IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 March 2024, accepted 21 April 2024, date of publication 8 May 2024, date of current version 15 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3398130

RESEARCH ARTICLE

Design of a Variable-Length Sequential Prediction Framework GTV-STP Based on Spatial and Temporal Water Quality Information of Taihu Lake

YUBO LEI, BO HU, HONGYUAN HUANG, AND YOUPING LIU
School of Civil Engineering and Architecture, Nanchang Institute of Technology, Nanchang 330000, China
Corresponding author: Youping Liu (csulyp@126.com)

This work was supported by the National Natural Science Foundation of China under Grant 51409140.

ABSTRACT Water quality predictions have a great importance in water resource managements and water pollution protections. Most the currently used water quality models can only predict time sequence of fixed length and ignore the real hydrological flow information. In this paper, we propose a new water quality prediction framework GTV-STP that adopts an embedding-Encoder-Decoder structure, in which the timestamps are introduced into the water quality information. The spatial embedding method is developed to introduce the spatial hydrological features into the water quality information, in order to achieve multisite parallel predictions. In addition, a variablelength decoder is put forward for the variable-length sequencial predictions. The practical water quality predictions of WT, pH, DO, COD_{Mn}, NH_3-N, P and N for Taihu Lake with GTV-STP framework are performed on two datasets of 6to12 in smaller-scale and 12to24 in larger-scale. Using the Huber Loss to balance the MAE and the RMSE, the val-Huber Loss of the GTV-STP framework is used to compare with that of the baseline models such as LSTM-CNN-ATT, LSTMNet and MLP in predicting accuracy. Results show that the GTV-STP framework has the highest accuracy with the val-Huber Loss of 0.079183 on 6to12 and 0.033561 on 12to24. It is shown that the GTV-STP framework has highly accurate not only for smaller-scale water quality predictions but also suitable for larger-scale predictions. In future, water quality predictions using other more precise frameworks containing spatial and time series information under deep learning will be one of the research directions.

INDEX TERMS GTV-STP, spatial and temporal water quality information, Taihu Lake, variable length sequential predictions, water quality predictions.

I. INTRODUCTION

In recent decades with explorations of prediction problems, the integrating considerations of time and spacial properties can give better results in solving such problems. Li et al. make a better prediction through adding the spatial attributes into the different combined data information of the port flow and time in a certain wide spatial spans [1].

Recently, in prediction of water quality indicators, the RNN model was used by Heesung et al. [2] and the NARX network model was used by Chang et al. [3]The NARX network model is also a subclass of RNN, in which the long-term relationships can be established between time information and water quality indicators and give better predictions. Shi et al. Used CNN+LSTM method to improve the prediction accuracy through spatial relation convolutions between the geographically adjacent regions [4].

Chen et al. delve deeply into the extensive application and significance of ANNs in the realm of water quality prediction [5]. They provide a comprehensive analysis of recent studies, detailing various types of ANN models and how they are utilized to tackle a myriad of issues within

The associate editor coordinating the review of this manuscript and approving it for publication was Liandong Zhu.

water quality prediction, including but not limited to the prediction of pollutant concentrations, assessment of water body nutritional status, and early warning of pollution events. The article also discusses the advantages of using ANN models in water quality prediction, the challenges faced, and the direction of future developments. It emphasizes the potential of deep learning technologies to enhance the accuracy and efficiency of water quality predictions while pointing out limitations related to data availability, model generalization capabilities, and interpretability. Through this review, readers can gain a thorough understanding of the current application status and future prospects of ANNs in water quality monitoring and management.

In this paper, a new water quality prediction framework of GTV-STP (Geography-graph and Time-graph were used by Variable-length Spatio-Temporal sequence Predict) is proposed, in which we use the structure of Embedding-Encoder-Decoder to encode the spatio-temporal water quality information. It will be used to predict the water quality index. The main contributions of this paper are as follows:

1) Predicting the overall water quality of a watershed based on the modeling of water quality data from a single site is challenging. To overcome this limitation, we employed hydrologic flow diagrams to model water quality data from multiple sites, aiming to provide a more accurate prediction of the overall watershed water quality. It is difficult to predict overall watershed water quality by modeling water quality data from a single site, and we modeled water quality data from multiple sites by introducing hydrological flow for the purpose of predicting overall watershed water quality.
2) Due to the sequential models having the information degradation phenomena with the increase of sequential length when we captured the long-term dependent water quality information. To avoid this phenomenon, we designed a method to introduce the timestamps into the water quality information.
3) A variable-length decoder is designed to adapt to the variable-length sequential prediction. For example, we can use the input sequential length as T_1 and the output sequential length as T_2.
4) The data sets of actual water quality information are collected and will be used to the verify the proposed model in this paper. Merits of this model are having better prediction results than the other baseline models.

II. RELATED WORK

In recent decades, researchers have conducted extensive research works on the prediction of water quality indicators. Antar et al. proposed a rainfall-runoff model based on the artificial neural network (ANN) for the Blue Nile Basin [6].Their research work was conducted in seven subareas of the Nile Basin. The main inputs for the ANN model was the average amount of precipitations within a certain time. Although the work have showed enormous potential on the predictions with ANN technique, evident bias will also exist in the predictions of mass data. The reason is that the data processing capability of the ANN model is not strong for available records.

Khani and Rajaee has proposed two hybrid models: the regression model based on the wavelet (WR), and the artificial neural network model based on wavelet (WANN) which is based on the RNN technique [7].The research works show that the results from the WR model are superior to the other models for short interval modeling. But for a long interval modeling in spring and summer seasons, the WANN model can give a better prediction on DO than the WR model.

Antanasijevic et al. has built an ANN model using non-specific water quality parameters, and compared the accuracy among the following three kinds of ANN architectures to predict the dissolved oxygen (DO) concentrations in the Danube River: the general regression neural networks (GRNN), the back propagation neural networks (BPNN) and the recursive neural networks (RNN) [8]. Using the data sets of water flow, temperature, pH and conductivity as input variables from the year of 2004 to 2008, result comparisons show that the RNN model has much better performance than the MLR model.

In recent years, the hybrid artificial neural network models have been used more and more frequently and play a significant roles in predictions. Tian et al. optimized the FNN, RNN and LSTM models with different structures using the transfer learning (TL) [9], and use them to predict and analyze the chlorophyll-related data sets at certain time intervals on Hekou reservoir in eastern China. The results show that the hybrid artificial neural network model can maintain a higher prediction accuracy.

Up to now, most methods of water quality prediction often use the individual data or average values from the specific hydrometric stations ignoring the hydrological flow relationships among the different stations. Different from methods mentioned above, we propose a new model GTV-STP, in which the Embedding- EncoderDecoder architecture is used and the spatial hydrological characteristics and time stamp are integrated into the water quality informations, to predict the water quality indicators. This method can realize the multistation parallel predictions through the hydrological flow directions and the spatial information modeling. The addition of timestamp enables the model to make differentiated predictions of water quality indicators. Thereby, we can also enrich and improve the basic data of model training to obtain more accurate predictions.

III. DATASETS

We chose a public dataset of the Taihu Lake basin to evaluate the performance of the GTV-STP framework. This dataset was converted to a 3D tensor $N \times T \times F$ substituting into the GTV-STP model, where N denotes the number of water quality monitoring stations, and T denotes the total number of samples observed at the water quality station

TABLE 1. Basic information of water quality variables.

