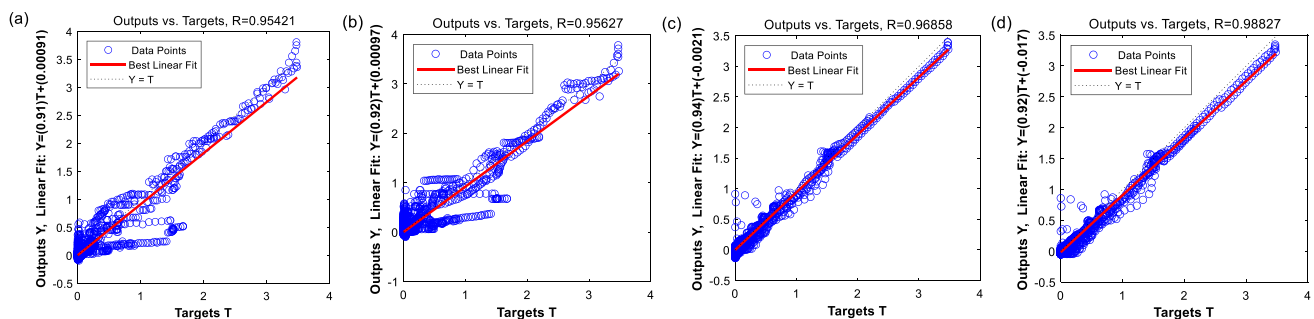


Fig. 2 Scatter plot of training model with different number of features. a 15; b 25; c 35; d 19

is the scatter diagram of a total of 15 input features for selecting z and $u_{(t-0*60min)}$. Figure 2b is the scatter diagram of adding 10 input features of $u_{(t-1*60min)}$ based on Fig. 2a. Figure 2c is the scatter diagram after adding 10 additional input features of $u_{(t-2*60min)}$. The model prediction performance improves significantly as the time-lagged features are added to the input. The distribution of the training sample points is scattered on both sides of the trend line, gradually becoming more tightly clustered, and the distribution trend is consistent with the trend line. At this point, adding more input variables is no longer meaningful, and may make the model more complex. Next backward selection is employed resulting in the removal of sixteen irrelevant input variables, causing the retention of 19 significant input variables due to the HNPSFS method. Figure 2d shows the training results of the model. Compared with Fig. 2c, d shows an equivalent regression effect, but with fewer input variables and a simpler model.

Figure 3 shows the intensity values W_{IO} for the 35 feature variables derived during the HNPSFS process. From the value of W_{IO} , inputs with positive effects on flood

depth prediction include: the invariant feature is the longitude of submerged area (L_o), the latitude of the submerged area (L_a). The effect of time-lagged features on the model output is dynamic during different time-lagged periods. The urban rainfall $R, R(t-1), R(t-2)$ is selected as the input during the three time-lagged periods, indicating that the important impact of rainfall on the output is continuous. In two time-lagged periods, typhoon motion longitude ($T_{yLo}, T_{yLo}(t-1)$), typhoon center pressure ($T_{yp}(t-1), T_{yp}(t-2)$), typhoon center wind speed ($T_{yw}(t-1), T_{yw}(t-2)$), urban tide of Zhuhai ($T_{iz}(t-1), T_{iz}(t-2)$), flood depth of submerged area ($D(t), D(t-2)$) are chosen for input. In one time-lagged period, urban wind speed (W_s), urban wind direction ($W_d(t-1)$), urban tide of Macao (T_{im}) are chosen for input. The value of W_{IO} of other features is negative, indicating that they have an inhibitory effect on model performance and cannot be selected for model input.

Table 2 further presents the W_{IO} values of the 19 significant input variables. Among them, the W_{IO} value of $D_{(t-2)}$ is maximal, which is up to 1.502 and 3 to 22 times the value of

