

Article

Not peer-reviewed version

Flood Forecasting Method and Application based on Informer Model

[Yiyuan Xu](#) , [Jianhui Zhao](#) ^{*} , Biao Wan , Jinhua Cai , Jun Wan

Posted Date: 26 January 2024

doi: 10.20944/preprints202401.1867.v1

Keywords: flood forecasting; seq length; LSTM; Informer



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Flood Forecasting Method and Application based on Informer Model

Yiyuan Xu ¹, Jianhui Zhao ^{1,*}, Biao Wan ², Jinhua Cai ³ and Jun wan ⁴

¹ School of Computer Science, Wuhan University, Wuhan, China

² School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan, China

³ School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan, China

⁴ Wuhan Luoshui Intelligent Technology Co., Ltd

* Correspondence: jianhuizhao@whu.edu.cn

Abstract: Flood forecasting helps anticipate floods and evacuate people, but due to the access of a large number of IoT data acquisition devices, the explosive growth of multidimensional data and the increasingly demanding prediction accuracy, classical parameter models and traditional machine learning algorithms are unable to meet the high efficiency and high precision requirements of prediction tasks. In recent years, deep learning algorithms represented by convolutional neural networks, recurrent neural networks and Informer models have achieved fruitful results in time series prediction tasks. The Informer model is used to predict the flood flow of the reservoir. At the same time, the prediction results are compared with the prediction results of the traditional method and the LSTM model, and how to apply the Informer model in the field of flood prediction to improve the accuracy of flood prediction is studied. The data of 28 floods in the Wan'an Reservoir control basin from May 2014 to June 2020 were used, with areal rainfall in five subzones and outflow from two reservoirs as inputs and flood processes with different sequence lengths as outputs. The results show that the Informer model has good accuracy and applicability in flood forecasting. In the flood forecasting with sequence length of 4, 5 and 6, Informer has higher prediction accuracy, and the prediction accuracy is better than other models under the same sequence length, but the prediction accuracy will decline to a certain extent with the increase of sequence length. The Informer model stably predicts the flood peak better, and its average flood peak difference and average maximum flood peak difference are the smallest. As the length of the sequence increases, the number of fields with a maximum flood peak difference less than 15% increases, and the maximum flood peak difference decreases. Therefore, the Informer model can be used as one of the better flood forecasting methods, and it provides a new forecasting method and scientific decision-making basis for reservoir flood control.

Keywords: flood forecasting; seq length; LSTM; Informer

1. Introduction

Our country is one of the countries with frequent flood disasters, different types and different degrees of floods may happen on about two-thirds of the land area. According to the statistics of the National Center for Disaster Reduction under the Ministry of Water Resources and the Ministry of Emergency Management, from 1991 to 2020, the number of people killed or missing due to floods in China reached 2,020, with a total of more than 60,000 deaths. Therefore, we need to predict the flood flow of the reservoir according to the information of rainfall and flow in the upstream of the basin through time series prediction, so as to give early warning and try to avoid safety accidents caused by sudden flood peaks [1,2]. In the past, hydrological models were generally used, but these models required a lot of parameters, such as temperature, soil moisture, soil type, slope, terrain, etc., and different parameters also contained very complex relationships [3]. In recent years, machine learning technology has developed rapidly, and many researchers have found that its efficient data parallel processing ability can be applied to the field of flood prediction [4,5].

The Artificial Neural Networks (ANN), one of the data-driven techniques, have been widely used in hydrology as an alternative to physical-based and conceptual models [6,7]. Liu Heng et al. [8]

proposed a flood classification forecast model, which greatly improved the forecast accuracy. Li Dayang et al. [9] proposed the VBLSTM model combining Stochastic Variational Inference (SVI) and Long Short-Term Memory (LSTM), and experiments found that the prediction accuracy is higher than that of traditional hydrological models. XU et al. [10] proposed to use Particle Swarm Optimization (PSO) to optimize LSTM hyperparameters for flood prediction tasks in Fenhe Jingle Basin and Luohe Luoshi Basin, improving the accuracy of LSTM model for flood prediction. Cui Zhen et al. [11] constructed a mixed model of GR4J (modele du Genie Rurala4 parametres Jour-nalier) and LSTM based on the conceptual hydrological model to forecast the inflow of land water reservoir. The results show that GR4J-LSTM has better flood forecasting performance than a single model. Compared with traditional hydrological models and shallow machine learning models [12,13,14], machine learning models show superior performance in flood prediction.

The Transformer model [15] proposed in 2017 supports parallel computing, is faster to train, and can simultaneously model long-term and short-term dependencies, and has shown good results in processing temporal data series [16,17]. In order to better realize time series prediction under different tasks, scholars in related fields have improved it in many ways. For example, The MCST-Transformer [18] (Multi-channel spatiotemporal Transformer) is used to predict the traffic flow, and the XGB-Transformer (Gradi-ent Boosting Decision) is used Tree transformer model [19] has been used for power load prediction, solving the problem of Transformer model being insensitive to local information in time series prediction tasks [20], and Transformer-based dual encoder model for the prediction of monthly runoff of the Yangtze River [21].

However, Transformers have three problems: high computational complexity, large memory usage, and low prediction efficiency. In 2021, Zhou [22] from Beihang University proposed the Informer model based on the classic Transformer encoder-decoder structure solve these problems to some extent. In this study, the data collection interval was long, and the Informer model was used for prediction and compared with other models.

2. Methodology

2.1. Development of the Approach

Figure 1 shows the process of the reservoir flood discharge prediction method. Firstly, the flood discharge data of Wan 'an Reservoir over the years were collected and pre-processed, including dividing the data into a training set (70%) and a prediction set (30%). Then the data data normalization is carried out, and the flood flow prediction is realized by using the three models of ANN, LSTM and Informer. NSE, RMSE, R2 and RE were calculated as evaluation criteria according to the flood flow prediction results. At the same time, the number of flood fields with flood peak gap less than 15%, total flood gap 15%, NSE greater than 0.8, and the maximum flood peak difference were counted as reference.

