Theory and Practice of Science and Technology
2025, VOL. 6, NO. 3, 61-65
DOI: 10.47297/taposatWSP2633-456910.20250603

A flood occurrence probability prediction model based on convolutional neural network algorithm

Xiangyang Gao

School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning, 125000, China

ABSTRACT

In order to cope with the situation where the frequency and severity of floods have been increasing in recent years[1]. Based on the hydrological data monitored by the Lijin Hydrological Station of the Yellow River and the China Water Pattern Data Center, this paper uses the traditional neural network model as the benchmark model[2], then prunes the model to reduce the channels, and establishes a convolutional neural network mode to predict the flood occurrence probability[3]. Finally, it compares the prediction accuracy with that of the traditional neural network model, calculates the accuracy indicators, and judges the adaptability and stability of the model. The convolutional neural network model exhibits superior performance in predicting the probability of flood occurrence compared to the traditional neural network model. Compared with the traditional neural network model, the MSE of this model has decreased by 81.04%, the RMSE has decreased by 56.58%, and R^2 has increased by 0.665. Therefore, he adoption of convolutional neural network model can effectively enhance the prediction performance of neural networks. Compared with the traditional neural network model, the convolutional neural network model performs better in the accuracy of flood prediction and is more adaptable in flood prevention and disaster reduction work.

KEYWORDS

Flood; Convolutional Neural Network; Probability prediction; Adaptability

Floods, as common and destructive natural disasters, have caused huge economic losses and social impacts all over the world. With the influence of climate change and human activities, the frequency and severity of floods are showing an increasing trend. Therefore, it is particularly important to have an in-depth understanding and effective prediction of flood occurrence. How to propose more precise and efficient forecasting models for different river basins and environments is an urgent task for hydrological forecasters.

At present, the research on flood prediction at home and abroad can be divided into two types: one is the mechanism-driven model based on physical understanding, such as using the storm flood management model as a basis to conduct regional storm simulation and evaluate its impact. Another research is to establish a dynamic programming model for the joint scheduling of flood control systems. However, mechanism-driven models often have problems such as insufficient understanding of the mechanism and significant influence of human subjective factors. The other is the data-driven model based on statistical science. Data-driven models can be divided into single models and hybrid models. For example, he linear ARX system is applied to calculate the correlation between regional heavy rain and runoff in the single model. And then analyze the river runoff to achieve the effect of flood prediction. For example, SOEBROTO uses the Support Vector Regression Model (SVR) to predict the probability of mountain flood outburst in the next 1-3 hours[4]. Some studies predict floods by analyzing relevant data from geographic information systems, using elevation data as input and runoff as output, and using artificial neural networks (ANN) for prediction[5]. However, single models often have problems such as low prediction accuracy and poor model stability. Hybrid models can effectively solve the shortcomings of single models. For example, Particle Swarm Optimization (PSO)[6] and Genetic Algorithm (GA)[7] can effectively solve the problem of neural networks being prone to falling into local optima. Using the Ensemble Kalman Filter to consider the uncertainty of initial conditions and water level drop can effectively improve the accuracy of the prediction model. Combining the random forest model[8] with the flood water level map model shows that compared with traditional precipitation index models, the hybrid model has higher accuracy.

Based on the above discussion, this article proposes a convolutional neural network model based on an improved neural network model. This model conducts pruning by sparse training to reduce redundant connections. It uses the original model as the teacher network and employs the knowledge distillation method to guide the pruned student network, thereby enhancing the performance and adaptability of the compressed model. The proposed method is tested using classification models and object detection models in the convolutional network. The experimental results show that this model compression framework can effectively reduce the storage and computing requirements of the model, with a reduction in model size exceeding 90% on multiple test models, an increase in inference speed by 3 to 4 times, and a precision loss controlled within 2%. The proposed multi-method fusion model compression framework ensures the performance of the convolutional neural network model while reducing the model size and improving the inference speed, and is suitable for the efficient deployment of convolutional neural networks in resource-constrained environments.

1 Research Methodology

1.1 Data Processing

This study is based on the hydrological data monitored by the Limingzhixin Hydrological Station of the Yellow River. Data processing is carried out, including outlier treatment and missing value treatment. Data points that deviate greatly from the overall data situation and those exceeding a certain unreasonable range will be removed. Outliers will be nullified or filled with other valid values. Input one or more quantitative or categorical variables, and the program will output the sequence after filling in the missing values.