Fig. 3 Intensity values for the 35 feature variables

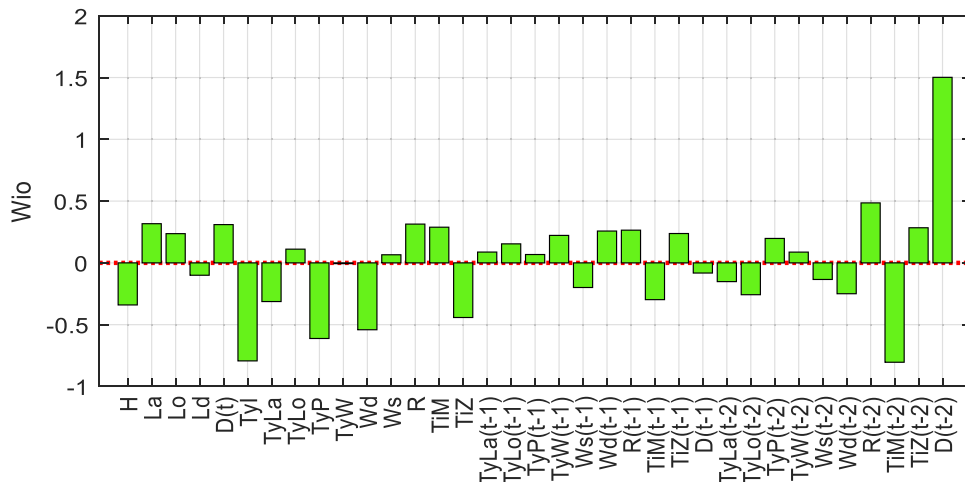


Table 2 W_{IO} value of the 19 significant input variables

Relevant variables	Invariant features	Lagged features
Longitude of submerged area	L_o : 0.237	
Latitude of submerged area	L_a : 0.317	
Typhoon motion longitude		T_{yLo} : 0.111 $T_{yLo}(t-1)$: 0.154
Typhoon motion latitude		$T_{yLa}(t-1)$: 0.087
Typhoon center pressure		$T_{yp}(t-1)$: 0.068 $T_{yp}(t-2)$: 0.198
Typhoon center wind speed		$T_{yw}(t-1)$: 0.222 $T_{yw}(t-2)$: 0.087
Urban wind speed		W_s : 0.066
Urban wind direction		$W_d(t-1)$: 0.258
Urban rainfall		R : 0.314 $R(t-1)$: 0.265 $R(t-2)$: 0.485
Urban tide(M)		T_{im} : 0.289
Urban tide(Z)		$T_{iz}(t-1)$: 0.238 $T_{iz}(t-2)$: 0.284
Flood depth of submerged area		$D(t)$: 0.310 $D(t-2)$: 1.502

other important input variables, indicating that flood depth of submerged area has the greatest influence on the predicted output. This result may also serve as a reminder to city managers that it is crucial to accurately and frequently collect data on changes in flooded areas in urban flood prediction scenarios. Urban weather and tidal during the typhoon event, as well as longitude and latitude of the submerged area, are more important than the optimal typhoon path. Meaning that the closer the typhoon path, such as the longitude and latitude of the moving path, is to the city, the more likely the typhoon directly affects the urban meteorology and offshore tides, and the greater the possibility of flood, which has an important impact on the model output.

Different from other input variable selection methods (such as correlation analysis, principal component analysis, information entropy calculation), the input selection is based on the performance of the artificial neural network. Thus, the HNPSFS method is used to help to identify important input parameters and reduce the number of input variables.

Structure selection

The choice of the number of neurons in the hidden layer n_H is important for determining the neural network architecture. Too high of an n_H value will increase the system's complexity. In reverse, too small of an n_H value results in insufficient modeling capabilities of nonlinear systems and increases the network architecture uncertainty. n_H is closely related to the model output and the input dimension. The 19 input variables are used together with the coupling method to select the best network architecture. The BPNN30 model is trained by a training set of 104,220 samples, where the n_H values are between 1 and 80. Figure 4 shows the change in the MSE and epochs for the training dataset.

In formula 3, $n_H = 15$ is calculated. In Fig. 4a, when $n_H \leq 10$, the MSE increases. When $n_H = 22$, the MSE will peak. When $n_H > 23$, the MSE of the dataset changes decreases. An increase in n_H indicates an increase in the model's complexity. This indicates that there is a best performance window when $10 \leq n_H \leq 21$. in Fig. 4b,

when $n_H \leq 5$, the model is the simplest, but the model also requires many iterations and a longer training time. When $n_H > 5$, the epochs maintains low values. When $10 \leq n_H \leq 19$, there is an interval with small epochs value and gentle change. Therefore, the negative search is conducted with initial n_H value as the starting point, the best network architecture is in the $10 \leq n_H \leq 15$ window. The results show that for a given input, output and n_H , when the smaller value of MSE and epochs (or training time) is found, the optimal parameters and network architecture can be objectively selected.