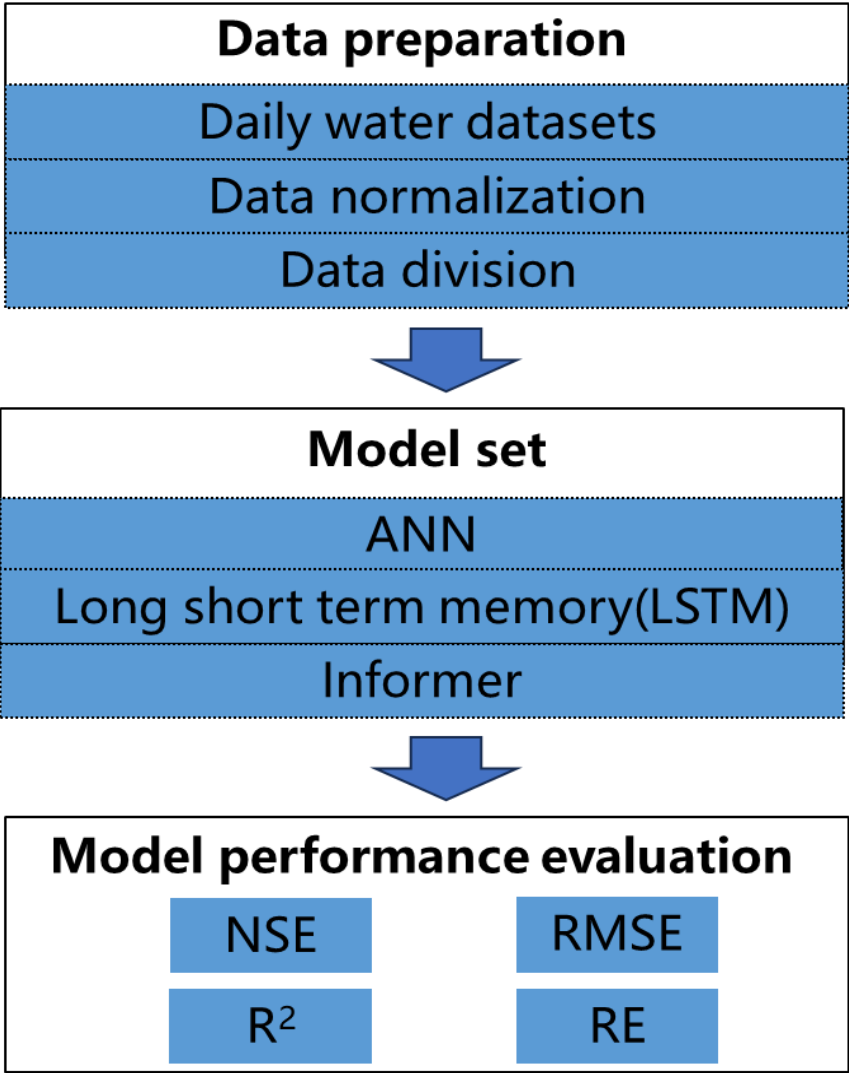


Figure 1. The flow chart of the proposed approach for forecasting.

2.2. *Deep Learning Models*

2.2.1. ANN

Artificial Neural Network (ANN) is a powerful tool for dealing with machine learning problems in the computer field. It is widely used in regression and classification problems. It simulates the operation principle of biological nerve cells, and forms a network structure of artificial neurons with hierarchical relationship and connection relationship. By means of mathematical expression, the signal transmission between neurons can be simulated, so as to establish a nonlinear equation with input and output relationship, and can be visualized through the network, we call it artificial neural network. Generally speaking, ANN can fit any nonlinear function through reasonable network structure configuration, so it can also be used to deal with nonlinear systems or black box models with complex internal expression.

2.2.2. Long Short-Term Memory

The LSTM network is a modified recurrent neural network proposed by Hochreiter and Schmidhuber. In recent years, the research on sequence prediction problem mainly focuses on the prediction of short sequences. LSTM network is used to conduct experiments on a set of data for short time series (12 data points, 0.5 days of data) and long time series (480 data points, 20 days of data). The results show that with the increase of sequence length, the prediction error increases

significantly, and the prediction speed decreases sharply. And LSTM models have several serious problems: High prediction error: When dealing with long time series data, the prediction error of LSTM model is high, which makes it perform unsatisfactory in some application scenarios; Slow prediction speed: The LSTM model has a relatively slow prediction speed, which is mainly caused by its internal complex calculation process; More model parameters: The LSTM model has more parameters to train, which makes it require more computing resources and time when dealing with large-scale data; Prone to mode switching: When dealing with non-stationary time series data, the LSTM model is prone to mode switching, which will lead to instability of the model prediction results.

2.2.3. Informer

Recent studies have shown that compared with RNN type models, Transformers show high potential in the expression of long-distance dependencies. However, Transformer has the following three problems: the quadratic computational complexity of self-attention mechanism: the time complexity and memory usage of each layer due to the dot product operation of self-attention mechanism; High memory usage problem: When stacking long sequence inputs, the stack of J encoder-decoder layers makes the total memory usage as, which limits the scalability of the model when accepting long sequence inputs. Efficiency in predicting long-term outputs: The dynamic decoding process of the Transformer, where the output becomes one after another, and the subsequent output depends on the prediction of the previous time step, results in very slow inference.

The authors of Informer target Transformer models with the following goal: Can Transformer models be improved to be more computationally, memory, and architecture-efficient while maintaining higher predictive power? To achieve these goals, this paper designs an improved Transformer based LSTF model, namely Informer model, which has three notable features: a ProbSparse self-attention mechanism, which can achieve a low degree of time complexity and memory usage; In the self-attention distillation mechanism, a Conv1D is set on the results of each attention layer, and a Maxpooling layer is added to halve the output of each layer to highlight the dominant attention, and effectively deal with too long input sequences. The parallel generative decoder mechanism outputs all prediction results for a long time sequence instead of predicting in a stepwise manner, which greatly improves the inference speed of long sequence prediction in Figure 2.