Water Quality Variables	Full Name	Categories	Unit
WT	water temperature	physical	°C
pH	pondus Hydrogenii	physical	none
DO	dissolved oxygen	chemical	mg/L
COD_{MN}	Permanganate index	chemical	mg/L
NH_3-N	Ammonia nitrogen content index	chemical	mg/L
P	Phosphate relative	physical	mg/L

during the specified time period. The Taihu Lake dataset was then subjected to the data preprocessing in the following Section VI. Three following parts of the data were obtained after processing: the training part (80%), the validation part (20%).

The Taihu Lake dataset was collected from National Real-Time Data Dissemination System for Automatic Surface Water Quality Monitoring of China, with environmental data processing support provided by Qingyue Data (data.epmap.org). The dataset was divided into following two parts: water quality indicator data and monitoring station location data. Among them, the frequency of water quality indicator data from each water quality monitoring station is 4h, and 16 kinds of indicator data including dissolved oxygen are measured. Considering most of the water quality assessment scenarios, we selected only 6 indicators as follow: WT,pH,DO,COD_{MN},NH_3-N,P.(As shown in Table 1)

The monitoring station location data include latitude and longitude information of 73 monitoring stations. Comprehensively considering the actual situation of the Taihu Lake basin, such as the sudden changes of water quality indicators in estuaries and major rivers affecting the Taihu Lake basin, we selected latitude and longitude data of 14 stations and their water quality indicator data. Table 1 and Table 2 show the information about the selected water quality indicators and monitoring station locations, respectively.

Taihu Lake locate at between latitudes 30° 55' 40'' and 31° 32' 58'' north and longitudes 119° 52' 32'' and 120° 36' 10'' east. There are many estuaries with more than 50 major inlet and outlet rivers, which have great significance for the predictions of water quality indicators. We used 14 station locations, which are located in different inlets and protected areas of Taihu Lake, and the geographic information of the 14 stations is shown in Table 2. Additionally, the number of missing data entries corresponding to each water quality variables is also displayed in Table 2. In the training part(80%, Data entries = 27866), the maximum number of missing entries for a single indicator is 1904. In the validation part (20%, Data entries = 6966), the maximum number of missing entries for a single indicator is 476.The maximum number of missing entries for a single indicator is less than 20% of the total number of data, so the KNN(K-Nearest Neighbors) method can be used.

The locations of these sites and their actual flows of water between them are shown in Fig. 1.

IV. OVERVIEW

Fig. 2 shows the framework of the workflow and data types of the GTV-STP which consists of the following three parts: the spatiotemporal diagram embedded, the dual-channel spatiotemporal encoder and the variable spatiotemporal decoder. The following 4 types of data related to water quality sequences are used as inputs for the spatiotemporal diagram embedding module:

1) Spatial graph G_p, in which the three dictionary types of data that represents the nodes, the edges and the adjacencies between nodes respectively.
2) Time period graph G_T, consistent in data structures with that of the spatial graph, but different from the spatial graph in time graph with a cyclic graph structure.
3) Time period stamp label $L_T \in \mathbb{R}^{T\times 1}$, a one-dimensional vector where the list element represents timestamp classifications under specified periods. For example, when the time period is weekly, the time label corresponds the value of the timestamp taken from the list $[0, 1, \ldots, i]]$, where i is the week i.
4) Water quality time-sequence X, a three-dimensional tensor $\mathbb{R}^{N\times T\times F}$ where N, T and F are the water quality monitoring sites, the time and the water quality characteristic dimensions respectively.

These different types of data will input into the Spatio-Temporal Graph Embedding module, which mainly performs the following two tasks:

1) The timestamp labels will be embedded into the time period graph, and the embedded results are used to replace the time stamp of water quality time sequences X. So that, a three-dimensional tensor $X^C \in \mathbb{R}^{N\times T\times(F+D)}$is obtained to simulate the period correlation of the historical data.
2) The spatial graph will be embedded, and the spatial features are introduced into the water quality time sequence X using the embedding results. Thus, a three-dimensional tensor $X^P \in \mathbb{R}^{N\times T\times(F+D)}$ is obtained to simulate the spatial correlation of the historical data.

The input the output results of X^C and X^P into the Dual Channel Spatio-Temporal Encoder module which has two Temporal Attention components with the same structures, and capture the complex time dependent relations between X^C and X^P respectively. After that, the integration of X^P into X^C with the Cross Attention component is used to incorporate the spatial coding based on the captured temporal dependencies. We will input the integration result into the Spatio Attention to capture the complicated spatial dependent relations based on the time dependent relationship, So that, we can obtain the fusion encoding result $H \in \mathbb{R}^{N\times T\times F}$. We will use the Variable Spatio-Temporal Decoder module to decode the H, and obtain the 3 dimensional tensor $\hat{Y}$ to predict the water quality time sequences for different sites. The Variable Spatio-Temporal Decoder module has three components of FC, TCN and Spatio Attention. Among them,

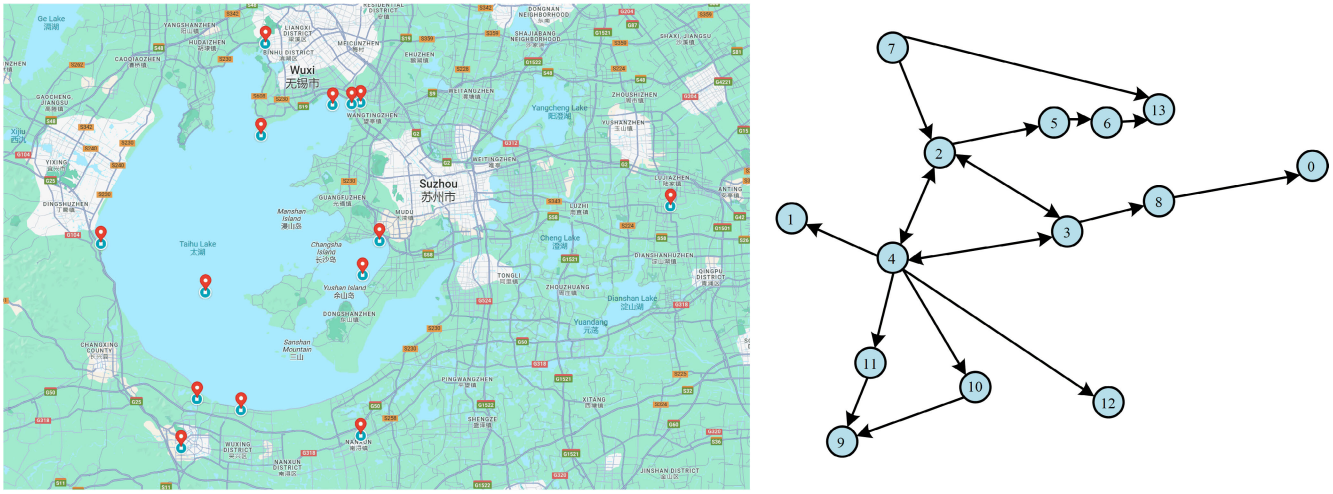

FIGURE 1. Spatial map G_P, modelled from hydrological flow and spatial information.

TABLE 2. Taihu lake, china.