1.2 Calculation of Spearman's Rank Correlation Coefficient

The Pearson correlation coefficient between two variable is defined as the quotient of the covariance and the standard deviation of the two variables[9]:

$$\rho_{X,Y}=\frac{cov(X,Y)}{\sigma_X\sigma_Y}=\frac{E\left[(X-\mu_x)(Y-\mu_Y)\right]}{\sigma_X\sigma_Y} \tag{2.1}$$

μ_x represents the average value of the X set, while μ_Y represents the average value of the Yset. σ_X denotes the variance of the X set, and σ_Y represents the variance of the Y set.

The Spearman's rank correlation coefficient is defined as the Pearson correlation coefficient between two rank variables. For a sample with sample size n, n original data are transformed into rank data, The formula for calculating the correlation coefficient ρ is as follows:

$$\rho=\frac{\Sigma_i(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma_i(x_i-\bar{x})^2\Sigma_i(y_i-\bar{y})^2}} \tag{2.2}$$

Based on the hydrological data monitored by the Lijin Hydrological Station of the Yellow River after data processing, the Pearson correlation coefficients between each influencing factor and the probability of flood occurrence were calculated. The Pearson correlation coefficient results are shown in Figure 1:

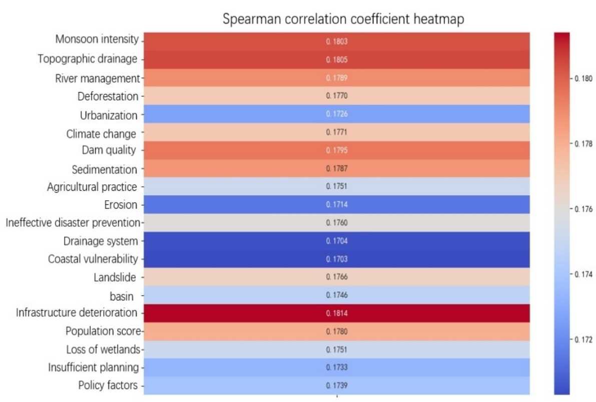

Figure 1 Visual graph showing the Spearman coefficient relationship between each indicator and the probability of flood occurrence

Based on the above results, the study found that indicators such as infrastructure deterioration, monsoon intensity, terrain drainage, dam quality, and river management have a close relationship with the occurrence of floods, while indicators such as drainage system, coastal vulnerability, erosion, urbanization, and watershed have little correlation with the occurrence of floods.

2 Results and Discussion

2.1 Predicting flood probabilities based on neural networks

This paper analyzes the hydrological data monitored by the Yili Jin Hydrometric Station of the Yellow River. An appropriate neural network structure is selected to process the data. The model uses the hydrological data monitored by the Yili Jin Hydrometric Station of the Yellow River after data processing as the training data to train the neural network model. During the training process, the weights are optimized through the backpropagation algorithm to minimize the loss function. Finally, the flood in the hydrological data monitored by the Yili Jin Hydrometric Station of the Yellow River is predicted. The flood probability prediction histogram is drawn using Python as shown in Figure 2:

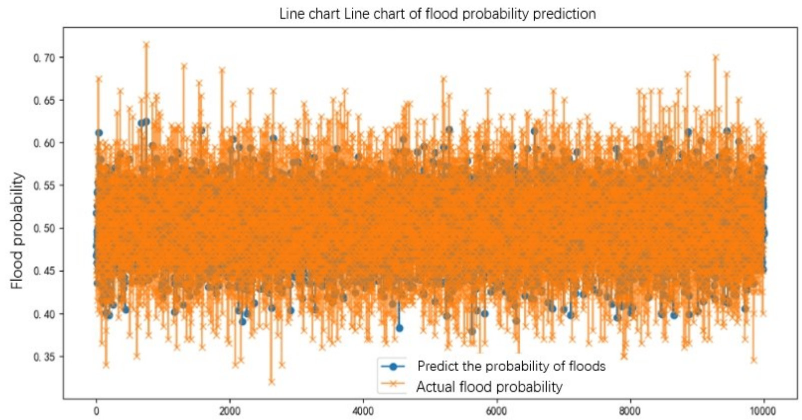

Figure 2 Line Chart of Flood Probability Prediction

Observation of the image reveals that the blue line represents the predicted flood probability while the orange line represents the actual flood probability. The closer the overlap between the two lines is, the higher the accuracy of the model is. After calculation, the accuracy of the model is 81%.