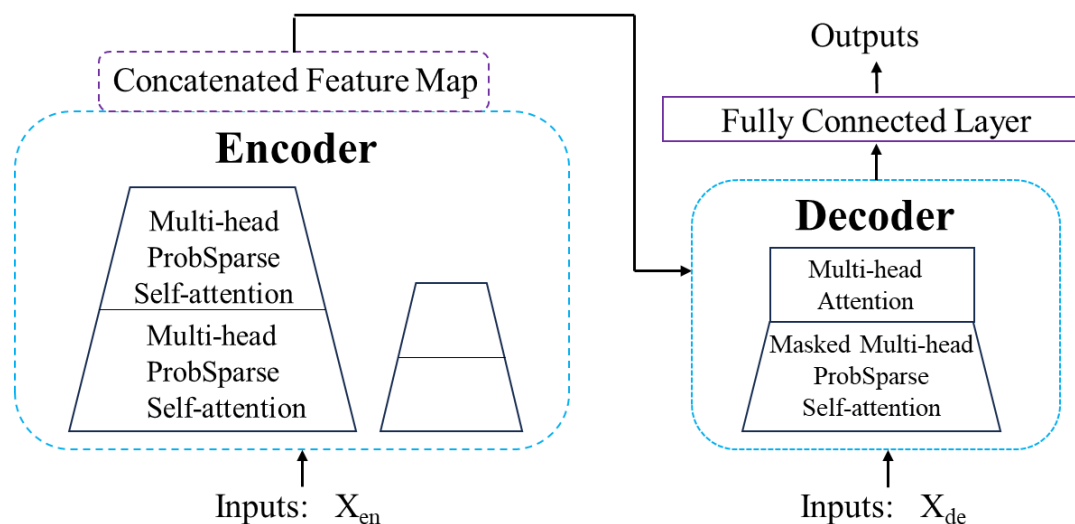


Figure 2. Conceptual diagram of Informer.

The model belongs to the Encoder-Decoder structure, in which the self-attention distillation mechanism is located in the Encoder layer. This operation essentially consists of several encoders and aims to extract stable long-term features. The network depth of each Encoder gradually decreases, and the length of the input data also decreases, and finally the features extracted by all encoders are

concatenated. It should be noted that in order to distinguish encoders from each other, the authors gradually decrease the depth of each branch by determining the number of repetitions of the self-attention mechanism of each branch. At the same time, in order to ensure the size of the data to be merged, each branch only takes the second half of the input value of the previous branch as input.

The self-attention distillation mechanism is an efficient method to improve the efficiency and accuracy of the model by training a smaller model to guide the learning of a larger model. The core idea of the self-attention distillation mechanism is to use the attention distribution of the large model to guide the training of the small model when training the small model, so that the small model can better imitate the large model.

The principle of the self-attention distillation mechanism is that it captures the global characteristics of the input data by computing the attention distribution of the input data. Then, these features are passed to the small model so that the small model can better imitate the large model.

The generative Decoder is located in the Decoder layer. In the past time series prediction, it is often necessary to perform multi-step prediction separately to obtain the prediction results of several future time points. However, with the continuous increase of the prediction length, the cumulative error will become larger and larger, resulting in the lack of practical significance of long-term prediction. In this paper, the authors propose a generative decoder to obtain the sequence output, which only requires a single derivation process to obtain the prediction result of the desired target length, effectively avoiding the accumulation of error diffusion during multi-step prediction.

Another innovation is the probabilistic sparse self-attention mechanism, which is proposed from the author's thinking about the feature map of the self-attention mechanism. The authors visualize Head1 and Head7 of the first layer of self-attention mechanism, and find that there are only a few bright stripes in the feature map. At the same time, a small part of the scores of the two heads have large values, which is consistent with the distribution characteristics of long-tailed data, as shown in the figure above. The conclusion is that a small fraction of the dot product pairs contribute the main attention, while the others can be ignored. According to this characteristic, the authors focus on the high-scoring dot product pairs, trying to calculate only the high-scoring parts in each operation of the self-attention module, so as to effectively reduce the time and space cost of the model. It allows the model to automatically capture the relationship between the individual elements of the input sequence, so as to better understand the input sequence. This mechanism enables the model to process all elements of the input sequence in parallel, avoiding the problem of computational order dependence in sequence models such as RNN/LSTM, and thus processing more efficiently.

2.3. Model Application

The water flow data of Wan 'an Reservoir is used as the data set for prediction. The dam site of Wan 'an Reservoir is located 2 kilometers upstream of Furong Town, Wan 'an County, Jiangxi Province, at 114°41 'east longitude and 26°33' north latitude. The upstream is 90 kilometers away from Ganzhou City and the downstream is 90 kilometers away from Ji 'an City, and the control basin area is 36900 square kilometers, seeing Figure 3.