No	The Name Of Water Quality Monitoring Site	Latitude	Longitude	Data entries		Missing data entries of each water quality variables					
						WT	pH	DO	COD_{MN}	NH_3-N	P
0	Wu-Song River	31.2748	121.0440	2488		113	113	113	246	125	202
1	Lanshanzui	31.2118	119.9118	2488		97	111	97	155	110	117
2	Shazhu	31.3962	120.2296	2488		132	132	132	179	135	137
3	Xuhuxin	31.1593	120.4318	2488		218	218	218	235	229	222
4	Western Hills	31.1300	120.1200	2488		101	107	101	100	100	100
5	Xidong waterworks	31.4483	120.3722	2488		173	173	184	183	181	182
6	Wangyu River	31.4494	120.4101	2488		139	141	210	157	164	173
7	Liangxi River	31.5525	120.2380	2488		130	131	134	141	146	159
8	Xu River	31.2173	120.4655	2488		87	87	88	125	115	121
9	West of the City Bridge	30.8654	120.0726	2488		218	218	218	238	227	245
10	Daqian	30.9300	120.1909	2488		136	136	138	184	138	146
11	Xiaomeikou	30.9485	120.1036	2488		33	33	33	49	35	46
12	Ditang	30.8870	120.4282	2488		229	229	229	244	235	211
13	the Grand Canal	31.4520	120.4276	2488		110	110	109	144	136	117
total				34832	Max	229	229	229	246	235	245

the FC is the fully connected layer. The function of FC layer is to learn to map complex patterns in the input data to the output space, thereby extracting and representing intricate features. Such feature representations aid in simplifying the decoding or encoding tasks of subsequent modules, making the training and inference processes of the entire deep learning model more efficient and effective. It maps the time length T of H to the time length $T'$to achieve the purpose of changing the time length prediction.TCN is a time convolutional network used to decode the 3 dimensional tensor T' with time length in time dimension. The Spatio Attention has the same structures as the Spatio Attention of the Dual Channel Spatio-temporal Encoder, and it is used to spatially decode a three-dimensional tensor containing temporal decoding information on the spatial dimension N.

V. MODEL

In this section, we will describe the detailed processing flow of every component in each module containing in the GTV-STP framework, as shown in Fig. 2. For ease of description, here we will introduce three formal definitions:

1) Time periodogram: directed cycle graph used to represent the time flow in time period and under time gap constraints. In recent years, such graphs have been used in traffic flow predictions and obtained excellent performance [10], as shown in Fig. 3, in which G_T represents the time periodogram with size $len(sc) * len(ts)$,where sc is a periodic sequence $sc = [sc_0, sc_1, \ldots, sc_j, \ldots, sc_m]$ (sc_j is cycle number, for example, if period is c, then $sc = [0, 1, \ldots, j, \ldots, c-1]$); $ts = [ts_0, ts_1, \ldots, ts_i, \ldots, ts_n]$ is an interval sequence of a fixed time length (ts_i is the time interval between the moments of $i-1$ and $i-1+f$, where f is the measured frequency). If the measured duration is Δt, then $n = \lfloor \Delta t/f \rfloor$. The black directed lines represent the time flow over a periodic sequences, and the red directed lines represent the time flow over

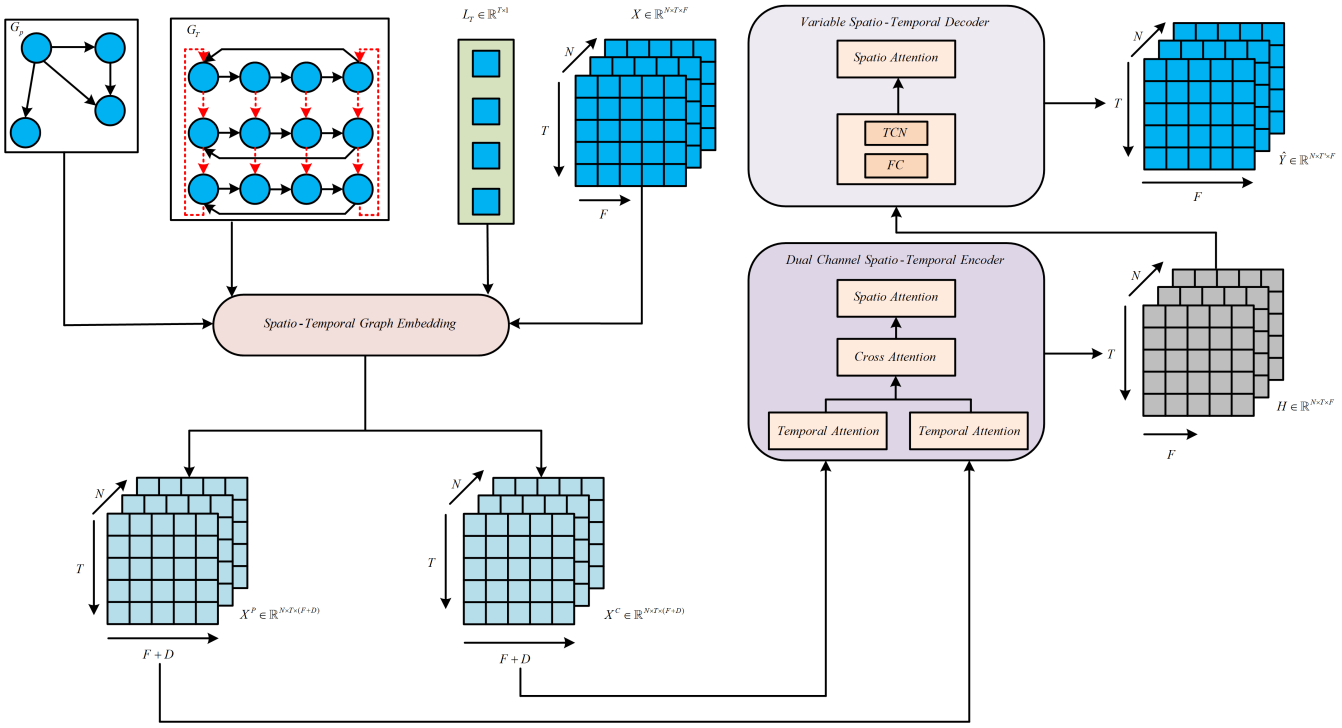

FIGURE 2. GTV-STP framework.

a interval sequences, and blue circles represent the time nodes. The node numbers tn_{ij} are calculated by $t_{ij} = i * (m + 1) + j$.

2) The periodic timestamp label: A periodic timestamp label is a classification label that uses the periodic sequential definition 1) to classify the timestamp of a time sequence, and replaces the timestamp with the corresponding classification label, in order to express the similarity of the time series in different period segments trend. The period timestamp label is the category of the timestamp obtained from the *sc*, which makes the data with a period interval have the same timestamp label.
3) The basin water quality prediction task: The basin water quality prediction task is to: (I) design an algorithm F whose input is historical water quality time sequence data $X = \{X_0, X_2, \ldots, X_{t-1}\}$, where $X_i = \{x_1, x_2, \ldots, x_N\}$ represents the data from N water quality monitoring stations at time i. Three additional information G_p, G_T and L_T are respectively the directed weighted maps representing the spatial location of water quality stations calculated according to latitude and longitude, the time period map established according to water quality sequence, and the cycle time stamp label.(II) Find the optimal learn parametersΘ, such that F at this Θ make the output of the calculated $\hat{Y}$ close to $Y = \{Y_t, Y_{t+1}, \ldots, Y_{k-1}\}$, where the time length of the predicted target Y is $k - t$ and the time length of the historical series X is t.

