Journal of Signal Processing Systems (2023) 95:351–362
https://doi.org/10.1007/s11265-023-01839-x

DA-Net: Dual Attention Network for Flood Forecasting

Qian Cheng[1] · Yirui Wu[2,3,7] · Aniello Castiglione[4] · Fabio Narducci[5] · Shaohua Wan[6,8]

Received: 8 November 2021 / Revised: 9 December 2022 / Accepted: 5 January 2023 / Published online: 24 January 2023

Abstract

Flood is difficult to predict due to its extreme runoff values, short duration and complex generation mechanism. In order to reduce the negative effects of flood disasters, researchers try to forecast flood by utilizing deep learning technology. Essentially, historical flood data can be regarded as sequential data with sets of flood factors. Facing challenges brought by inherent characteristics of flood forecasting, this paper proposes a dual attention embedding network, i.e., DA-Net, to achieve accurate prediction results. The proposed attention mechanism not only embeds a convolution self-attention module (CSA) on Temporal Convolutional Network (TCN) for description of local context information, but also constructs a Temporal-related Feature Attention (TFA) Module to assign time-varying weights for different features in a global sense. Specifically, CSA offers additional and local context information to help predict extreme runoff values even within a small period, meanwhile TFA improves global modeling capability of TCN for construction of data-driven generation mechanism in our method. Experiments on Changhua and Tunxi watershed dataset show the proposed method achieves superior prediction performance than current deep learning based methods.

Keywords Flood forecasting · Attention mechanism · Temporal convolutional network · Data-driven model

1 Introduction

Due to global warming and extreme climate, flood disasters occasionally happen and have brought great damage to human society, resulting in huge economic losses and safety hazards. It's essential to carry out accurate flood forecast to avoid the risk brought by flood disaster, which is one of the key issues in both machine learning and hydrological research.

Generally, current flood prediction models can be roughly divided into two categories, i.e., traditional physical models by simulating hydrological process[1, 2], and data-driven model based on machine learning algorithms [3, 4]. It's noted traditional physical models require hydrological experts to define model parameters,

✉ Shaohua Wan
shaohua.wan@ieee.org

Qian Cheng
chengqian-research@outlook.com

Yirui Wu
wuyirui@hhu.edu.cn

Aniello Castiglione
castiglione@ieee.org

Fabio Narducci
fnarducci@unisa.it

1 Jiangsu Hydraulic Research Institute, Nanhu Road, Nanjing 210017, Jiangsu, China

2 Key Laboratory of Water Big Data Technology of Ministry of Water Resources, Hohai University, Fochengxi Street, Nanjing 210096, Jiangsu, China

3 College of Computer and Information, Hohai University, Fochengxi Street, Nanjing 210096, Jiangsu, China

4 Department of Science and Technology, University of Naples Parthenope, Naples, Italy

5 Department of Computer Science, University of Salerno, Fisciano, Italy

6 Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, 518110, Guangdong, Shenzhen, China

7 Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Qianjin Road, Changchun 130015, Jilin, China

8 Key Laboratory of AI and Information Processing, Hechi University, Guangxi 546300, Yizhou, China

since performance of such models is quite sensitive to the pre-defined parameters. In other words, one specifically designed physical model by experts, which is limited to be applied in a well-examined watershed.

With significant progress of deep learning algorithm, data-driven models[5–9] have achieved remarkable development by exploring inherent relationship between run-off values and multiple flood factors. Essentially, data-driven models become the most promising ways to forecast floods, since large distributed sensors offer accessible means to flood-related factors, such as rain, upstream water and others. However, directly applying deep learning algorithms, such as LSTM, GRU, etc, for flood forecasting could result in low accurate performance.

Flood data is characterized by high-dimensional complexity, embedding highly non-linear relationship among flood factors. In other words, it's essential to build temporal relationship among flood factors in a local sense, thus better describing instantaneous characteristics of flood. Since traditional sequential data processing model pay little attention on modeling local temporal relationship, we adopt Temporal Convolutional Network (TCN) to extract temporal feature by enlarging the size of receptive field. On the basis of TCN, we propose CSA-TCN (short for Convolution Self-Attention-TCN) structure by embedding a self-attention mechanism, which successfully extracts local context information from sequential data with different receptive field sizes.

Moreover, different flood factors own different importance or weights on generation of flood in different process of flood. For example, soil moisture, one of the flood factors, has a great influence on the run-off values at the initial stage of flood. After middle stage of flood, soil can be filled with water and factor of soil moisture keeps consistent. In other words, we should assign a higher weight to factor of soil moisture at the beginning, meanwhile offering a quite small weight to it in middle or final stage of flood. Inspired by such fact, we propose a temporal-related feature attention (TFA) Module, which is capable to assign time-varying weights to high-dimensional features corresponding to different factors. Such weighting scheme focuses on the global hidden states of the LSTM in terms of temporal and feature dimensions, which offers a novel concept to build attention scheme for LSTM structure in global sense.

Based on the above consideration, we propose a dual attention network (DA-Net) for flood forecasting, which not only embeds a convolution self-attention module on TCN for description of local context information, but also constructs a temporal-related feature attention (TFA) Module to assign time-varying weights for different features. Both attention models serve as a part of the proposed dual attention mechanism in either local or global sense. The main contributions can be summarized as follows:

- We propose a lightweight CSA-TCN on the basis of TCN structure, which not only successfully extracts temporal features during a short-time period of floods, but also enhances local context information of feature by embedding a novel convolution self-attention mechanism.
- The proposed TFA module is capable to assign time-varying weights to different features, which serves as a weighting scheme for temporal and feature related information in a global sense.

The rest of this paper is arranged as follows. In the second part we introduce the correlation work for flood prediction methods. In the third part, details of our model structure and algorithm are presented. The fourth part introduces our comparative experiment, and the last part summarizes the paper.

2 Related work

This part will introduce relevant research results that inspire us to design models, mainly including data-driven models, temporal convolutional networks and attention mechanisms.

2.1 Data-Driven Model

With the advancement and development of deep learning technology[10–12] and big data technology[13–15] , more and more data-driven models are applied to flood prediction. Toth et al. [16] comprehensively analyzed and compared the Auto Regressive Moving Average Model(ARMA), artificial neural network (ANN) and nonparametric nearest neighbor method applied to 1h-6h short-term flood prediction. Based on statistical theory, Yu et al. [17] used a support vector machine approach to determine the lag time of the input variables using the concept of hydrologic response time to develop an effective flood prediction model. Biondi and De Luca [18] discussed the exploration of Bayesian systems applied to flood forecasting. The importance of using different diagnostic methods to analyze the quality of forecasts is emphasized. Ding et al. [19] used LSTM and attention mechanism to dynamically extract key feature vectors from various hydrological information to improve the accuracy rate in flood prediction. Song et al. [20] designed a multivariate time-stepped LSTM network for flood forecasting using information on the spatio-temporal dynamics of observed forecasted precipitation and early flows as input. Since current data-driven algorithms cannot achieve the expected detection effect in complex environments, such as background clutter, noise inundation or very small targets, Liu et al. [21] have designed an image enhancement-based detection algorithm for solving through detail enhancement and target expansion.