The number and quality of samples in the training dataset directly affect the model's performance. It is another source of uncertainty in the network architecture. The degree of data quality is often unknown, and the network performance can only be tested and analyzed, either by splitting samples or extracting samples to build training datasets of different lengths. The flooding data used in this study are gathered in a record for each minute, and other input variables are also 1 record per minute after interpolation, forming a training set of 104,220 sample pairs. Data was extracted at 5-min, 15-min, 30-min, and 60-min intervals, resulting in five training subsets. The sampling data length is 104,220, 20,844, 6,948, 3,474, and 1,737, which corresponds to 100%, 20%, 6.67%, 3.33%, and 1.67% of the total sample size, respectively. Figure 5 shows the impact of the long and short training dataset on the BPNN30 model performance. When $n_H = 15$ and the number of training samples is reduced from 100.00 to 20.00% of the total, the MSE of the model increases. When it is further reduced to 6.67% of the total, the MSE of the model decreases. The epochs show a decreasing trend and the network performance becomes better. As the number of training samples decreases further, the MSE and training time of the model increase, and the network performance becomes worse. Urban flooding is a time series problem, compared to the all sample data of 1 record per minute, the sample data drawn from 15-min intervals reduce redundant information and briefly describe the complex relationship of the flooding process. The sample data drawn from 30

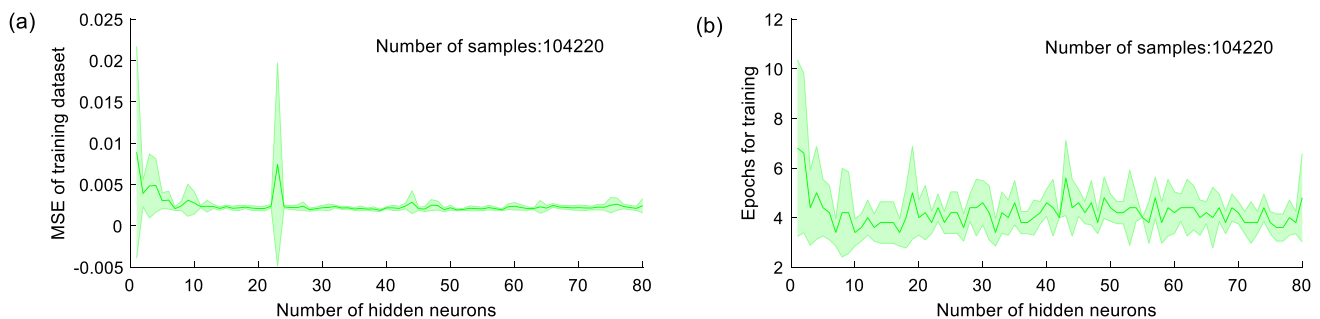


Fig. 4 A comparison of the performance of the ANN model with increasing model complexity. **a** MSE; **b** epochs

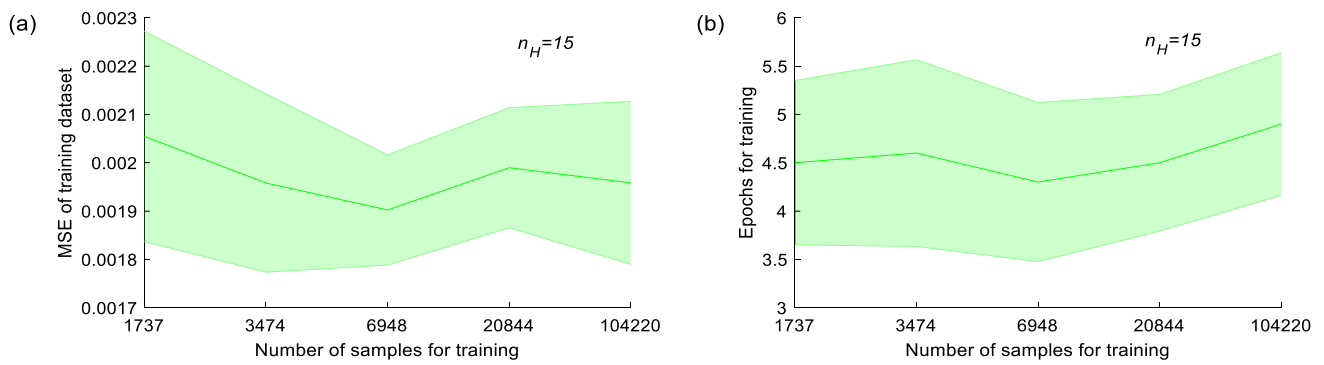


Fig. 5 A comparison of the performance of the ANN model with using long and short training datasets. **a** MSE; **b** epochs