A. SPATIO-TEMPORAL GRAPH EMBEDDING (FEATURE EXTRACTION)

The water quality time sequences usually have periodic characteristics. Introducing such information into the water quality prediction can guide the model to learn the periodic trends and correct the prediction results. Moreover, basin water quality data usually contain water quality time sequences of different spatial locations which exist co-relationships of upstream and downstream among them. Effects the change of water quality indicators and have influences on the water quality indicators. The introduction of co-relationships between upstream and downstream can enable the model to learn the influences on the water quality time sequence. Therefore, we can get differential predictions for different locations.

In order to introduce the above two features, the timestamp in water quality time sequence data X and the spatial graph G_p composed of every water quality station can be encoded into a fixed-length vector. The vector then will be integrated into the water quality time sequential data X to form new data X^P and X^C, shown in Fig. 4.

We create a time graph G_T and a periodic timestamp label L_T for X based on definitions 1) and 2), to form a time feature $X^{(T)} \in \mathbb{R}^{T \times 1}$. The spatial graph G_p and the time graph G_T are wanderingly embedded with node2vec respectively to obtain

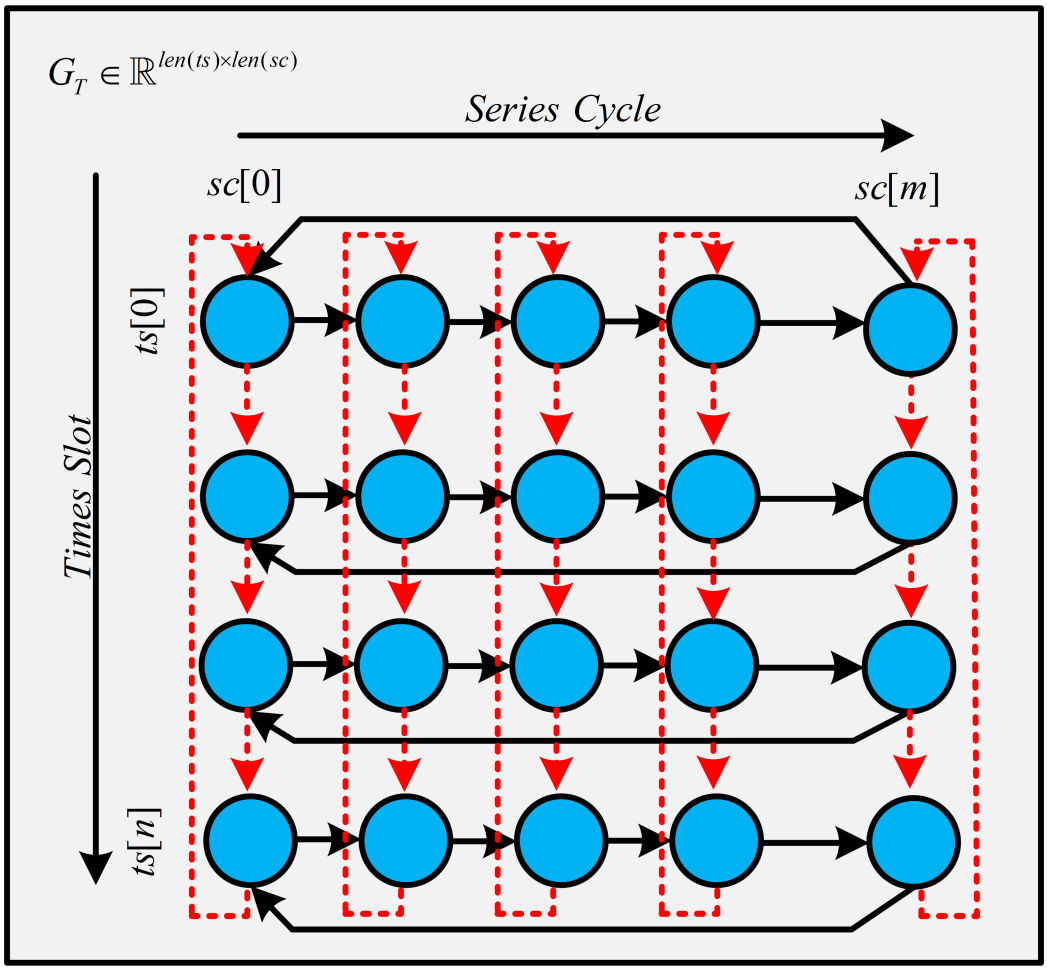

FIGURE 3. Time graph.

the positional embedding matrix $P \in \mathbb{R}^{N\times D}$ and the periodic embedding matrix $W_T \in \mathbb{R}^{(len(ts)*len(sc))\times D}$. W_T is regarded as the initialization parameter matrix $X^{(T)}$ to be Embedded as a label, in order to acquire the timestamp code $tcode \in \mathbb{R}^{T\times D}$. The P and $tcode$ will be made N copies and spliced them with X in F dimension, to Obtain the 3D tensors of X^P and X^C. Details of the node2vec and Word Embedding are as follows:

1) NODE2VEC

The node2vec is a random walk graph embedding method, which generates the node sequences according to the walk probability of (1), where $\pi_{tn_i\to tn_j}$ is the transition probability from node tn_i to tn_j; $\alpha_{pq}(tn_k, tn_i)$ is a tendency control parameter to guide the next migration tendency by calculating the migration tendency of the previous step; tn_k is the time node just visited in the previous step; tn_i is the time node currently visited; tn_j is the time node that may be visited in the next step; d is the shortest path from time node tn_k to tn_i; p and q are visit weights; $w_{tn_itn_j}$ is the weight of the edge $< tn_i, tn_j >$ and maps the generated sequence into vectors based on the Word2Vec in order to appropriately represent the relationships and features among nodes [11].

$$\pi_{tn_i\to tn_j} = \alpha_{pq}(tn_k, tn_i)\cdot w_{tn_itn_j}$$
$$\alpha_{pq} = \begin{cases} \frac{1}{p} & d_{tn_ktn_i} = 0 \\ 1 & d_{tn_ktn_i} = 1 \\ \frac{1}{q} & d_{tn_ktn_i} = 2 \end{cases} \quad (1)$$

In our task, the characteristics of water quality index is subject to great environmental changes, the cycle is determined difficultly and selected randomly. From tn_{k-m-1}, you can choose randomly to walk along the red arrows or the black arrows with, and we use definition 1) to build the time cycle graph as shown in Fig. 5. The algorithm of node2vec to the graph, where $len(sc)$ and $len(ts)$ are set to adjustable parameters, so that the model can be fine-tuned in different basins. This figure shows the time period diagram when $len(sc) = 7$ and $len(ts) = 6$.