2.2 Verification of Model Accuracy

Compare the prediction results with those of the Yellow River, calculate the values of mean square error, root mean square error and R^2, and the results are presented as shown in Table 1 below:

Table 1 Accuracy Indicators

Accuracy indicators	Size
Mean Squared Error (*MSE*)	0.00049
Root Mean Square Error (*RMSE*)	0.02206
R^2 value	0.81103

When the training data is reduced and five feature values are adopted, the prediction is made again. The new flood probability prediction line chart obtained is shown in Figure 3:

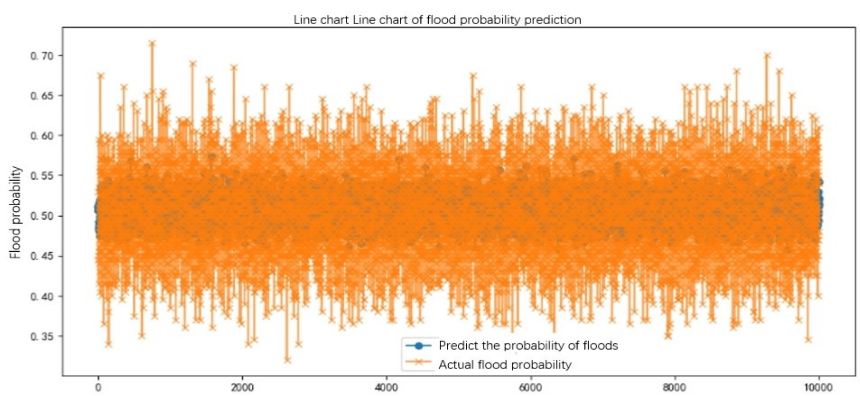

Figure 3 Line chart of flood probability prediction with five characteristic values

Once again, the predicted results were compared with the hydrological data monitored by the Limingjiang Lijin Hydrological Station. The mean square error, the root mean square error and the value of R^2 were calculated.

$$MSE=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\hat{x}_i\right)^2 \#(3.1)$$

The mean squared error is the average of the squares of the distances between the predicted values and the true values. The smaller its value is, the closer the predicted results of the model are to the true values.

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\hat{x}_i\right)^2} \#(3.2)$$

The root mean square error is the square root of the mean square error. If the root mean square errors are similar, then the prediction effects of them may also be similar.

$$R^2 = 1 - \frac{SSE}{SST} \#(3.3)$$

The coefficient of determination R^2 reflects the proportion of the total variation of the dependent variable that can be explained by the independent variable through the regression relationship. The closer the value of R^2 is to 1, the smaller the error of the predicted values of the model.

The result is as shown in Table 2 below:

Table 2 Accuracy Indicators for Five Characteristics

Accuracy indicators	Size
Mean Squared Error (MSE)	0.00211
Root Mean Square Error (RMSE)	0.04595
R^2 value	0.18005

Upon observing the table, it is found that the R^2 value is too small. The reason for this phenomenon is that when the number of selected feature values is limited, the model cannot be accurately trained, which leads to a decrease in accuracy and makes it impossible to establish a neural network well. Therefore, we changed the model and chose to use the convolutional neural network model to make another prediction on the results.

2.3 Establishing Convolutional Neural Network for Flood Probability Prediction

This paper uses the preprocessed data as the dataset for training the convolutional neural network model. We select five features with the strongest correlation, set appropriate input layer and convolutional layer quantities, and adjust the size and quantity of filters. Through techniques such as cross-validation, we adjust the filter size, layer number, learning rate, etc., to improve the generalization ability and performance of the model. Finally, we predict the probability of flood occurrence. We obtain the histogram and line graph of the flood occurrence probability, as shown in Figure 4:

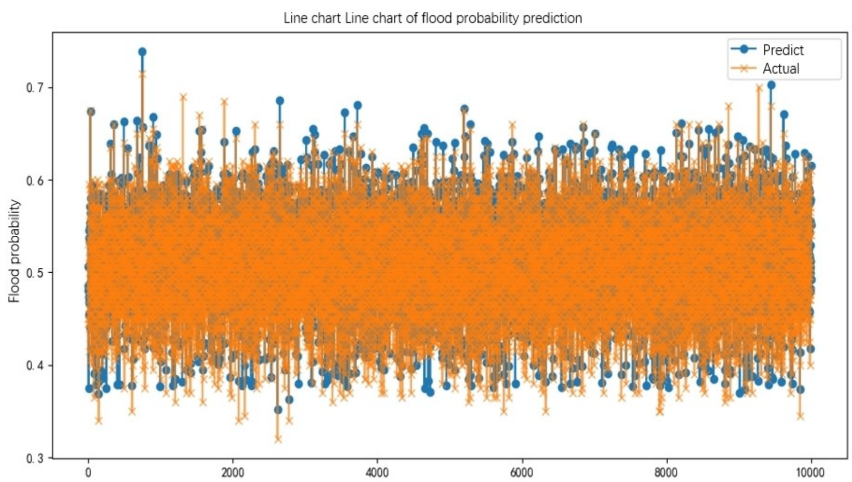

Figure 4 Line chart of flood probability prediction by convolutional neural network

Observing the results in Figure 4: The predicted values of the convolutional neural network are more consistent with the actual values compared to the traditional neural network model. Finally, the model achieves an accuracy of 84% under the condition of five feature values.