to 60-min intervals lose a lot of details and are no longer good representatives of the data, which leads to poor performance in the data-driven model.

The ANN model training is not to obtain a suitable prediction network, but to select the best network architecture. Here the best network architecture is chosen as $n_H = 15$ and the 15-min interval sampling subset is selected for the training dataset. In terms of MSE, the current model demonstrates a low value with minimal variability, indicating the model resides within the optimal performance window. As for the number of training epochs, the model exhibits low values, reflecting its simplicity and avoidance of extensive training durations. Additionally, increasing the amount of training data will not lead to significant improvements in the model's performance, further validating its adequacy for the given task. These attributes collectively contribute to the superiority of the current network. Compared with the trial-and-error method, the proposed method chooses to design the network architecture parameters based on the objective results, which can be used to reduce the uncertainty of the neural network architecture.

Network output

Based on the determined 19 inputs and best network architecture parameters, BPNN30, BPNN60, and BPNN120

models are trained and tested. Figure 6 shows the test results for the three networks. Compared with the observed samples, the prediction values of BPNN30 and BPNN60 accurately capture the peak flood occurrence. However, their prediction results for the flood peak above 1.5 m are too high and good time synchronization for both the flood rise and descent changes is presented. The prediction results of BPNN120 show multiple shocks and jumps. Its prediction for the peak flood level above 1.5 m is close to the actual value but with great delays or advances in time. It can be seen that the prediction output of the individual model is insufficient and can be compensated by using integration technology.

The integration learning framework is centered on the meticulous selection and seamless integration of multiple base learners, forming a close alliance with classical model selection theory. The methodologies utilized for model selection include cross-validation, grid search, and model evaluation. Typically, the quest is to identify an aggregated model whose generalization error significantly outperforms any standalone model. In our investigation, BPNN30 emerged as the preeminent individual model. Therefore, we embarked on a comparative analysis of evaluation indicators, pitting the three integrated models against BPNN30 as the yardstick for selecting the optimal integrated model. This process ensures that we select the most effective and robust

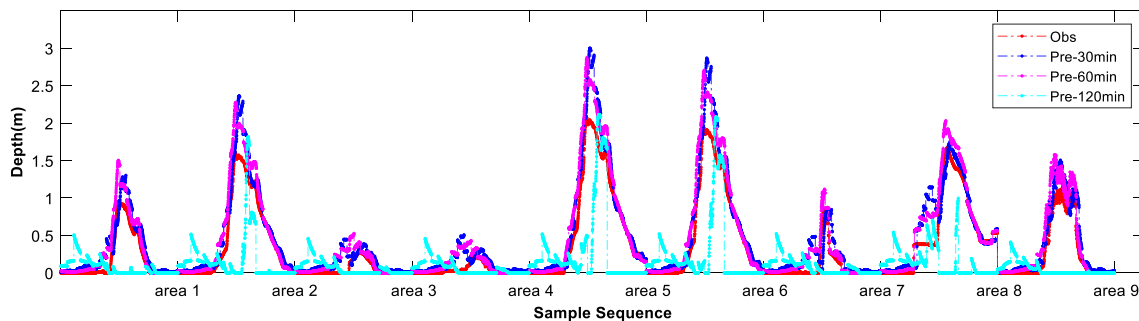


Fig. 6 Prediction results of the BPNN models on the test set

Table 3 Error statistic of the models on the test set

	BPNN30	SAE	WAE	NNE
R ²	0.96356	0.93969	0.98011	0.98367
RMSE (m)	0.20593	0.18701	0.10243	0.093381

ensemble model, tailored to achieve optimal performance. Table 3 presents the error statistics of the BPNN30 model and EANN models on the test sets. R² and RMSE is used to evaluate the model performance. Specifically, the SAE model has the lowest fit, with an R² value of 0.93969. The WAE and NNE models have a higher fit, approximately 2% higher than the BPNN30 model. All three integrating models obtain smaller RMSE values than the BPNN30 model. The SAE model decreases by 10%, the WAE model by 101%, and the NNE model by 120%. The integrating model reduces the variance of individual models and yields better prediction performance.