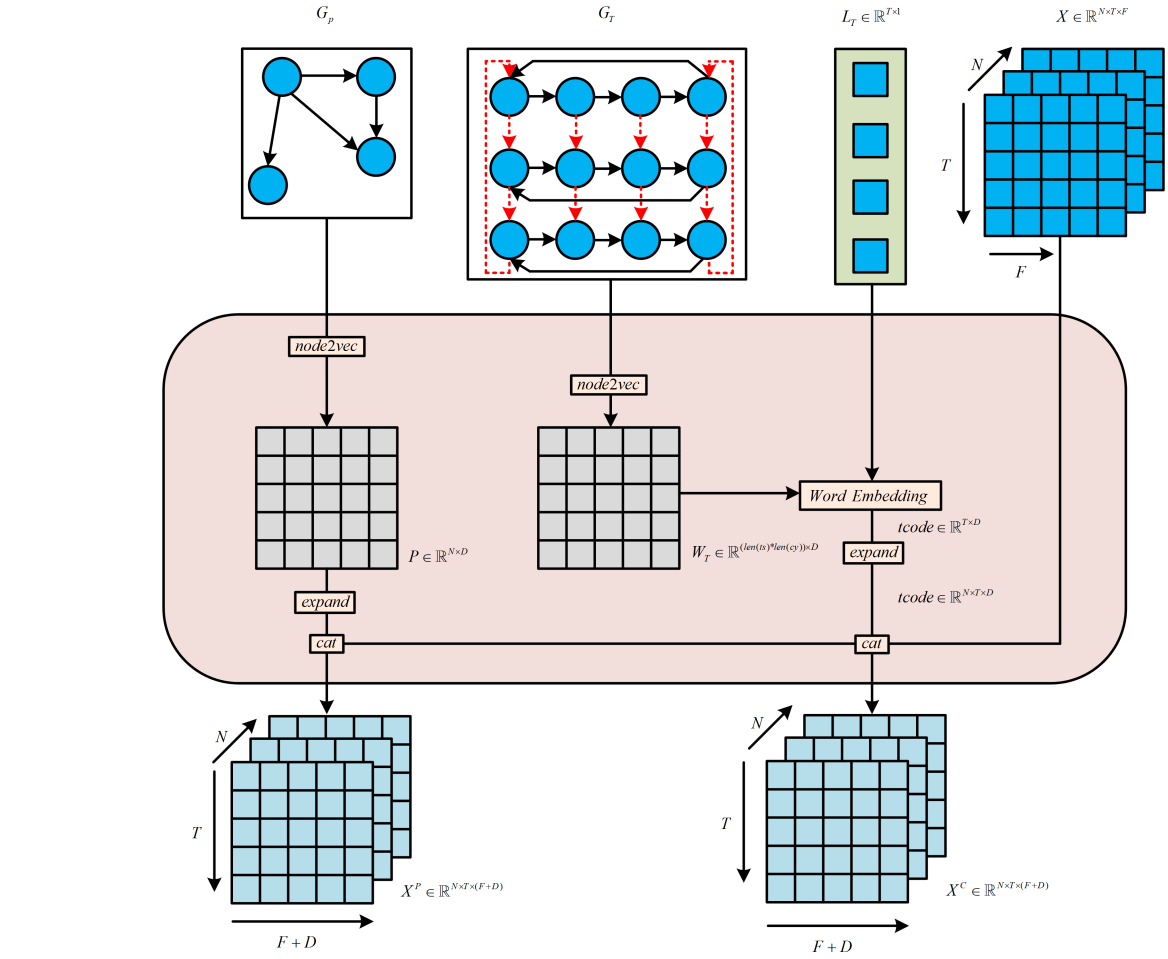

FIGURE 4. Time-Spatio graph embedding layer.

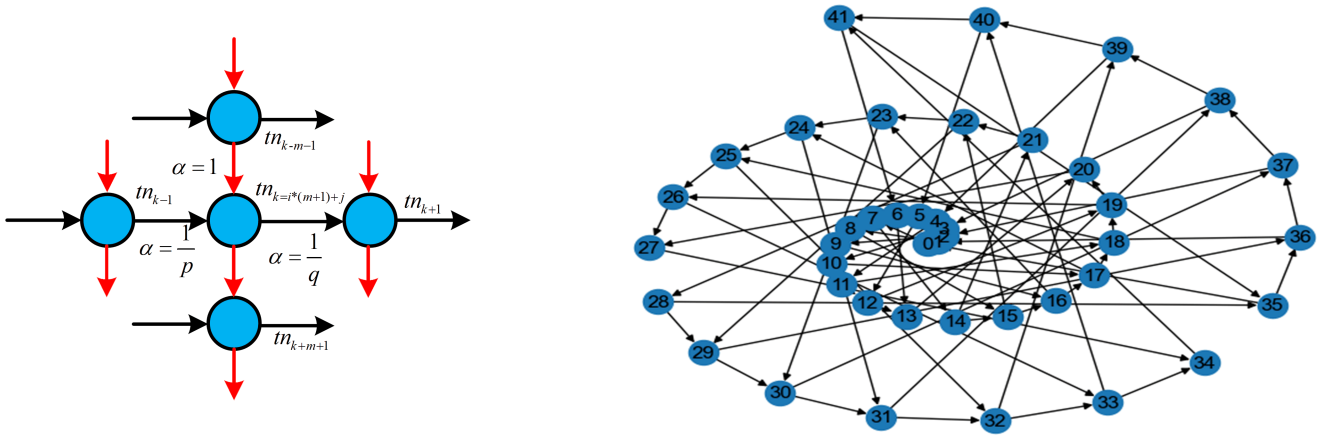

FIGURE 5. Node2vec.

2) WORD EMBEDDING

In order to embed a time label as a vector, we refer to the practice of embedding a one-hot vector as a dense vector [10]. Unlike previous work, our embedding method takes into account the problem of cycle similarity. In our task, if the period is t, the water quality data X_t will show the same trend as X_{2t}. Therefore, the timestamps of X_t and X_{2t} should be embedded as similar vectors. As shown in Fig. 6,