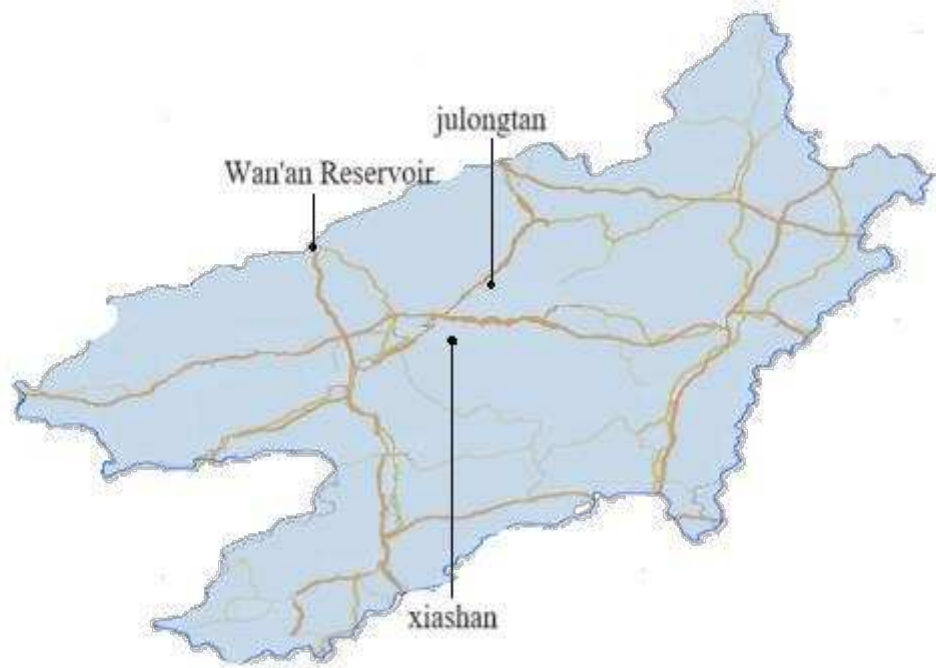


Figure 3. Map of the Wan'an Reservoir basin.

There are several hydrological stations and rainfall stations in the basin of Wan 'an Reservoir. In this paper, the rainfall information of five regions around the reservoir (PQJ1, PQJ2, PQJ3, PQJ4, PQJ5) and the flow information of Xishan hydrological Station and Julongtan hydrological station are selected to predict the flow of Wan 'an Reservoir. Rainfall and discharge data were collected in the region from May 17, 2014 to June 14, 2020, and the collection interval was 6 hours. See Figure 4.

FDNO	TM	PQJ1	PQJ2	PQJ3	PQJ4	PQJ5	xiashan	julongtan	OT
20140523	2014/5/17 14:00	3.6	6.1	0.3	6.8	1.8	1000	422	1680
20140523	2014/5/17 20:00	0.3	4.2	1.3	1.1	2	1000	426	1690
20140523	2014/5/18 2:00	0.3	0.2	1.8	0	0.4	1000	426	1703
20140523	2014/5/18 8:00	7.4	4	0.2	4	1	1245	424	1723
20140523	2014/5/18 14:00	1.3	6	0	0.2	0.2	1191	424	1760
20140523	2014/5/18 20:00	6.7	13.1	2	10.9	3.2	1107	421	1813
20140523	2014/5/19 2:00	17.5	7.1	6.2	0.7	0.6	844	442	1870
20140523	2014/5/19 8:00	0	0	0	0	0	1302	441	1973
20140523	2014/5/19 14:00	0.1	0.5	0	0	0	1272	561	2240
20140523	2014/5/19 20:00	17.8	13.1	34.7	5.9	11	1356	887	2603
20140523	2014/5/20 2:00	14.3	4.2	1	2.2	0.8	1491	1004	2927
20140523	2014/5/20 8:00	0.2	0.4	0	0	0	1593	877	3083
20140523	2014/5/20 14:00	0.4	0.1	0	0	0	2074	601	3167
20140523	2014/5/20 20:00	0	0	0	0	0	1997	425	3160
20140523	2014/5/21 2:00	0	0	0	0	0	1875	543	3083
20140523	2014/5/21 8:00	0	0	0	0	0	1516	610	2960

Figure 4. Dataset of Wan'an reservoir.

In the above figure, FDNO is the flood number, TM is the record date, PQJ1, PQJ2, PQJ3, PQJ4, and PQJ5 are the rainfall information of the five rainfall stations in the reservoir area, xiashan and julongtan are the discharge information of the two discharge stations in the reservoir area, and OT is the actual discharge of Wan 'an Reservoir.

Due to the noise and jitter of the initial data, the smooth function is used to smooth the data, and the value at time t after processing is the average value at time t-1, t and t+1. In addition, the flow at the flood peak needs to be recorded, so the flow at the flood peak is not smoothed.

In this experiment, considering the long sampling time, the step size is set to 4, 5, 6, and the output length is 1, respectively. When the step size is 4, four data are used to predict one data in the future.

The commonly used Loss functions in machine learning regression problems include mean absolute error (MAE), mean square error (MSE) and Huber Loss. MSE loss usually converges faster than MAE, but MAE loss is more robust to outliers, that is, less susceptible to outliers. Huber Loss is a Loss function that combines MSE and MAE and takes the advantages of both, also known as Smooth Mean Absolute Error Loss. The principle is simple: MSE is used when the error is close to 0 and MAE is used when the error is large.

In this paper, more attention is paid to the peak flood flow since it has the greatest impact on reality, and MAE, MSE and Huber Loss functions are used in this experiment to compare the prediction effect.

In the predictions of the two models, the common parameters are learning rate, training epochs, patience and batch_size, respectively 0.0005,100,10,10.

All numerical experiments in this study were implemented on a Windows system (CPU: Intel i7-12700H, GPU: NVIDIA GeForce RTX 3070 using Python (3.9) based on the Pytorch (1.8.0).

2.4. Model Performance Measures

Four indicators are finally used to measure the prediction results, namely NSE (Nash coefficient), R^2 (determination coefficient), RMSE (mean square error), and RE (mean difference).

Four hydrological concepts were used to measure the prediction results, namely, the number of fields with the peak discharge gap less than 0.15, the number of fields with the total flood gap less than 0.15, the number of fields with the Nash coefficient greater than 0.8 and the maximum flood peak gap. The calculation equations are as follows:

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

$$R^2 = \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (3)$$

$$RE = \frac{\sum_{i=1}^n y_i - \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i} \quad (4)$$

where x_i and y_i are observed and simulated values, respectively; \bar{x} and \bar{y} are average observed and simulated values, respectively and n is the number of samples.