Figure 7 shows the regression results of the BPNN30 model and EANN models on the test set. It can be seen that the distribution trend of the four models' scatter is consistent with the trend line. The sample of the BPNN30 model mostly falls to the left of the trend line, meaning the majority of predictions are higher than the expected value. The distribution of sample points in the SAE model becomes better. After the observed value is greater than 1.5 m, the prediction sample points distribute more scattered and the prediction value is higher than the observed value. Sample points are more uniformly clustered on the trend line for the WAE model and the NNE model, indicating that the predicted values are closest to the observed values, meaning that the regression is more accurate.

Figure 8 shows the test results of flood prediction thirty minutes in advance for the BPNN30 model and the EANN models. Compared with the BPNN30 model, the flood peak predicted by the SAE model has smaller errors, but the hopping phenomenon of flood timing during flooding was not

improved. The WAE and the NNE model have close performance and perform excellent predictions on flooding depth and time synchronization. In the nine inundation areas, the WAE and the NNE model predict the flooding process more smoothly and continuously, especially for peaks in flooding. This indicates the reliability of the output of the ensemble neural network model. In areas three and four, the rising edge of the predicted flooding is advanced. In area eight, they are slightly lower predictions for flood peaks than the observed values. The reasons behind these differences need to be further explored to be better understood.

Conclusion

Artificial neural networks are widely used in flood prediction, and the adverse effects of uncertainty sources on the model need to be improved. This paper studies the application of the EANN model to improve the certainty of flood prediction in coastal cities, proposes the HNPSFS method to select appropriate inputs, uses the coupling method to select network architecture and parameters, uses nonlinear neural integration technology to capture the uncertainty of output, and realizes the objective method based on the neural network itself to reduce model uncertainty and improve model prediction performance. The model achieves accurate results in the early prediction of floods in the Macao region of China. The prediction output can provide effective guidance for city managers to issue flood warnings. It also has been found that ANN modeling can utilize shore training datasets sampled at appropriate time intervals and that they can have similar or better performance compared to long training dataset models.

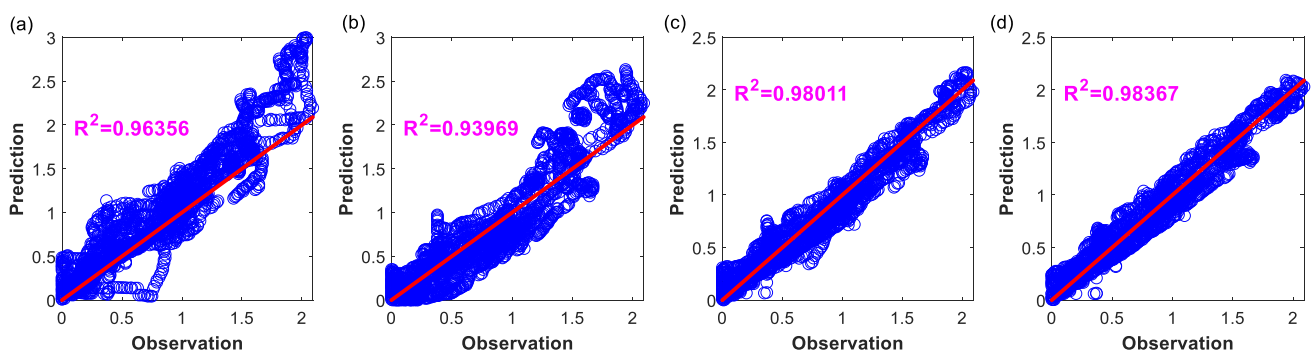


Fig. 7 Result of multivariate regression analysis of the BPNN30 model and the ensemble models. **a** BPNN30 model; **b** SAE model; **c** WAE model; **d** NNE model