Calculate the numerical values of the accuracy indicators and obtain Table 3:

Table 3 Standard Indicators of Convolutional Neural Network

Accuracy indicators	Size
Mean Squared Error (*MSE*)	0.00040
Root Mean Square Error (*RMSE*)	0.01995
R^2 value	0.84538

3 Conclusion

This paper establishes a convolutional neural network model that integrates multiple methods to predict the probability of flood occurrence. The hydrological data monitored at the Lijin Hydrological Station of the Yellow River is used as the input of the model, and the predicted water levels are compared with those of the traditional neural network model. The research results show that, compared with the traditional neural network model, the convolutional neural network model performs better in flood prediction. Its predicted values are closer to the observed values and the prediction accuracy is higher. This indicates that in time series prediction, the convolutional neural network model has stronger compatibility and can accelerate data processing speed and improve accuracy, thereby enhancing the efficiency of flood forecasting.

The model is based on the hydrological data monitored at the Lijin Hydrological Station of the Yellow River. It calculates the Spearman correlation coefficients between each factor and the probability of flood occurrence, and obtains the top 5 factors with the highest correlation values. This paper establishes a convolutional neural network model to predict the probability of flood occurrence based on these 5 highly correlated factors. The core of the convolutional layer is multiple convolution kernels, which slide over the input data and perform weighted summation operations to extract local features. Each convolution kernel generates a feature map, revealing specific local patterns in the input data. These features are provided by the calculated correlation coefficients. The convolutional layer significantly reduces the complexity and training parameters of the model through the parameter sharing mechanism. This means that the same convolution kernel is reused throughout the input data, thereby reducing the risk of overfitting. Existing prediction models face challenges in capturing sudden changes in hydrological data and their subsequent impacts, especially during extreme climate events. The current convolutional neural network model has certain deficiencies in predicting sudden data, and further improvement and optimization are needed. Future improvement directions include, but are not limited to, enhancing the model's adaptability to sudden data, possibly requiring more flexible model structures or strengthening data preprocessing capabilities, to further improve the model's robustness and accuracy. Additionally, the response mechanism for extreme climate events needs to be strengthened to more effectively predict and respond to sudden changes in the water situation of the Yellow River.

References

[1] Li Chunyu. Flood Disaster Risk Assessment and Optimization of Flood Control Engineering Design [J]. Maritime Safety, 2024, (24): 190-192.

[2] Tang Ning. Research on Hydrological Forecasting Method Based on Artificial Neural Network [D]. Nanchang University, 2024. DOI: 10.27232/d.cnki.gnchu.2024.002175.

[3] Guo Kai-tai, Li Yu-zhe, Fu Dong-hao, et al. A Compression Method for Convolutional Neural Network Model Based on Multi-method Fusion [J/OL]. Journal of Xidian University, 1-11 [2025-01-21].

[4] Li Xiaoli, Zhou Xiaojian, Shen Ganggang, et al. Application of Uncertain Support Vector Machine in Flood Prediction Model [J]. Journal of Lanzhou University of Technology, 2012, 38(03): 107-110.

[5] Wu Xunjun, Zhao Xiaodong, Ding Xi, et al. Data-driven CNN-based method for flood evolution prediction [J]. Journal of Hydropower Engineering, 2021, 40(05): 79-86.

[6] Wu Shaohui, Lu Donghui, Wang Lingmin, et al. Research on Joint Flood Control Optimization Operation of Reservoir Groups Based on Improved Genetic Algorithm [J/OL]. China Flood Control and Drought Relief, 1-14 [2025-02-13].

[7] Tang Peiyu. Research on Optimization Scheduling of Chengbi River Reservoir throughout Its Entire Operating Cycle Based on Particle Swarm Algorithm [D]. Guangxi University, 2024. DOI: 10.27034/d.cnki.ggxiu.2024.000550.

[8] Xiao Peng, Song Wanli, Xia Anyu. Research on the Application of Random Forest Model in Visualization of Cascade Flood Control [J]. Automation Technology and Application, 2022, 41(11): 125-128 + 183. DOI: 10.20033/j.1003-7241. (2022) 11-0125-05.

[9] Zhang Weifeng. Statistical Characteristics Analysis of Spearman's Simple Rank Correlation Coefficient and Gini-Gamma Correlation Coefficient [D]. Guangdong University of Technology, 2020. DOI: 10.27029/d.cnki.ggdgu.2020.001333.