NSE (Nash coefficient): It is used to verify the quality of hydrological model simulation results. The value of NSE is negative infinity to 1, and NSE is close to 1, indicating that the model quality is good.

R^2 (Coefficient of Determination): The proportion that reflects the total variation of the dependent variable can be explained by the independent variable through the regression relationship. It ranges from 0 to 1. The greater the coefficient of determination, the better the prediction effect.

RMSE (mean square error): It is the square root of the ratio between the squared deviation of the predicted value from the true value and the number of observations. It tells you how discrete a dataset is.

RE (mean difference): It is the ratio of the absolute error caused by a measurement to the measured (agreed) true value multiplied by 100%, which reflects the confidence of the measurement.

Peak discharge: the maximum instantaneous discharge in a flood discharge process, that is, the highest discharge on the flood process line. It may be the measured value, or it may be the calculated value using the water-flow relationship curve or the calculation value of the hydrodynamic formula.

Total flood water: The total amount of flood water flowing from the outlet section of the basin in a certain period of time. The total amount of a flood caused by a rainfall is often calculated in the forecast of rainfall runoff, which can be obtained from the area between the beginning time of the flood flow and the end time on the retreating section of the flood process line.

3. Result

Using thirty percent of the data as the prediction set, the different seq lengths and prediction accuracy metrics based on the loss types (MAE, MSE, and Huber) of the three deep learning models (ANN, LSTM, and Informer) are shown in Table 1.

Table 1. Wan'an reservoir forecast results statistical table (prediction).

Accuracy metrics	Loss	Seq len	ANN	LSTM	Informer
NSE	MAE	4	0.487	0.407	0.675
		5	0.492	0.500	0.639
		6	0.416	0.436	0.522
	MSE	4	0.628	0.711	0.746
		5	0.476	0.473	0.643
		6	0.494	0.304	0.665
	Huber	4	0.596	0.610	0.710
		5	0.641	0.665	0.719
		6	0.672	0.715	0.744
R ²	MAE	4	0.832	0.878	0.859
		5	0.836	0.882	0.851
		6	0.842	0.833	0.855
	MSE	4	0.847	0.895	0.873
		5	0.859	0.898	0.874
		6	0.827	0.883	0.889
	Huber	4	0.857	0.832	0.858
		5	0.842	0.877	0.884
		6	0.859	0.886	0.893
RMSE	MAE	4	754	782	615
		5	794	788	680
		6	905	993	790
	MSE	4	723	591	571
		5	755	789	665
		6	822	936	602
	Huber	4	789	704	590
		5	689	640	574
		6	707	603	576
RE(%)	MAE	4	15.3	18.8	12.5
		5	17.9	17.2	13.4
		6	21.1	20.9	19.5
	MSE	4	14.4	16.5	11.4

Huber	5	18.5	16.9	16.9
	6	16.6	20.1	14.3
	4	16.1	17.2	11.7
	5	17.5	15.9	13.5
	6	16.2	14.8	12.5

In addition, considering that this study is application-oriented, the trained model is used to predict all data sets, and the results are obtained and compared with the original data, as shown in Table 2.

Table 2. Wan'an reservoir forecast results statistical table.

Accuracy metrics	Loss	Seq len	ANN	LSTM	Informer
NSE	MAE	4	0.689	0.672	0.822
		5	0.652	0.566	0.743
		6	0.656	0.594	0.684
	MSE	4	0.674	0.734	0.775
		5	0.628	0.473	0.765
		6	0.632	0.587	0.705
	Huber	4	0.666	0.743	0.753
		5	0.612	0.747	0.792
		6	0.636	0.736	0.757
R ²	MAE	4	0.885	0.887	0.923
		5	0.842	0.910	0.898
		6	0.866	0.874	0.912
	MSE	4	0.863	0.910	0.916
		5	0.892	0.907	0.915
		6	0.832	0.902	0.899
	Huber	4	0.869	0.909	0.917
		5	0.886	0.910	0.915
		6	0.885	0.905	0.906
RMSE	MAE	4	575	560	449
		5	597	662	513
		6	612	604	540
	MSE	4	655	591	473
		5	667	789	479
		6	692	655	512
	Huber	4	753	521	512
		5	711	507	470
		6	759	517	497
RE(%)	MAE	4	16.6	12.6	9.9
		5	17.5	16.2	10.2
		6	17.1	13.4	12.2
	MSE	4	15.9	16.5	11.3

Huber	5	16.1	16.9	11.4
	6	17.5	15.8	11.2
	4	17.5	11.9	14.4
	5	17.7	11.8	10.7
	6	14.6	12.0	10.5

The number of the following four indicators are shown in Table 2: difference values less than 15%, flood peak difference less than 15%, NSE more than 0.8, max flood peak gap.

4. Discussion

In the Tables 1 and 2, the best prediction results of the three models are marked in red, and the best prediction results are based on the statistical results of Informer. The result with the highest NSE (Nash coefficient) is selected, that is, the step size is 4 and the loss function is MAE. Two floods (No. 1 and No. 2) are selected to draw the rainfall flow diagram, see Figures 4 and 5.