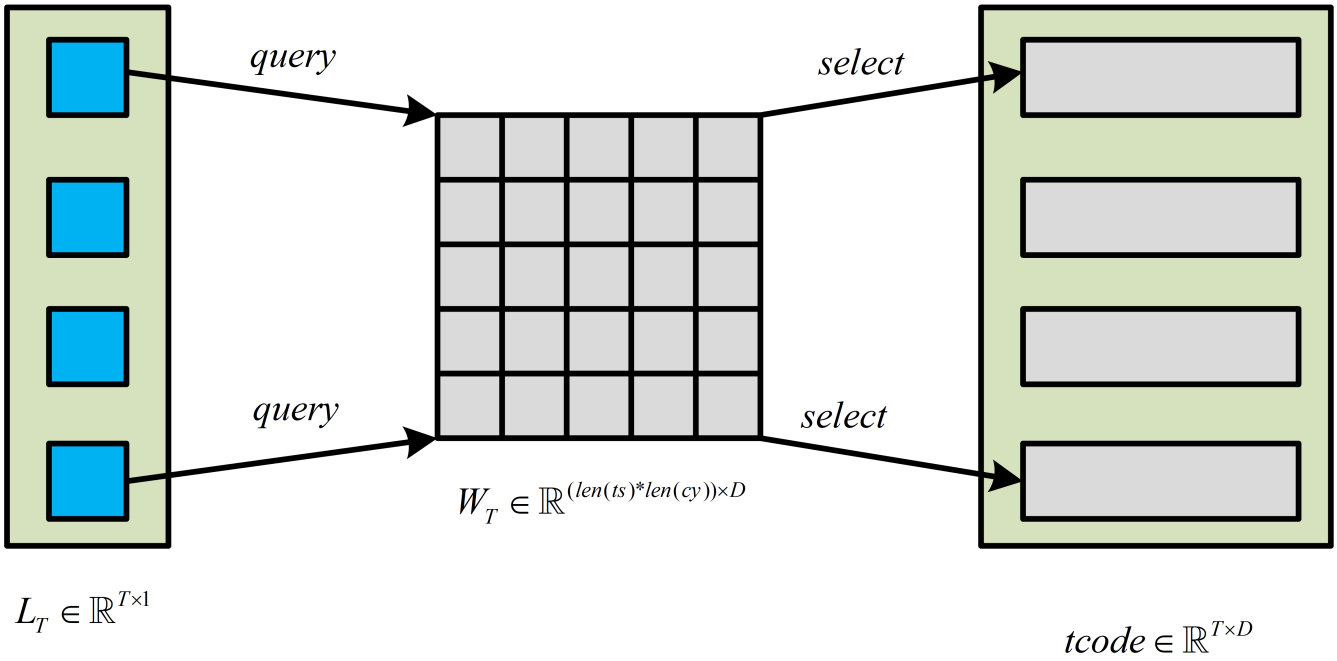

FIGURE 6. Word embedding.

We use the method of using indexes to find the result of node2vec embedding. The specific operation are as follows: the time feature tag $X_t^{(T)}$ at time t is used to take the row vector corresponding to the label value in W_T, and all the obtained vectors are splicing into the time feature encoding $tcode \in \mathbb{R}^{T\times D}$ in dimension 0.

B. DUAL CHANNEL SPATIO-TEMPORAL ENCODER

According to the Spatio Temporal Graph Embedding, we obtain two types of modal data.$X^P \in \mathbb{R}^{N\times T\times(F+D)}$ with position encoding and $X^C \in \mathbb{R}^{N\times T\times(F+D)}$ with period encoding. To integrate the long-term spatio-temporal dependence of these two modal data to capture water quality indicators, we designed a Dual Channel Spatio-temporal Encoder (Cross+Temporal+Spatio Attention), as shown in Fig. 7. Each of these three types of Attention introduces the attention mechanism of Formula 2 into their respective tasks. Among them, Temporal Attention and Spatio Attention can filter out non-critical information through time and space dependence and adapt to different time and spatial scales, which have been proved to significantly improve model performance in CV and sequence modeling tasks [12], [13]. Cross Attention can compress data in the fusion process and use the attention mechanism to dynamically calculate the impact factors between different modal data to achieve cross-data source interaction. This structure has been proved to improve the accuracy of tasks such as multi-modal matching [14]. In our encoder, Temporal Attention has Mask and Residual Block, using Mask can force the model to follow causality, for example, make the model pay attention to the data between t moments, but not the data after t moments. The use of residuals can effectively prevent the model from forgetting the historical data before the t moment due to the long time. In the coding process, the 3 dimensional tensors X^P and X^C capture the hidden representation of the three dimensional tensors $H^C \in \mathbb{R}^{N\times T\times D}$ and $H^P \in \mathbb{R}^{N\times T\times D}$ through the residual of the Mask Temporal Attention of the same structure respectively. Through Cross Attention, the processing of X^C is the input of Q in (2), X^P is the input of K and V in (2), and the fusion encoded 3 dimensional tensor $H^{CP} \in \mathbb{R}^{N\times T\times D}$ in time dimension is obtained. The spatial dependence is captured through SpatioAttention, and the fusion-encoded 3 dimensional tensor $H \in \mathbb{R}^{N\times T\times F}$ is obtained.

$$\begin{cases} Q = FC(X) \\ K = FC(X) \\ V = FC(X) \end{cases}$$
$$Att = softmax(\frac{QK^T}{\sqrt{d_k}})$$
$$H = FC(att \cdot V), \tag{2}$$

where FC is the fully connected layer and d_k is the attention factor

C. VARIABLE SPATIO-TEMPORAL DECODER

As shown in Fig. 8, we use a combination of a temporal convolutional network (TCN) and a full connection layer (FC) to achieve variable long time decoding, and use the Spatio Attention structure as same as that of the Encoder to achieve spatial decoding, resulting in a final prediction

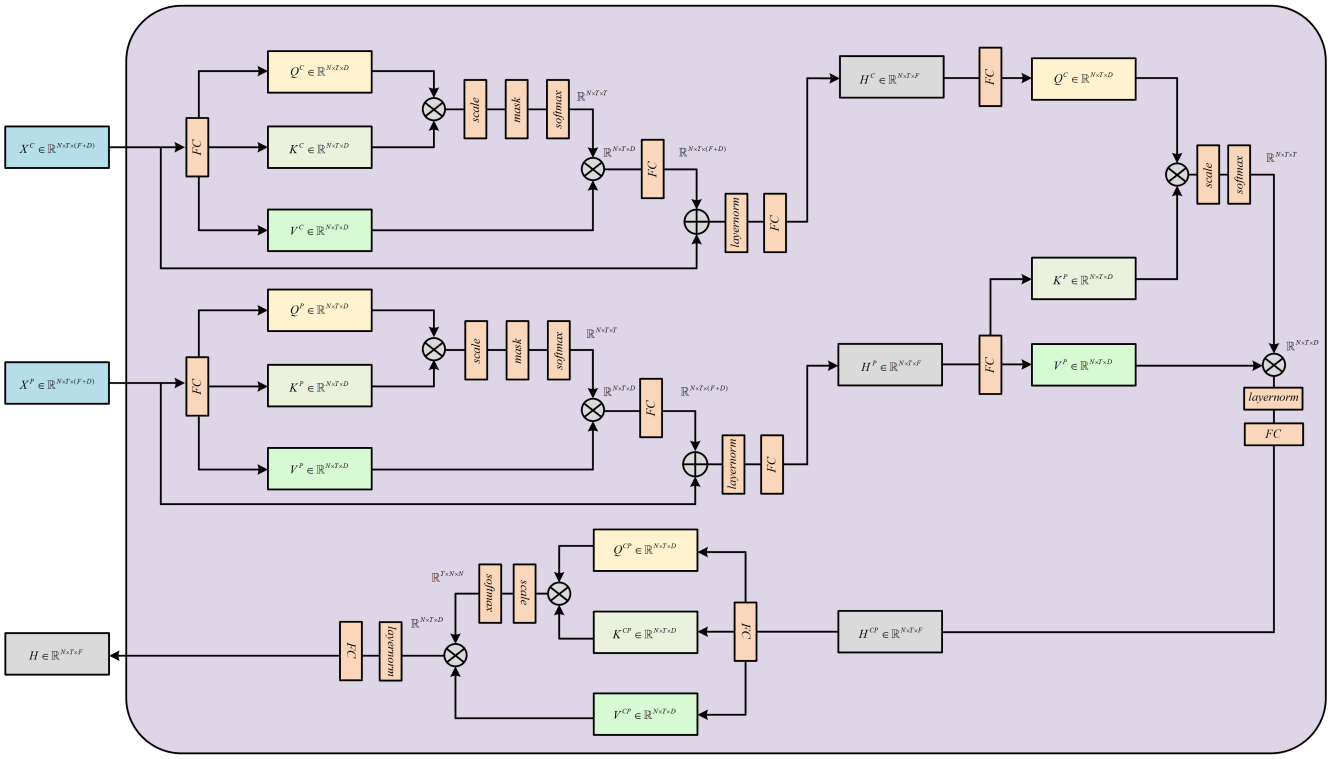

FIGURE 7. Dual channel spatio-temporal encoder.