2.2 Temporal Convolutional Network

TCN is a popular model recently used to deal with time series tasks [22]. Its basic idea is to rely on a sample of a known sequence to model the target sequence that one wants to generate. Traditional convolutional neural networks are generally considered less suitable for modeling time-series problems, mainly due to the limitation of their convolutional kernel size, which cannot capture long-time dependency information well. TCN is based on two main principles: the output length of the model is kept consistent with the input length, and it needs to ensure that future data information during the computation does not leak into the past in advance.

To realize the first point, TCN uses a one-dimensional fully convolutional network architecture in which each hidden layer has the same length as the input layer, and zero-padding length is added to keep the length of subsequent layers the same as before. To realize the second point, TCN uses causal convolution, where the output at time t is only convolved with elements in the previous layer at time t or earlier. It differs from traditional convolutional neural networks in that it uses internally dilated convolutions.

TCN obtains different sizes of the receptive field size by changing the dilation factor d. Generally speaking, the larger the dilation factor, the larger the perceptual field can be obtained. Chen et al. [23] proposes a probabilistic prediction framework based on a temporal convolutional network that captures the dependencies of time series based on stacked residual blocks of dilated causal convolutions. Based on temporal convolutional network, Shen et al. [24] designed a general model for sequence modeling based on a temporal convolutional network, which is able to fully learn the features of sequential data with different interval lengths, and the structure of this model gives us great inspiration. Huang and Hain [25] builds on TCN by adding self-attentive attention blocks to highlight target-related features and mitigate the interference of irrelevant information. Since Correlation Filter based algorithms usually failed to track objects in complex environments, Liu et al. [26] propose a fuzzy detection strategy to prejudge the tracking result for temporal information modeling. Most recently, Liu et al. [27] propose a template update mechanism in temporal domain to improve the accuracy of visual tracking, in order to solve the problem of tracking failure in clutter background.

2.3 Attention Model

The attention mechanism can effectively help our model to focus on important information in the data. This mechanism is similar to the human visual system, which focuses selectively on the present information. Attention mechanism has been widely used in sequence tasks such as flood prediction [28] and speech recognition [29].

Attention mechanism can be strictly divided into Soft attention [30] and Hard attention [31]. Soft attention will pay attention to all the data, weighting each key after weight calculation. Hard attention will filter out some of the unqualified attention and focus directly on certain keys. Fan et al. [32] combines high-performance multi-level forecasting with an interpretable understanding of temporal dynamics to improve prediction accuracy by learning underlying patterns in historical data. Zang et al. [33] adds a temporal attention mechanism to the video streaming task, which effectively improves the recognition performance of video images through a time-weighted attention model. Shih et al. [34] focused on the design of TPA-LSTM (Temporal Pattern Attention-LSTM) that can weigh the selection of relevant variables at different times. In addition, Vaswani et al. [35] abandoned the traditional CNN and RNN models, and the entire network structure consists entirely of attention mechanism. Li et al. [36] improved the Transformer for time series prediction by reducing the memory complexity and by convolutional design so that the model can capture the local context.

3 The Proposed Method

In this section, we introduce the overall architecture of DA-Net, the structure design of CSA-TCN and TFA module in detail.

3.1 Overview

We first introduce the overall calculation flow of DA-Net in a general way. Fig. 1 shows the internal details of the DA-Net. The input data I of our model is specific to four hydrological factor sequences, which are the current evaporation, former rainfall, former rainfall factor and former flood flow. The objective of our model is to make short and medium term flood flow forecasts from 1 to 6 hours.

Our CSA-TCN structure is mainly composed of temporal convolutional network and convolution self-attention. The core part of temporal convolutional network is dilated causal convolution. The receptive field of the dilated convolution depends on the network depth , the filter size k and the dilation factor d. For the factor sequences under different time intervals after TCN calculation, we use the convolutional self-attention to enhance their local contextual information, which can fully strengthen the association information on each time interval. The main calculation process of convolution is as follows:

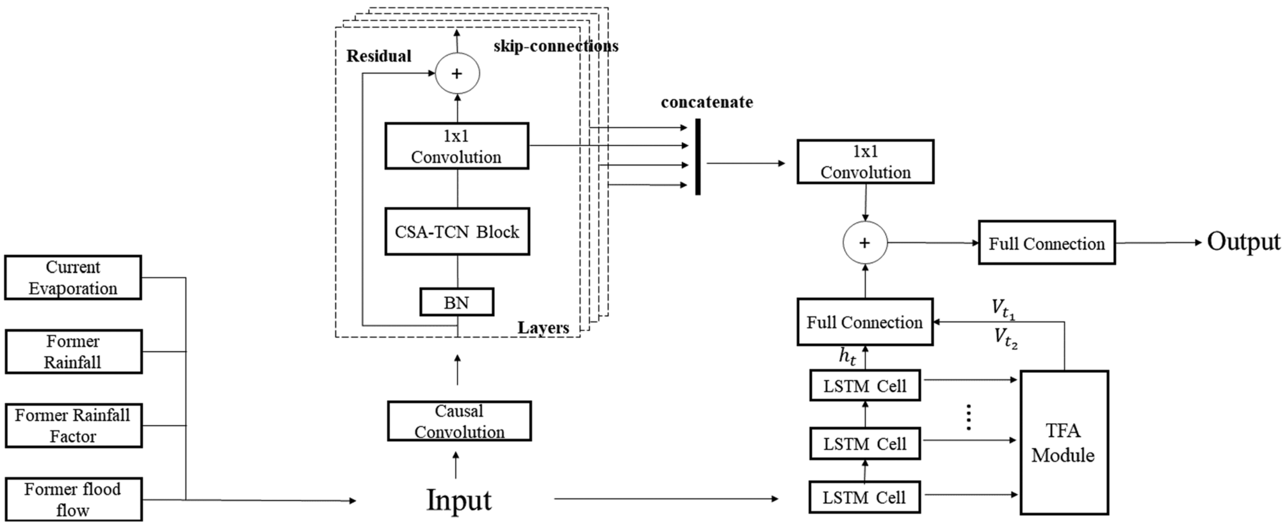

Figure 1 Structure design of the proposed DA-Net (Dual-attention Network).

$$y_c = f_c(TC(I)) \tag{1}$$

where I is the input sequence, TC is the function of temporal convolutional network, and f_c is the function of CNN self-attention.