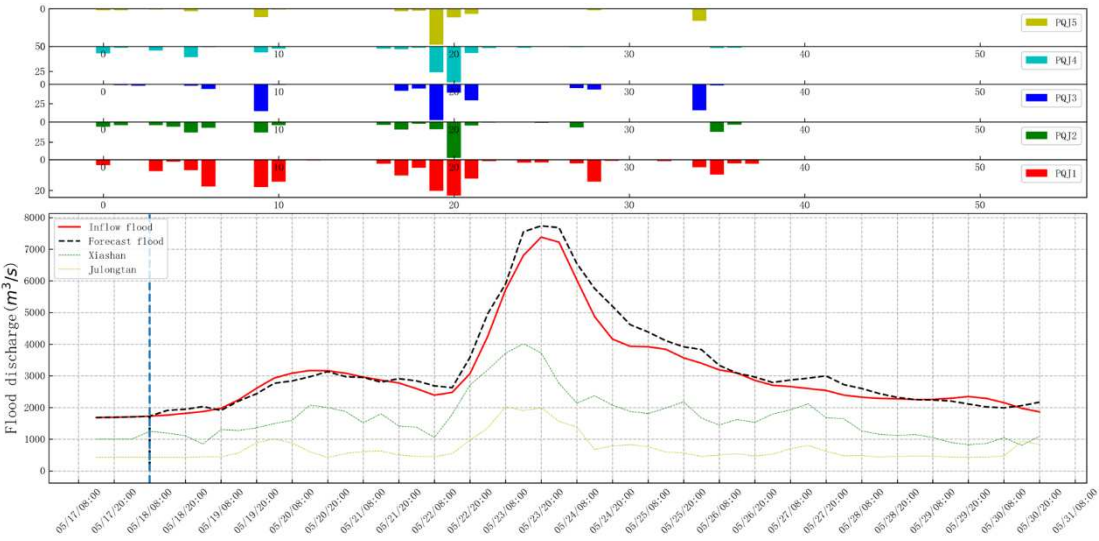


Figure 4. No. 1 Flood rainfall flow chart(Informer).

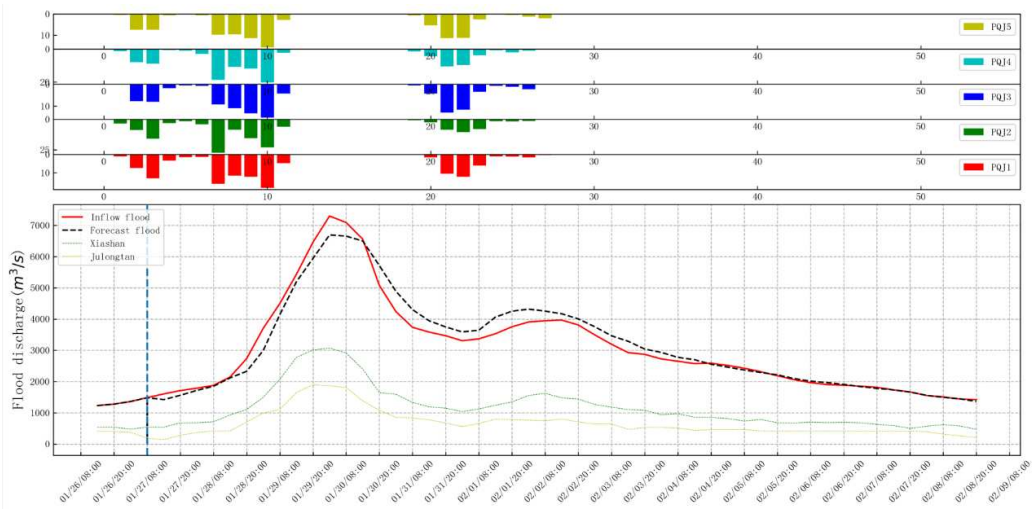


Figure 5. No. 2 Flood rainfall flow chart(Informer).

From Figures 4 and 5, it can be found that the predicted flood process line has a high degree of agreement with the actual flood process, and compared with the flood peak of Xiashan and Julongtan, the flood peak is lagged behind, which better reflects the spatiotemporal information of the flood.

According to Table 1, 30% of the data is taken as the prediction set, and the average value of the nine experimental results is taken as the result to draw the graph, as shown in Figure 6, where the horizontal axis is the four prediction indicators, and the vertical axis is the average value of the experimental results after appropriate processing.

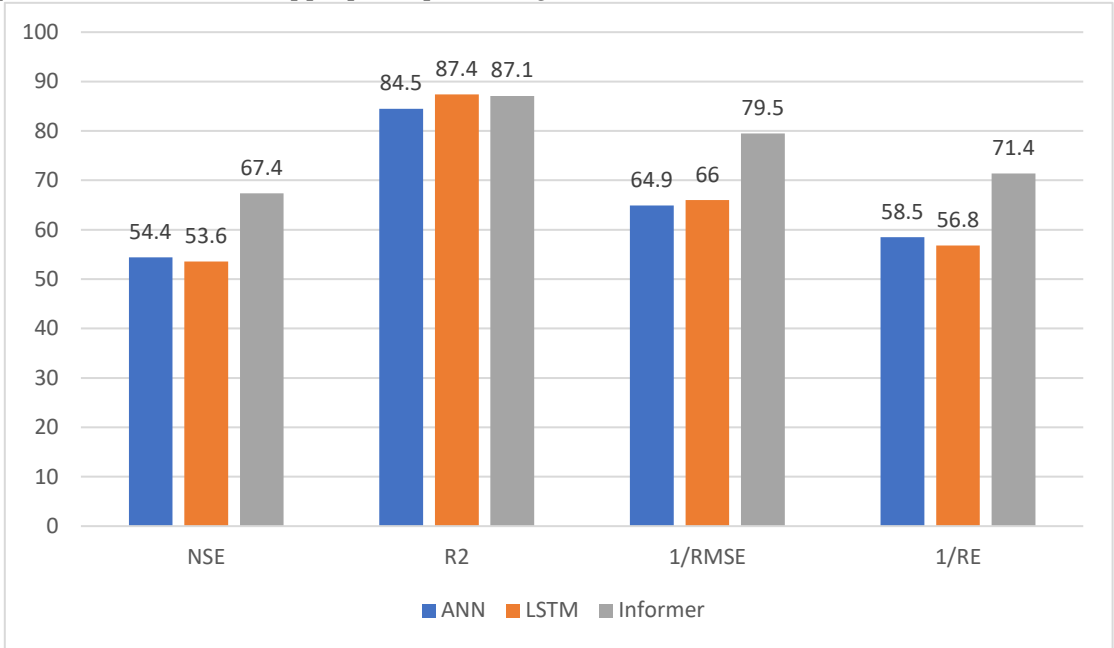


Figure 6. Prediction results indicator bar chart(prediction set).