target of the $\hat{Y} \in \mathbb{R}^{N\times T\times F}$. TCN is a network structure that introduces the ideas of CNN into the time sequence modeling [15]. It changes the size of its receptive field for time sequence by inflating convolution. For example, the receptive field can be controlled by increasing the inflating factor or stacking convolution layers. Secondly, TCN adopts residual blocks to prevent information loss as network depth increases, which has been verified to achieve ideal results on multiple general tasks [16]. In our prediction task, due to the use of the fully connected layer, the three-dimensional tensor $H \in \mathbb{R}^{N\times T\times F}$ is mapped to $H \in \mathbb{R}^{N\times T'\times F}$, in the time dimension. Some information may be lost in the process, but the inflation mechanism of TCN allows to focus on valuable information through the inflation factor and accept data of any length by sliding one-dimensional convolution. Therefore, here we adopt TCN as the time decoding component in the decoder.

$$H^l(s) = \sum_{i=0}^{k-1} f(i)X_{s-d\cdot i}$$
$$O(H^{l-1}, H^l) = \sigma(H^{l-1} + Conv1d(H^l)). \tag{3}$$

Equation (3) demonstrates the calculation method of layer l in TCN. The input time sequence s is convolved at the i th time step with a convolution kernel of size k. The convolution interval is d and the $s - d \cdot i$ ensures the effects of the causal convolution. 1 dimensional convolution is used to obtain the input H^l of layer l and taking the H^l transform and adding the residuals to the output H^{l-1} of the layer $l - 1$.

VI. DATA PREPROCESSING

In this section, we processed the water quality indicator data and water quality monitoring station location data separately. Considering the noise and error in the original water quality indicator data, we performed data normalisation. In order to adapt to variable length prediction, we performed sliding window segmentation on the raw water quality indicator data. In order to adapt to the input structure of the GTV-STP model, we constructed the temporal and spatial maps as well as the timestamp data mentioned in the model section as additional data.

A. DATA NORMALISATION

Considering that the dataset is small-scale and only a small percentage of data is missing, we used the KNN method to fill in this missing data. Additionally, in order to deal with the significant data differences among different indicators, the Z-Score normalization operations were applied to each water quality indicator separately [17], as shown in (4).

$$x' = \frac{x - \mu}{\delta}, \tag{4}$$

where μ is the mean value and δ is the standard deviation of each indicator.

B. SLIDING WINDOW SEGMENTATION

The sliding window aims to perform sliding intercepts of data X along T, intercepting data of window M size

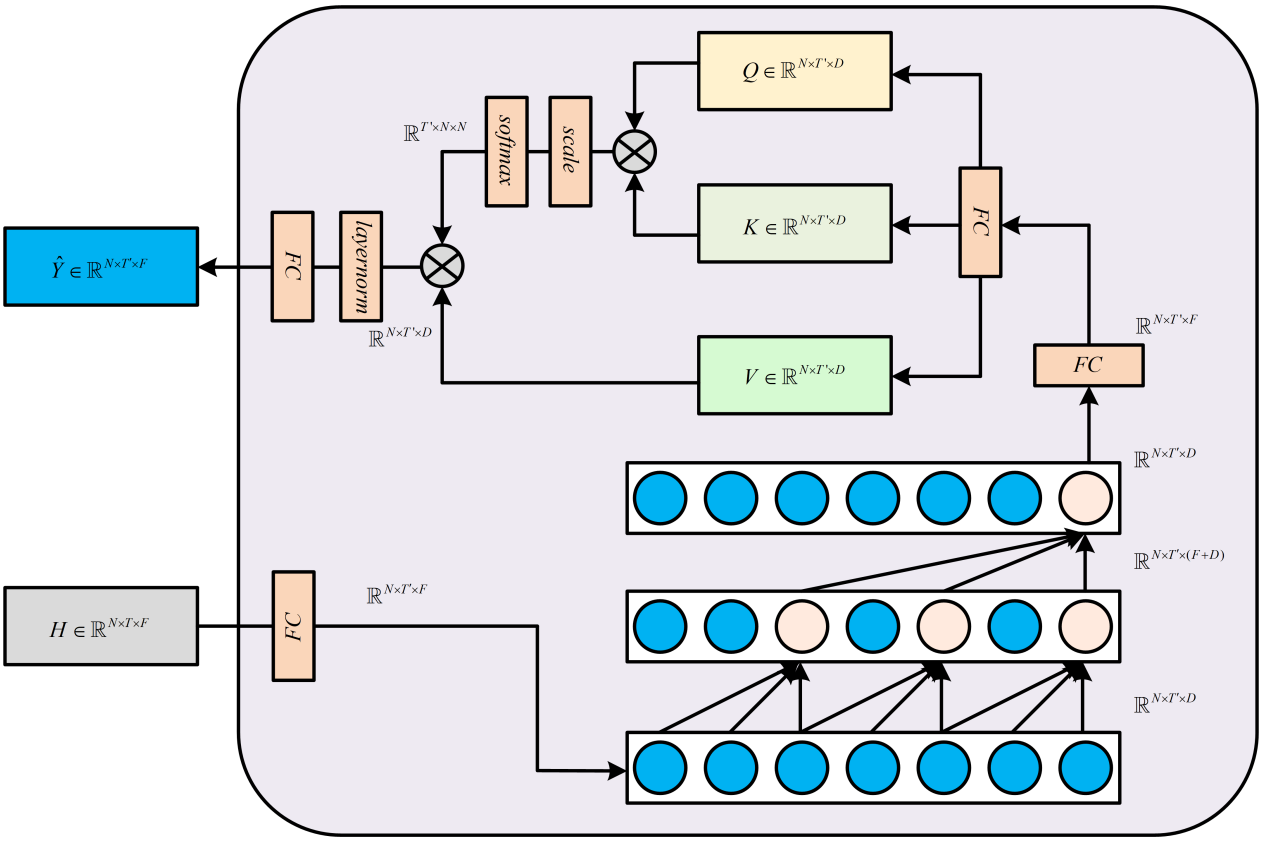

FIGURE 8. Variable spatio-temporal decoder.

each time. Because the model belongs to a variable-length prediction one, choosing the appropriate input and output lengths is important for model performances. Using too short input lengths will make it difficult to learn the feature representations of different water quality indicators, whereas using too long input lengths will impose a heavy computational burden as well as prediction delays. In order to obtain better prediction accuracy with lower computational costs and prediction delays, we use an overlapping sliding window segmentation method to intercept different window sizes for the model inputs and outputs, and construct two datasets under two variable-length datasets which consists of dataset 6to12 (6 sequence lengths predicting 12 sequence lengths) and dataset 12to24 (12 sequence lengths predicting 24 sequence lengths). Fig. 9 shows an example of sliding window segmentation, where the window $M = \{M^1 + M^2\}$ containing the input window $M^1 = [X_i : X_{i+p-1}]$ with size p and the output window $M^2 = [X_{i+p} : X_{q+i+p-1}]$ with size $p + q$. During the sliding process shown in Fig. 9, windows of M_1 and M_2 are partially overlapped, which means that the window M_2 contains parts of the history information in the window M_1.