In addition, we use the design of residual blocks and skip connection to fuse the features of different sequences after convolutional self-attention calculation, so that local information at different levels of time intervals can be obtained in the global sense.

Our TFA module focuses on the temporal and feature dimensions based on the hidden state matrix of LSTM network. By focusing on the different weights of multivariate features at each point in time and the influence weights of different time series at past points in time, the module not only assigns time-varying weights to different features, but also reduces the interference of irrelevant noise information. At the same time, our TFA module also retains the temporal information of the original hidden state matrix in the LSTM for time-weighted attention calculation. The two complement each other effectively to improve the prediction accuracy. The calculation process is as follows:

$$y_d = l_c\left(V_{t_1}, V_{t_2}, h_t\right) \tag{2}$$

where V_{t_1} is the context vector that the module is concerned about in the feature dimension,V_{t_2} is the context vector that the module is concerned about in the temporal dimension, h_t is the hidden state of LSTM, and l_c is the linear layer calculation after vector splicing.

Our model uses the Back Propagation Through Time (BPTT) algorithm to minimize the loss function and uses MSE as our loss function.The loss function is defined as:

$$loss_{min} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2 \tag{3}$$

where $\bar{y}_i$ is the true flood flow, and y_i is the predicted value of our model.

3.2 Design of CSA-TCN Structure

We add the convolutional self-attention mechanism on the basis of temporal convolutional network. Temporal convolution network is composed of multi-level one-dimensional dilated convolution. Since different dilated factors in each layer can capture data features at different scales, flood feature information at different time intervals can be learned. Fig. 2 shows the structure of the CSA-TCN structure.

For input data X, the calculation formula F of dilated convolution is as follows:

$$F_{(X)} = \sum_{i=1}^{k} f(i) \cdot X_{t-di} \tag{4}$$

d is the size of the dilated factor and f is a filter of size k. $ReLU$ is used as the activation function, and the residual is used to avoid gradient. The multi-head self-attention mechanism calculates the weight of three vectors K, Q and V respectively after the input changes through the linear layer.

$$\text{head}_i = \text{softmax}\left(\frac{Q_i K_i^{\top}}{\sqrt{\tilde{d}_k}} \cdot \text{ mask }\right) V_i \tag{5}$$

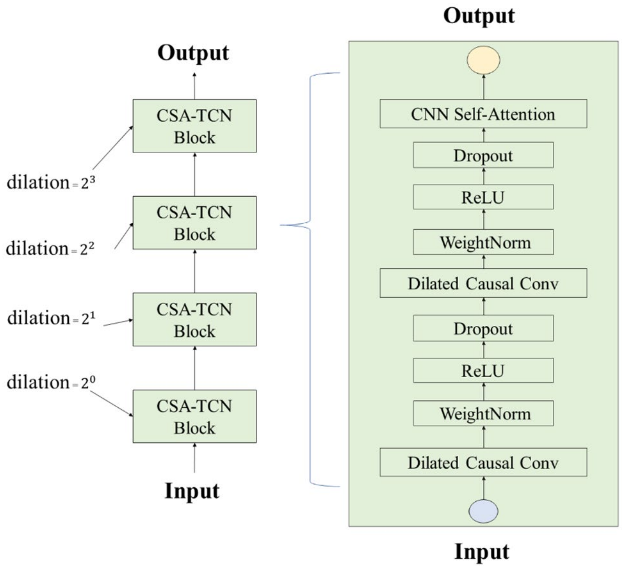

Figure 2 Illustration of the proposed CSA-TCN structure.

where Q_i, K_i and V_i are obtained by linear transformation of the input vector.

In order to establish the temporal relationship between the flood factors in a local sense, we no longer use the linear transformation method for the feature sequences after obtaining the results of the dilated causal convolution calculation. $Conv1D$ can make Q and K in attention computation have local context information compared to the linear layer.

The results of the dilated convolution at each layer in TCN will be output after the convolution from the attention weighting calculation, and the output results will be passed into the dilated convolution at the next layer as input. In this way, convolution self-attention calculation can be carried out for feature sequences at different scales. We believe that this method can effectively capture local context information while enlarging the receptive field.

Our model calculates the similarity of different feature sequences through Softmax scoring. In addition, the calculation process of self-attention mechanism will do attention calculation for the whole sequence. But in the process of prediction, we do not want the future information to be leaked. So we refer to the Mask matrix method in Transformer to ensure that the information after time t cannot be seen.

Due to its successive multiplications in modeling, gradient of RNN is dominated by dependence learning within a short distance, thus being difficult to learn the dependence within a long distance and leading to gradient disappearance. Gradient disappearance of RNN would result in non-updating of parameters in training process, where the trained RNN model could obtain low performance in prediction tasks.

In order to solve the problem of gradient disappearance encountered by RNN, LSTM improves the memory attention of long-distance information. But on the other hand, the retained long-distance information itself is not all valuable. LSTM has multiple gradient propagation paths through gating design, the structure of LSTM determines that the same weight is shared in each time step. Our temporal-related feature attention (TFA) module focuses on the key time step information and ignores the useless part from two directions of temporal and feature, so as to solve gradient disappearance by adding additional weights in training process. After times of trials, we observe that the involve of TFA leads to stable and improved performance of the proposed prediction model. Fig. 3 shows the structure of the TFA in detail.

The classical LSTM network is designed to solve the gradient vanishing and gradient explosion problems when RNN is applied in the long sequence training process. By adding additional attention mechanism, the weight of important information can be increased more effectively. The traditional temporal attention mechanism is based on the position of different time steps to carry out effective information weighting calculation. As shown in Fig. 3, the time is the core of the model attention mechanism. A hidden state matrix H is obtained by LSTM calculation, $H \in \mathbb{R}^{m*t}$. Convolution operation is performed on the row vectors of the hidden state matrix from time 1 to time $t-1$. The specific operation is as follows:

$$H^P = \{h_1, h_2, \dots, h_{t-1}\} \tag{6}$$

$$H^C = \left\{h^C_{1,1}, \quad h^C_{i,j}, \cdots, h^C_{m.k}\right\} \tag{7}$$

The H^P matrix can be calculated by convolution to obtain the H^C matrix. $h^C_{i,j}$ means to convolve the ith dimension of the hidden state matrix H^P with the jth convolution check. $H^C \in \mathbb{R}^{m*k}$, which m said after convolution the vector dimension on each time step, and k represents the different convolutional kernels in the convolutional network.