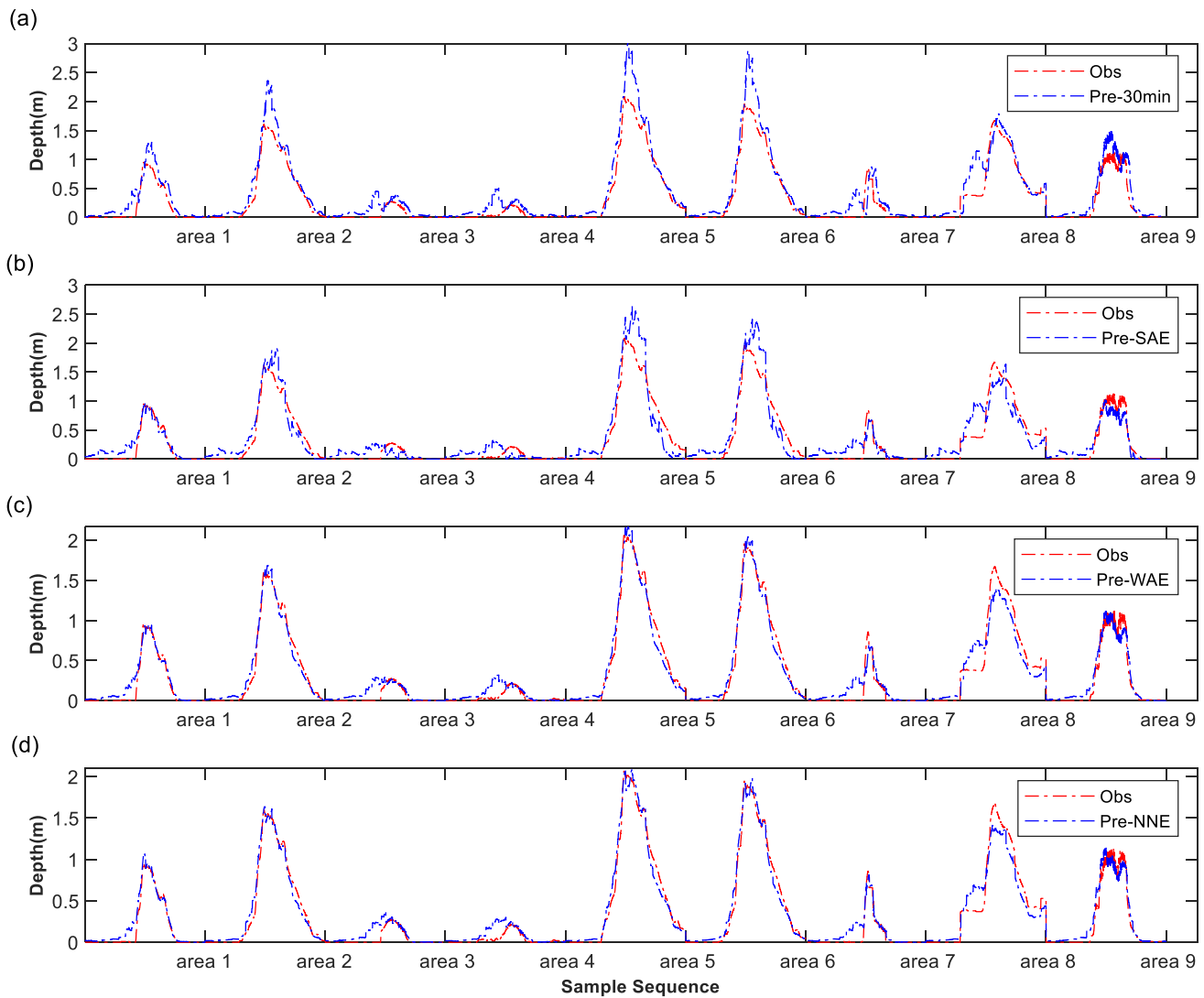


Fig. 8 A comparison of predicted and observed flood depth values. **a** BPNN30 model; **b** SAE model; **c** WAE model; **d** NNE model

Acknowledgements This study was supported by: National Natural Science Foundation of China and Macao Science and Technology Development Joint Project (0066/2019/AFJ): Research on Knowledge-oriented Probabilistic Graphical Model Theory based on Multisource Data. MOST-FDCT Joint Projects (0058/2019/AMJ, 2019YFE0110300): Research and Application of Cooperative Multi-Agent Platform for Zhuhai-Macao Manufacturing Service. Natural Science Characteristic Innovation Project of Guangdong General Universities (2022KTSCX323). Heyuan Social Development Science and Technology Project (2021/62/138).