According to Table 2, all data sets are used for prediction, and the bar charts of the four indicators are shown in Figure 7.

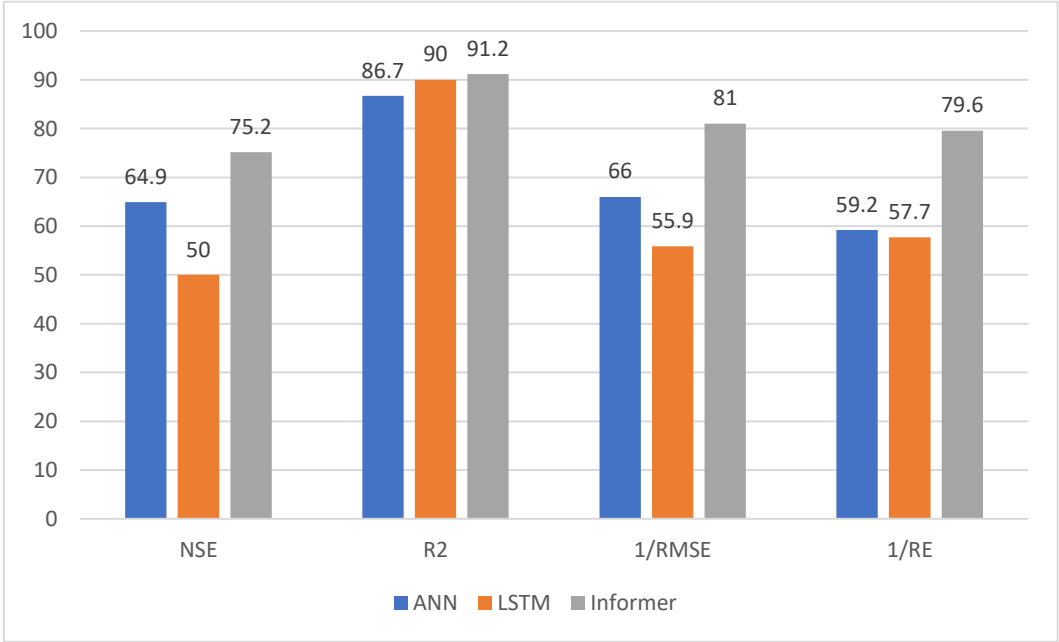


Figure 7. Prediction results indicator bar chart.

According to Table 3, for each hydrological index, the average value of the nine experimental results is taken as the result to draw a diagram, as shown in Figure 8, where the horizontal axis is the four hydrological indicators(A: difference values less than 15%, B: flood peak difference less than 15%, C: NSE more than 0.8, D: reciprocal of max flood peak gap) and the vertical axis is the average number.

Table 3. Hydrology table of Wan 'an reservoir prediction results.

Accuracy metrics	Loss	Seq len	ANN	LSTM	Informer
flood peak difference values less than 15%	MAE	4	18	23	19
		5	15	18	15
		6	17	17	24
	MSE	4	14	15	17
		5	14	13	17
		6	15	12	21
	Huber	4	13	15	13
		5	12	13	16
		6	15	13	21
flood peak difference less than 15%	MAE	4	19	17	23
		5	19	17	22
		6	17	14	20
	MSE	4	19	15	22
		5	15	12	21
		6	19	12	21
	Huber	4	14	14	16
		5	11	11	21
		6	13	15	21
NSE more than 0.8	MAE	4	14	14	18
		5	15	16	17
		6	16	13	19
	MSE	4	14	11	18
		5	13	8	18
		6	14	12	18
	Huber	4	15	15	16
		5	13	10	19
		6	11	12	20
max flood peak gap (%)	MAE	4	34.5	38.3	30.7
		5	36.7	27.4	31.5
		6	35.1	33.8	30.4
	MSE	4	47.3	57.2	35.1
		5	39.4	36.3	35.1
		6	42.1	49.2	32.7
	Huber	4	49.7	61.0	37.7
		5	42.4	45.3	29.7
		6	43.8	45.6	27.4

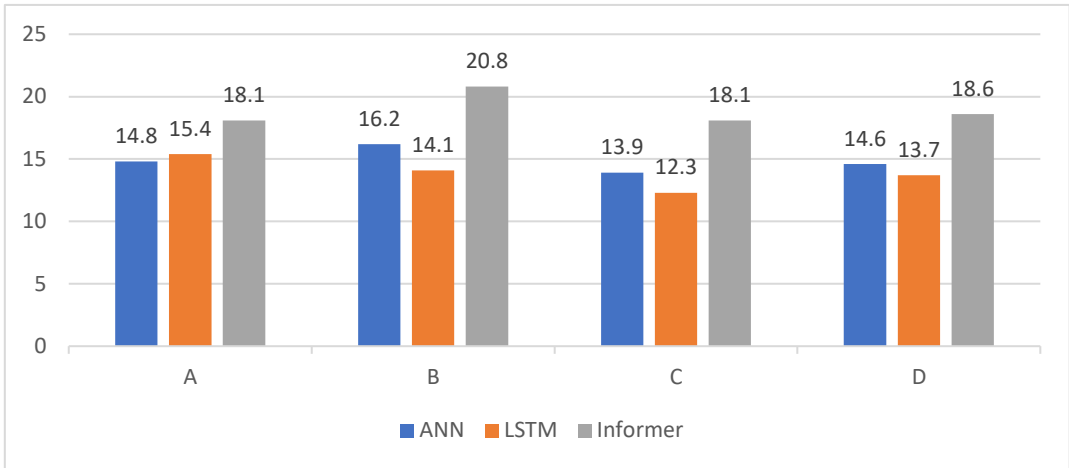


Figure 8. Histogram of hydrographic indicators.