Our feature attention mechanism focuses on the changes of different features in different time steps at the time of h_t, so we can get the attention size of the ith hidden feature vector in the H^C matrix:

$$a_i = \text{sigmoid}\left(\left(H^C_i\right)^T W h_t\right) \tag{8}$$

Among them, $W \in \mathbb{R}^{k*m}$ for the parameters of the corresponding weighting matrix, attention size a_i as in the sigmoid activation function under the action of the corresponding weights, the context vector V_{t_1} is obtained by weighting the attention score.

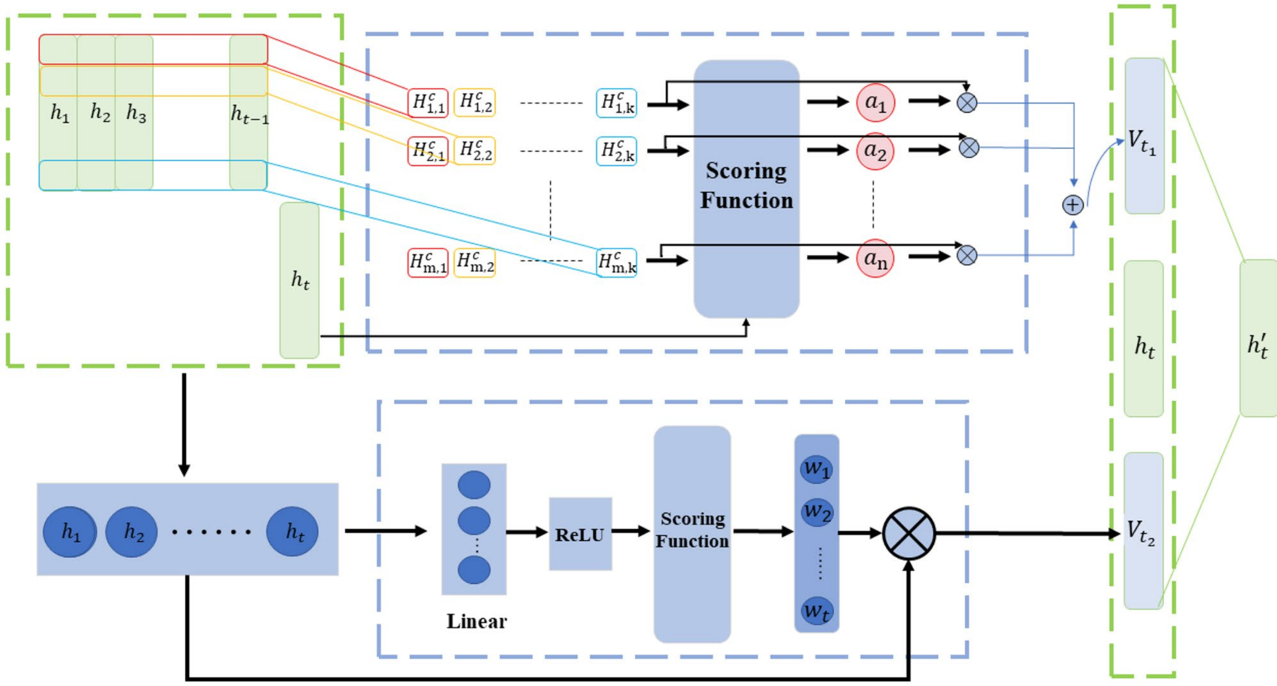

Figure 3 Illustration of the proposed TFA Module.

$$V_{t_1} = \sum_{i=1}^{m} a_i H_i^C \tag{9}$$

By dynamically calculating the weight of different flood factors in the corresponding time step, our model can well discover the internal relationship between different flood factors in each time step, and at the same time reduce the interference of non-critical factors to the prediction to the greatest extent.

In addition, we calculate the temporal attention of the hidden state matrix H to get the time weight. Firstly, we transform H through a linear layer to get the matrix H'. After the matrix H' is obtained, the weight of temporal attention is calculated. The specific process is as follows:

$$w_a = \mathrm{Softmax}\left(\mathrm{ReLU}\left(H'\right)\right) \tag{10}$$

where w_a is the corresponding temporal attention weight obtained by calculating the hidden state matrix H' through activation function Relu and Softmax function.

Then, the context vector V_{t_2} can be obtained through the product operation of w_a and the hidden state matrix H, where $V_{t_2} \in R^{n\times 1}$. The specific process is as follows:

$$V_{t_2} = \sum_{i=1}^{t} h_i w_i \tag{11}$$

After splicing V_{t_2} and the context vector V_{t_1} and h_t obtained previously, the final vector h_t was obtained through a full connection layer. The specific process is as follows:

$$h'_t = \mathrm{Linear}\left(\mathrm{concat}\left(V_{t_1}, h_t, V_{t_2}\right)\right) \tag{12}$$

The vector h'_t is obtained from the context vector of temporal and feature dimension respectively and from the hidden state h_t, which pays attention to the key factors while retaining the original time sequence information, and the two complement each other to effectively improve the prediction accuracy.

4 Experiments

4.1 Dataset and Measurement

We show the map of Tunxi and Changhua basin in Fig. 4. The hydrological station of Changhua Basin in Zhejiang Province has 31 hydrological flood field data measured from 1998 to 2010. Changhua Basin is located in the upper reaches of Fenshui River, with a total length of 96 kilometers and a basin area of 1376 square kilometers. The basin is a mountain stream river, the flood has the characteristics of sharp rise and fall, and the flood peak is high and large. A total of 8,553 data were collected from Changhua River Basin. We randomly selected 1,000 samples as the test set

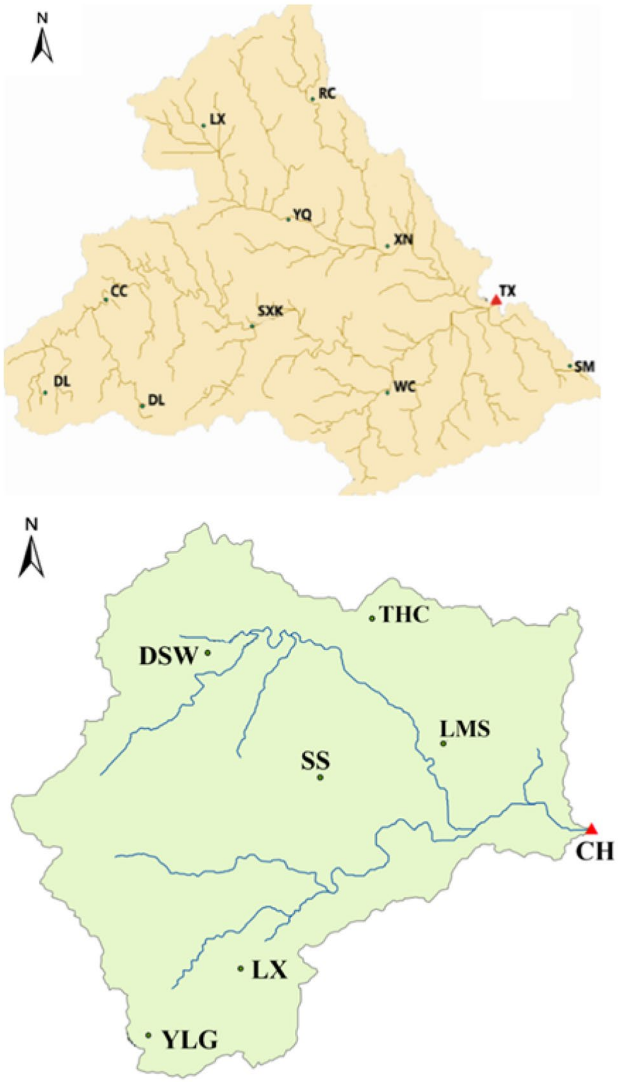

Figure 4 Map of each measuring station in Tunxi and Changhua, including rainfall station and river flow mapping station.