Funding National Natural Science Foundation of China and Macao Science and Technology Development Joint Fund, 0066/2019/AFJ, Zhiming Cai, Research and Application of Cooperative Multi-Agent Platform for Zhuhai-Macao Manufacturing Service, MOST-FDCT Projects 0058/2019/AMJ, Zhiming cai, 2019YFE0110300, Zhiming Cai, Natural Science Characteristic Innovation Project of Guangdong General Universities, 2022KTSCX323, Weijun Dai, Heyuan Social Development Science and Technology Project, 2021/62/138, Weijun Dai.

Declarations

Conflict of interest We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, “Urban Flood Prediction Using Ensemble Artificial Neural Network: An Investigation on Improving Model Uncertainty”.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abrahart RJ, Anctil FC (2012) Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Prog Phys Geogr* 36(4):480–513
- Berkhahn S, Fuchs L, Neuweiler I (2019) An ensemble neural network model for real-time prediction of urban floods. *J Hydrol* 575:743–754
- Bowden GJ, Dandy GC, Maier HR (2005a) Input determination for neural network models in water resources applications. Part 1—background and methodology. *J Hydrol* 301(1–4):75–92. <https://doi.org/10.1016/j.jhydrol.2004.06.021>
- Bowden GJ, Maier HR, Dandy GC (2005b) Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. *J Hydrol* 301(1–4):93–107. <https://doi.org/10.1016/j.jhydrol.2004.06.020>
- Bowden GJ, Dandy GC, Maier HR (2005c) Input determination for neural network models in water resources applications. Part 1 Background and methodology. *J Hydrol* 301(1–4):75–92
- Dai W, Cai Z (2021) Predicting coastal urban floods using artificial neural network: the case study of Macao, China. *Appl Water Sci* 11(10):161
- Dai W et al (2021) Ensemble learning technology for coastal flood forecasting in internet-of-things-enabled smart city. *Int J Comput Intell Syst* 14(1):1–16
- Dipu Kabir HM, Khosravi A, Kavousi-Fard A, Nahavandi S, Srinivasan D (2021) Optimal uncertainty-guided neural network training. *Appl Soft Comput* 99:106878. <https://doi.org/10.1016/j.asoc.2020.106878>
- Gomez M et al (2019) Skill of ensemble flood inundation forecasts at short- to medium-range timescales. *J Hydrol* 568:207–220
- Herzog L, Murina E, Dürr O, Wegener S, Sick B (2020) Integrating uncertainty in deep neural networks for MRI based stroke analysis. *Med Image Anal* 65:101790. <https://doi.org/10.1016/j.media.2020.101790>
- Jhong B-C, Wang J-H, Lin G-F (2017) An integrated two-stage support vector machine approach to forecast inundation maps during typhoons. *J Hydrol* 547:236–252
- Kasiviswanathan KS, Sudheer KP (2013) Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. *Stoch Env Res Risk Assess* 27(1):137–146
- Nourani V, Gozen Elkiran Abba SI (2018) Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach. *Water Sci Technol* 78(10):2064–2076. <https://doi.org/10.2166/wst.2018.477>
- Patel P, Ghosh S, Kaginalkar A, Islam S, Karmakar S (2019) Performance evaluation of WRF for extreme flood forecasts in a coastal urban environment. *Atmosph Res* 223:39–48. <https://doi.org/10.1016/j.atmosres.2019.03.005>
- Solomatine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. *J Hydroinform* 10(1):3–22. <https://doi.org/10.2166/hydro.2008.015>
- Tiwari MK, Chatterjee C (2010) Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). *J Hydrol* 382(1–4):20–33. <https://doi.org/10.1016/j.jhydrol.2009.12.013>
- Usman TK, Jianxun H, Caterina V (2017) River flood prediction using fuzzy neural networks: an investigation on automated network architecture. *Water Sci Technol* 2017(1):238–247
- Zhang B, Shin YC (2021) An adaptive Gaussian mixture method for nonlinear uncertainty propagation in neural networks. *Neurocomputing* 458:170–183

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.