5. Conclusions

In this study,Informer model and traditional hydrology methods were combined to study the effect of ANN,LSTM and Informer models in predicting the flood discharge of Wan 'an Reservoir, and NSE, RMSE, R² and RE were used to evaluate the accuracy and reliability of the prediction, all of which showed satisfactory results. Therefore, using machine learning method to predict flood flow is reliable. The results show that Informer is superior to ANN and LSTM in most cases.

The R² of LSTM model is better than Informer model except for 3 cases (loss=MAE, seq len=4),(loss= MAE, seq len=5),(loss=MSE, seq len=4). In other cases, the statistical indicators of the Informer model are better than the other two models. However, when predicting with the full dataset, the R² of the LSTM model (loss= MAE, seq len=5) is better than that of the Informer model in both cases (loss=MSE, seq len=6). In one case (loss=Huber, seq len=5), the LSTM model has better RE than the Informer model, and in the remaining cases, the Informer model predicts the best result. In terms of hydrological indicators, LSTM outperforms Informer in three cases, ANN outperforms Informer in four cases, and Informer predicts better in the remaining cases. Therefore, we can conclude that Informer performs better than ANN and LSTM models in predicting Wan 'an reservoir.

According to Figures 6, 7, 8, we can get the final conclusion that the average values of the four indicators (NSE, R², RMSE, RE) of the Informer model are better than those of ANN and LSTM under nine different conditions.

Acknowledgments: This paper was supported by the "Science and Technology + Hydrology" joint program project "Research and Application Demonstration of Reservoir Flood Forecasting and Scheduling Technology based on artificial intelligence Driven by big data" (2023KSG01007).

References

1. Mlv, M.; Todini, E.; Libralon, A. A Bayesian decision approach to rainfall thresholds based flood warning. Hydrol. Earth Syst. Sci. Discuss. 2006, 2, 413–426.
2. Bartholmes, J.C.; Thielen, J.; Ramos, M.H.; Gentilini, S. The european flood alert system EFAS-Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrol. Earth Syst. Sci. 2009, 13, 141–153.
3. Park, D.; Markus, M. Analysis of a changing hydrologic flood regime using the variable infiltration capacity model. J. Hydrol. 2014, 515, 267–280.
4. QIAO G C, YANG M X, LIU Q, et al.Monthly runoff forecast model of Danjiangkou Reservoir in autumn flood season based on PSO-SVRANN [J] . Water Conservancy and Hydropower Technology (Chinese and English) , 2021, 52 (4) : 69-78.
5. TAN Q F, WANG X, WANG H, et al. Application comparison of ANN, ANFIS and AR models in daily runoff time series prediction [J] .South-to-North Water Transfer and Water Conservancy Science and Technology, 2016, 14 (6) : 12-17+26

6. Salas, J.D.; Markus, M.; Tokar, A.S. Streamflow forecasting based on artificial neural networks. *Artif. Neural Netw. Hydrol.* 2000, 36, 23–51.
7. Tokar, A.S.; Johnson, P.A. Rainfall-runoff modeling using artificial neural networks. *J. Hydrol. Eng.* 1999, 4, 232–239.
8. LIU H. Research on flood classification and prediction based on neural network and genetic algorithm [J]. *Water Resources and Hydro-power Technology*, 2020, 51 (8) : 31-38.
9. LI D Y, YAO Y, LIANG Z M, et al. Hydrological probability prediction method based on variable decibel Bayesian depth learning [J/OL]. *Water science progress*: 1-10.
10. XU Y, HU C, WU Q, et al. Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation [J]. *Journal of Hydrology*, 2022, 608: 127-153.
11. CUI Z, GUO S L, WANG J, et al. Flood forecasting research based on GR4J-LSTM hybrid model [J]. *People's Yangtze River*, 2022, 53 (7) : 1-7.
12. OUYANG W Y, YE L, GU X Z, et al. Review of the progress of indepth study on hydrological forecasting II-Research progress and prospects [J]. *South-to-North Water Transfer and Water Conservancy Technology (Chinese and English)*, 2022, 20 (5) : 862-875.
13. YIN H, WANG F, ZHANG X, et al. Rainfall-runoff modeling using long short-term memory based step-sequence framework [J]. *Journal of Hydrology*, 2022, 610: 127-191.
14. LI B, TIAN F Q, LI Y K, et al. Deep-learning hydrological model integrating temporal and spatial characteristics of meteorological elements [J]. *Progress in Water Science*, 2022, 33 (6) : 904-913.
15. VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. *Neural Information Processing Systems*, 2017, 5998: 6-008.
16. SCHWALLER P, LAINO T, GAUDIN T, et al. Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction. [J]. *ACS Central Science*, 2019, 5 (9) : 1572-1583.
17. WANG W, XIE E, LI X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions [C] // 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 548-558.
18. ZHOU C H, LIN P Q. Traffic volume prediction method based on multi-channel transformer [J]. *Computer Application Research*, 2023, 40 (2) : 435-439.
19. DONG J F, WAN X, WANG Y, et al. Short-term power load forecasting based on XGB-Transformer model [J]. *Power Information and Communication Technology*, 2023, 21 (1) : 9-18.
20. LI S, JIN X, XUAN Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting [J]. *Neural Information Processing Systems*, 2019, 32. [21] LIU C, LIU D, MU L. Improved transformer model for enhanced monthly streamflow predictions of the Yangtze River [J]. *IEEE Access*, 2022, 10: 58240-58253.
21. ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting [C] // Proceedings of the 35th AAAI Conference on Artificial Intelligence, the 33rd Conference on Innovative Applications of Artificial Intelligence, the 11th Symposium on Educational Advances in Artificial Intelligence, Feb 2-9, 2021. Menlo Park: AAAI, 2021: 11106-11115.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.