and the rest as the training set. Our goal is to predict the flood flow one to six hours after the current moment.

Tunxi Hydrological Station has 30 hydrological flood field data from 1981 to 2003. The Tunxi Valley is in Zhejiang Province. The catchment area of the basin is 2696.76 square kilometers, located in the subtropical monsoon climate zone. The average annual precipitation in this basin is as high as 1600mm. The annual distribution of precipitation is very uneven. It is rainy from April to June every year, accounting for half of the annual precipitation, and it is prone to flood disaster. Zhejiang Province is located in the eastern region, where there is plenty of rain and the rainfall lasts for a long time, and the flood process lasts for a long time, usually about a week. Due to the long duration of the flood process, the randomness of the rainfall regime changes during the period is great, so the regularity of the flood pattern is difficult to grasp. There are 3765 pieces of Tunxi Basin data, and 500 samples are randomly selected as the test set and the rest as the training set. Our goal is to predict the flood flow one to six hours after the current moment.

We conducted the model experiment and evaluation in Changhua and Tunxi watershed, and we chose RMSE and MAE as our evaluation indexes.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - q_j\right)^2} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - q_j| \tag{14}$$

where j is the index for test samples, n is the number of test samples, y_i is the predicted result, and q_i is the groundtruth.

4.2 Implementation Details

All experiments are carried out on a Linux server equipped with 2.10 GHz 8-core Xeon CPU, 60GB RAM and Nvidia GeForce GTX 1080 Ti. A total of 3 layers of dilated convolution units are set in our network, and the dilated factors are 1,2,4 respectively. The hidden layer unit in LSTM is set as 128. The input sequence length is 32, the batch size is 64, the learning rate is 0.001, and the epoches design is 1000 iterations.

4.3 Ablation Experiments

In order to verify the effectiveness of our idea, we first conducted ablation experiments on Changhua and Tunxi data sets, and the experimental results are shown in Tables 1 and 2. It's noted that RMSE should be as low as possible, since we aim to simulate flood with few errors. T+1 means our prediction is designed to predict run-off value after one hour.

Table 1 Comparison of RMSE by conducting ablation experiments on Changhua dataset.

Method	T+1	T+2	T+3	T+4	T+5	T+6	Average
Basic	26.94	52.01	63.21	71.45	81.90	92.98	64.74
Basic with CSA-TCN	28.86	46.65	59.43	71.10	82.39	93.70	63.69
Basic with TFA	28.52	46.29	58.25	69.53	81.27	93.91	62.96
Our Method	27.64	46.91	55.27	64.76	75.62	89.80	60.18

Table 2 Comparison of RMSE by conducting ablation experiments on Tunxi dataset.

Method	T+1	T+2	T+3	T+4	T+5	T+6	Average
Basic	35.32	59.39	90.67	114.94	142.35	164.23	101.15
Basic with CSA-TCN	31.76	57.83	85.36	111.93	139.47	165.48	98.64
Basic with TFA	29.28	55.76	81.10	107.98	138.62	167.14	96.65
Our Method	36.14	54.80	80.88	106.27	135.90	162.57	96.09

Table 3 Comparison of rmse on changhua dataset with several comparative methods.

Method	T+1	T+2	T+3	T+4	T+5	T+6	Average
FCN [37]	37.30	41.88	56.87	84.32	110.80	125.3	74.41
LSTM [38]	54.50	57.67	61.93	71.41	83.32	96.12	70.83
SeriesNet [39]	26.94	52.01	63.21	71.45	81.90	92.98	64.74
CA-LSTM [40]	28.53	47.01	57.59	66.44	77.65	89.97	61.20
ST-GCN [41]	60.23	65.15	70.07	71.45	72.91	73.93	68.96
Our Method	27.64	46.91	55.27	64.76	75.62	89.80	60.18

Table 4 Comparison of MAE on Changhua dataset with several comparative methods.

Method	T+1	T+2	T+3	T+4	T+5	T+6	Average
FCN [37]	14.88	22.21	28.59	34.01	38.86	42.92	30.25
LSTM [38]	26.08	27.14	29.71	33.68	37.30	40.77	32.45
SeriesNet [39]	12.26	19.20	25.10	30.60	35.55	39.74	27.08
CA-LSTM [40]	9.84	18.75	25.58	31.41	36.46	41.69	27.29
ST-GCN [41]	19.04	23.59	29.98	30.18	30.15	30.24	27.20
Our Method	10.61	18.42	23.80	29.88	32.42	38.17	25.72

Table 5 Comparison of RMSE on Tunxi dataset with several comparative methods.

Method	T+1	T+2	T+3	T+4	T+5	T+6	Average
FCN [37]	48.78	68.93	93.25	120.85	148.91	177.47	109.70
LSTM [38]	40.22	67.47	90.69	117.56	146.28	170.91	105.52
SeriesNet [39]	35.32	59.39	90.67	114.94	142.35	164.23	101.15
CA-LSTM [40]	31.28	58.81	85.04	110.32	139.02	165.28	98.29
ST-GCN [41]	51.28	70.78	95.44	104.74	113.86	125.45	93.59
Our Method	36.14	54.80	80.88	106.27	135.90	162.57	96.09

Table 6 Comparison of MAE on Tunxi dataset with several comparative methods.

Method	T+1	T+2	T+3	T+4	T+5	T+6	Average
FCN [37]	31.48	43.06	55.32	71.21	85.56	101.46	64.68
LSTM [38]	23.08	36.25	47.70	63.30	78.76	91.95	56.84
SeriesNet [39]	19.17	30.37	52.37	65.31	80.60	90.36	56.36
CA-LSTM [40]	15.62	30.21	44.62	58.93	73.61	87.73	51.79
ST-GCN [41]	47.08	48.85	21.44	50.60	70.49	76.96	52.57
Our Method	24.26	28.37	40.28	54.66	70.76	86.48	50.80

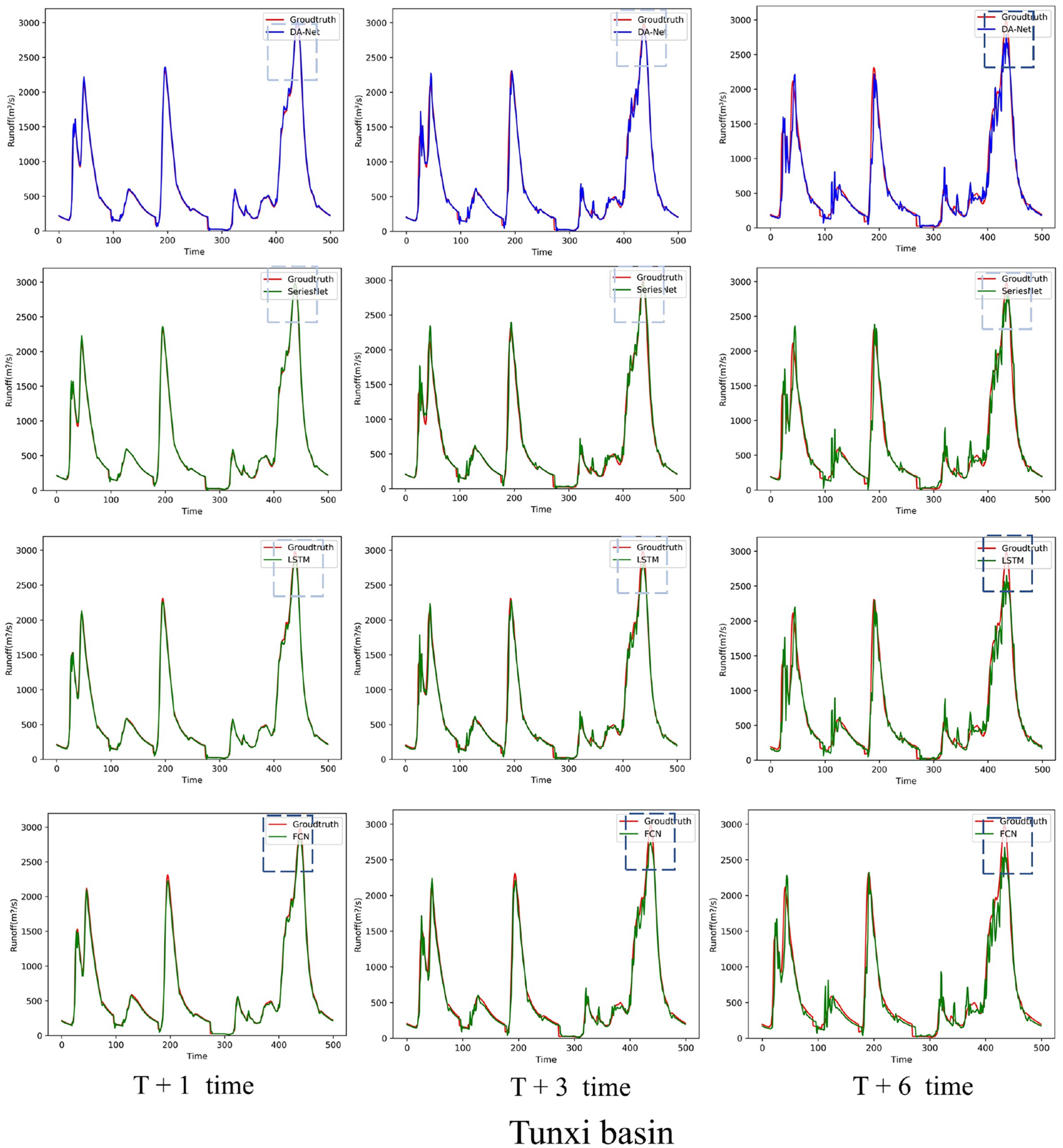

Figure 5 Comparison on Tunxi Basin among the ground truth flow rates (first row) and predicted flow rates computed by our method (second row), FCN (third row), and CA-LSTM (fourth row). Note that the rectangles indicates several obvious comparison results.

On the basis of basis, i.e., SeriesNet model, we added the designed CSA-TCN structure and TFA module in turn to experiment. The results of experiment show that our modules have improved to a certain extent in the prediction from T+2 to T+6. In the experiment of Tunxi basin, our module has the best performance in the middle period of time. It's not difficult to see from the experimental results that the proposed model has a good improvement effect on the prediction results, and is not weakened by the growth of the prediction time. After comparing average performance between Basic and basic with CSA-TCN or Basic with TFA, we can find both modules improve the prediction accuracy,

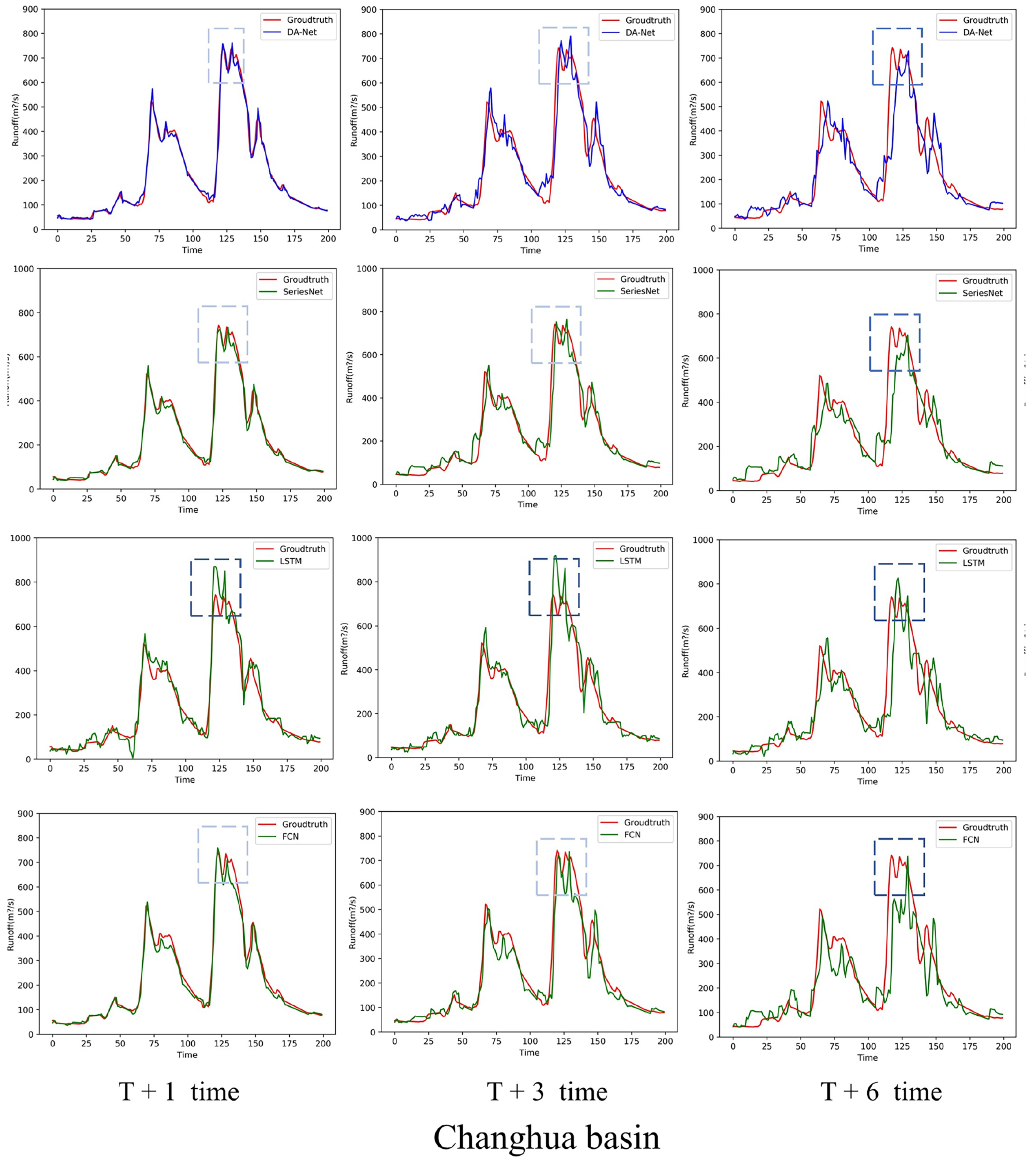

Figure 6 Comparison on Changhua Basin among the ground truth flow rates (first row) and predicted flow rates computed by our method (second row), FCN (third row), and CA-LSTM (fourth row).

which proves the role of both modules. Our dual attention network consists of CSA-TCN and TFA, which is superior to both basic with CSA-TCN and Basic with TFA. The reason is that our method properly embed both modules in our network, thus working closely to improve prediction.

4.4 Experiment Analysis

As shown in Tables 3 and 4, RMSE and MAE measurements are used to compare the proposed method with the implementations of CA-LSTM [40], SeriesNet [39], LSTM [38], FCN [37] and ST-GCN [41], where CA-LTSM tries to build attention mechanism for flood prediction offering innovations to inspire later researchers, and ST-GCN encodes geometrical information of watershed into modeling. Specially, our method is worse than SeriesNet in the first hour and slightly less effective than FCN in the second hour. Our method has a greater advantage in the last four hours of prediction. And the average RMSE value of our method is 60.18, which is the lowest prediction error among our comparison methods, expect for RMSE on Tunxi dataset. In the metric comparison of MAE, our method performs slightly lower than CA-LSTM at the first hour, but has better performance in the prediction from the second hour to the sixth hour. Moreover, we find ST-GCN generally performs stably with increasing of prediction hours, due to the encoding of geometrical information by graph neural network. In fact, the sufficient knowledge on watershed topography could greatly contributes to perdition with large hours, which could be our further improvement.

As shown in Tables 5 and 6, our method is worse than SeriesNet and CA-LSTM in the first hour and it has a greater advantage in the last five hours of prediction. The average RMSE value of our method is 96.09, which has the lowest prediction error among our comparison methods. In the comparison of the MAE metric, our method slightly underperforms in the first hour and it performs better in the second hour to the sixth hour of prediction. It's noted that the proposed method achieved 1.53s computing time for one sample prediction, and required 20.8 minutes on training dataset (including both tunxi and Changhua basins).

Figures 5 and 6 shows the actual prediction results of our comparison experiments on the two datasets, i.e., Tunxi and Changhua Basins. Observing the flooding process in the Changhua basin, we can find that there are two distinct flood peaks, with an average height between 600 and 700 cubic meters. Although the prediction accuracy of various methods inevitably decreases with time. Compared to other methods, our model is more stable over time and performs better in predicting flood peaks.

It is not difficult to see the Tunxi watershed flood flow fluctuates larger. The flood had three distinct peaks, with an average height of between 2,000 and 3,000 cubic meters. It can be seen that the overall fitting of our model is relatively smooth, and certain accuracy is guaranteed when predicting the height of flood peak in the first three hours. There is a significant decrease in the prediction accuracy at the sixth hour, but our method still has the best overall results.

5 Conclusion

In this paper, a flood forecasting network DA-Net based on dual attention mechanism is proposed on the basis of SeriesNet. Our model not only embeds a convolutional self-attention module on Temporal Convolutional Network(TCN) for enhancing local contextual information, but also constructs a time-dependent feature (TFA) module to assign time-varying weights to different features. These two components improve the short-term flood prediction accuracy on small and medium-sized rivers by complementing each other with global perspective and local context information. It's noted we generally name rivers with watershed areas smaller than 3000 square kilometers as small-sized rivers. Through the comparative experiments on Changhua and Tunxi river basins, it is proved that our model is superior to several existing methods, and our model has higher prediction accuracy. In the future research work, facing some basins with small data, we will try to add the idea of small sample learning to improve and optimize the model of flood prediction method.

Acknowledgements This work was supported by National Key R&D Program of China under Grant No. 2021YFB3900601, National Natural Science Foundation of China under Grant No. 62172438, the Fundamental Research Funds for the Central Universities under Grant B220202074, and the Fundamental Research Funds for the Central Universities, JLU.

Funding Laboratory of AI and Information Processing (Hechi University), Education Department of Guangxi Zhuang Autonomous Region under Grant 2022GXZDSY014, and Research Funding Project of Jiangsu Hydraulic Research Institute No. 2022Z028.

Data Availability The datasets generated during and analysed during the current study are not publicly available due to privacy reasons, but are available from the corresponding author on reasonable request.

References

1. Pierini, N. A., Vivoni, E. R., Robles-Morua, A., Scott, R. L., & Nearing, M. A. (2014). Using observations and a distributed hydrologic model to explore runoff thresholds linked with mesquite encroachment in the sonoran desert. *Water Resources Research, 50*, 8191–8215.
2. Rogger, M., Viglione, A., Derx, J., & Blöschl, G. (2013). Quantifying effects of catchments storage thresholds on step changes in the flood frequency curve. *Water Resources Research, 49*, 6946–6958.
3. Ding, Y., Zhu, Y., Wu, Y., Jun, F., & Cheng, Z. (2019). Spatio-temporal attention LSTM model for flood forecasting. In *Proceedings of International Conference on Internet of Things (iThings)* (pp. 458–465).
4. Liu, F., Xu, F., & Yang, S. (2017). A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with bp neural network. In *2017 IEEE third International conference on multimedia big data (BigMM)* (pp. 58–61). Ieee.
5. Bafitlhile, T. M., & Li, Z. (2019). Applicability of ε-support vector machine and artificial neural network for flood forecasting in humid, semi-humid and semi-arid basins in china. *Water, 11*, 85.
6. Biondi, D., & De Luca, D. L. (2012). A bayesian approach for real-time flood forecasting. *Physics and Chemistry of the Earth Parts A/B/C, 42–44*, 91–97.
7. Han, D., Chan, L., & Zhu, N. (2007). Flood forecasting using support vector machines. *Journal of Hydroinformatics, 9*, 267–276.

8. Le, X.-H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (lstm) neural network for flood forecasting. *Water, 11*, 1387.
9. Wu, Y., Cao, H., Yang, G., Lu, T., & Wan, S. (2022). Digital twin of intelligent small surface defect detection with cyber-manufacturing systems. *ACM Transactions on Internet Technology*, .
10. Gou, J., Sun, L., Yu, B., Wan, S., Ou, W., & Yi, Z. (2022). Multi-level attention-based sample correlations for knowledge distillation. *IEEE Transactions on Industrial Informatics*, .
11. Qiao, W., Tian, W., Tian, Y., Yang, Q., Wang, Y., & Zhang, J. (2019). The forecasting of pm2. 5 using a hybrid model based on wavelet transform and an improved deep learning algorithm. *IEEE Access, 7*, 142814–142825.
12. Wu, Y., Guo, H., Chakraborty, C., Khosravi, M., Berretti, S., & Wan, S. (2022). Edge computing driven low-light image dynamic enhancement for object detection. *IEEE Transactions on Network Science and Engineering*, .
13. Chen, C., Jiang, J., Zhou, Y., Lv, N., Liang, X., & Wan, S. (2022). An edge intelligence empowered flooding process prediction using internet of things in smart city. *Journal of Parallel and Distributed Computing, 165*, 66–78.
14. Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial marketing management, 90*, 493–507.
15. Zhang, Y., Zhang, F., Jin, Y., Cen, Y., Voronin, V., & Wan, S. (2022). Local correlation ensemble with gcn based on attention features for cross-domain person re-id. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*
16. Toth, E., Brath, A., & Montanari, A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of hydrology, 239*, 132–147.
17. Yu, P.-S., Chen, S.-T., & Chang, I.-F. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology, 328*, 704–716.
18. Biondi, D., & De Luca, D. (2013). Performance assessment of a bayesian forecasting system (bfs) for real-time flood forecasting. *Journal of hydrology, 479*, 51–63.
19. Ding, Y., Zhu, Y., Wu, Y., Jun, F., & Cheng, Z. (2019). Spatio-temporal attention lstm model for flood forecasting. In *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (pp. 458–465). IEEE.
20. Song, T., Ding, W., Wu, J., Liu, H., Zhou, H., & Chu, J. (2020). Flash flood forecasting based on long short-term memory networks. *Water, 12*, 109.
21. Liu, S., Chen, P., & Wozniak, M. (2022). Image enhancement-based detection with small infrared targets. *Remote. Sens., 14*, 3232.
22. Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv http://arxiv.org/abs/1803.01271arXiv:1803.01271
23. Chen, Y., Kang, Y., Chen, Y., & Wang, Z. (2020). Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing, 399*, 491–501.
24. Shen, Z., Zhang, Y., Lu, J., Xu, J., & Xiao, G. (2018). Seriesnet: a generative time series forecasting model. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE.
25. Huang, Q., & Hain, T. (2020). Improving audio anomalies recognition using temporal convolutional attention network. arXiv preprint arXiv:2010.11286, .
26. Liu, S., Wang, S., Liu, X., Lin, C., & Lv, Z. (2021). Fuzzy detection aided real-time and robust visual tracking under complex environments. *IEEE Trans. Fuzzy Syst., 29*, 90–102.
27. Liu, S., Liu, D., Muhammad, K., & Ding, W. (2021). Effective template update mechanism in visual tracking with background clutter. *Neurocomputing, 458*, 615–625.
28. Ding, Y., Zhu, Y., Feng, J., Zhang, P., & Cheng, Z. (2020). Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing, 403*, 348–359.
29. Li, R., Wu, Z., Jia, J., Bu, Y., Zhao, S., & Meng, H. (2019). Towards discriminative representation learning for speech emotion recognition. In *IJCAI* (pp. 5060–5066).
30. Yao, L., Torabi, A., Cho, K., Ballas, N., Pal, C., Larochelle, H., & Courville, A. (2015). Video description generation incorporating spatio-temporal features and a soft-attention mechanism. arXiv preprint arXiv:1502.08029, .
31. Serra, J., Suris, D., Miron, M., & Karatzoglou, A. (2018). Overcoming catastrophic forgetting with hard attention to the task. In *International Conference on Machine Learning* (pp. 4548–4557). PMLR.
32. Fan, C., Zhang, Y., Pan, Y., Li, X., Zhang, C., Yuan, R., Wu, D., Wang, W., Pei, J., & Huang, H. (2019). Multi-horizon time series forecasting with temporal attention learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2527–2535).
33. Zang, J., Wang, L., Liu, Z., Zhang, Q., Hua, G., & Zheng, N. (2018). Attention-based temporal weighted convolutional neural network for action recognition. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 97–108). Springer.
34. Shih, S.-Y., Sun, F.-K., & Lee, H.-Y. (2019). Temporal pattern attention for multivariate time series forecasting. *Machine Learning, 108*, 1421–1441.
35. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. arXiv preprint https://doi.org/10.48550/arXiv.1706.03762
36. Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., & Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. arXiv preprint https://doi.org/10.48550/arXiv.1907.00235
37. Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell., 39*, 640–651.
38. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*, 1735–1780.
39. Shen, Z., Zhang, Y., Lu, J., Xu, J., & Xiao, G. (2018). Seriesnet: A generative time series forecasting model. In *Proceedings of International Joint Conference on Neural Networks* (pp. 1–8).
40. Wu, Y., Liu, Z., Xu, W., Feng, J., Palaiahnakote, S., & Lu, T. (2018). Context-aware attention LSTM network for flood prediction. In *Proceedings of International Conference on Pattern Recognition* (pp. 1301–1306).
41. Feng, J., Wang, Z., Wu, Y., & Xi, Y. (2021). Spatial and temporal aware graph convolutional network for flood forecasting. In *Proceedings of International Joint Conference on Neural Networks* (pp. 1–8). IEEE.

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