C. EXTRA DATA CONSTRUCTION

We constructed a time graph as well as timestamps based on definition 1) in section V, as well as the water quality indicator data using week as a period and 6 weeks as a time slot. So the time graph has 42 nodes and timestamp categories are 0to6. A directed bandwidth-weighted spatial graph was constructed based on the relevant terrain characteristics of the Taihu Lake basin as well as monitoring station location data.

VII. EXPERIMENTAL SETUP

A. EXPERIMENTAL SETUP

We used the NetWorkX as the complex network construction tool [18], and used the time processing module Timestamp in the Pandas to construct the additional data in the data preprocessing. We adjusted the parameters using the Optuna package. The Pytorch deep learning framework and the two preprocessed variable-length datasets were utilized to train the GTV-STP model. The two datasets were trained by means of the Adam optimizer and the loss function of Huber Loss.

B. ADJUST THE PARAMETERS

Due to Optuna's employment of sampling and pruning algorithms to optimize hyperparameters, it is possible to prematurely terminate some sampling points with inferior intermediate results, thereby accelerating the search process. We adjusted the parameters using the Optuna package, where we set the number of experimental runs (train) to 100 (attempting to adjust parameters 100 times). The outlier

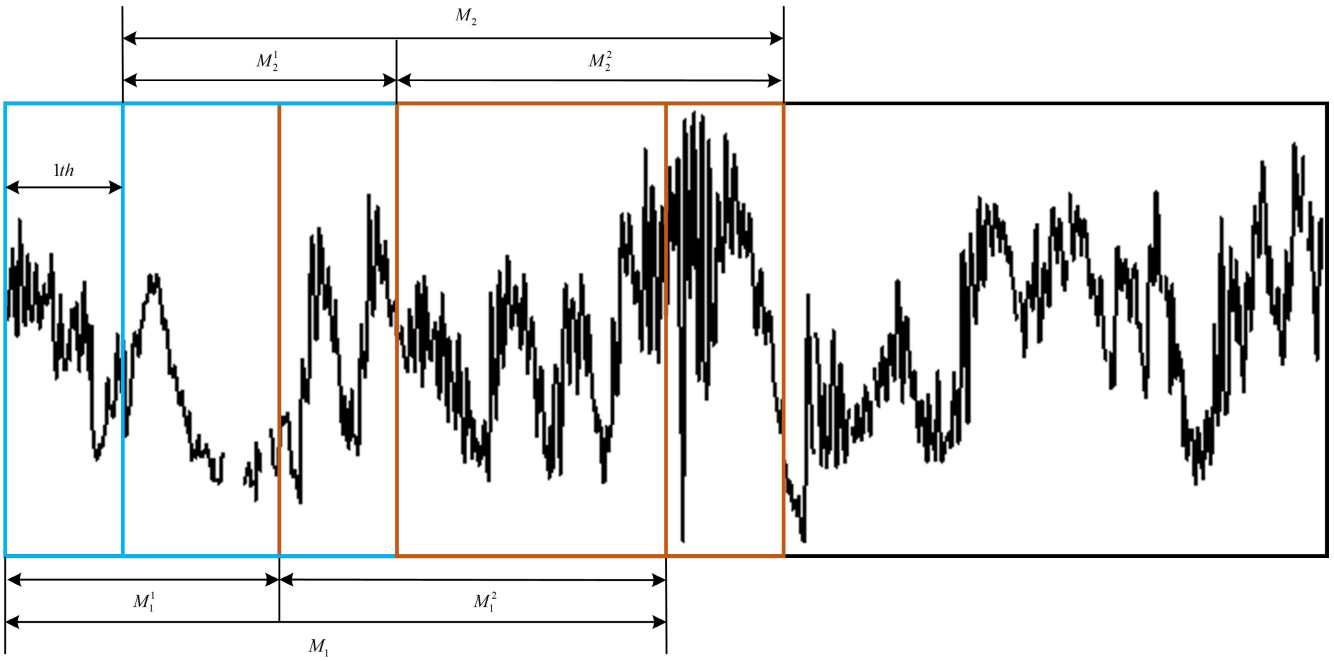

FIGURE 9. Sliding window segmentation.

sensitivity of the Huber Loss was set to 1.0, the epoch parameter was established at 500, and the number of samples per batch along with the Adam optimizer's learning rate were determined experimentally. The input data were randomized to prevent overfitting. We preserved the optimal parameters identified through this process.

VIII. RESULTS AND ANALYSIS

Results presented in this section have a range of fluctuations within ± 0.1. It is normal due to the influences of research platforms and other factors.

According to the GTV-STP framework and the two datasets of 6to12 and 12to24 used for training with epoch 500 and train 100, the related parameters obtained from experiments are shown in Table 3.

In order to evaluate the performances of the GTV-STP, its two components of train and val were evaluated using MAE, RMSE and Huber Loss. While Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are common metrics for evaluating model prediction performance, they are not sufficient in all contexts, leading to the introduction of Huber Loss as an alternative assessment tool. MAE, being a linear score, can underestimate the impact of outliers as it treats all errors equally, lacking sensitivity to the distribution of errors. On the other hand, RMSE, due to its quadratic nature, heavily penalizes larger errors, which can lead to an overestimation of the impact of outliers on model performance. Huber Loss, however, bridges the gap between MAE and RMSE by combining the best of both metrics. It behaves like MAE for small errors and like RMSE for larger errors, introducing a threshold (delta) to switch between the two behaviors. This makes Huber Loss more robust to outliers than RMSE while being less sensitive to errors than MAE for large deviations.

TABLE 3. Model parameters and training parameters.

training parameters(6to12)	interpretations
hdim =16	The embedding dimension for node2vec, which is also the embedding dimension for attention mechanisms
TCNdim = [16,22,16]	The embedding dimensions for TCN
delta = 0.0010260719787192482	HuberLoss loss parameter, balancing MSE and MAE
learning rate = 0.004586593782220711	The learning rate for the optimization process
eps = 0.0026844107678024158	The epsilon parameter for the Adam optimizer
batchsize = 32	The size of data batches for training
training parameters(12to24)	**interpretations**
hdim =140	The embedding dimension for node2vec, which is also the embedding dimension for attention mechanisms
TCNdim = [140,146,140]	The embedding dimensions for TCN
delta = 0.0024512784957632866	HuberLoss loss parameter, balancing MSE and MAE
learning rate = 0.0015447670497578904	The learning rate for the optimization process
eps = 0.006474305690193212	The epsilon parameter for the Adam optimizer
batchsize = 32	The size of data batches for training

The GTV-STP was trained with epoch 500 and train 100. Change curves of the Huber Loss with different epochs are shown in Figure 10.

It is shown that the GTV-STP can give more precise predictions to the most indicators by means of the 6to2 dataset when the train reaches 28 and the epoch reaches 95. After a comprehensive evaluation, we take epoch=95 and train=28

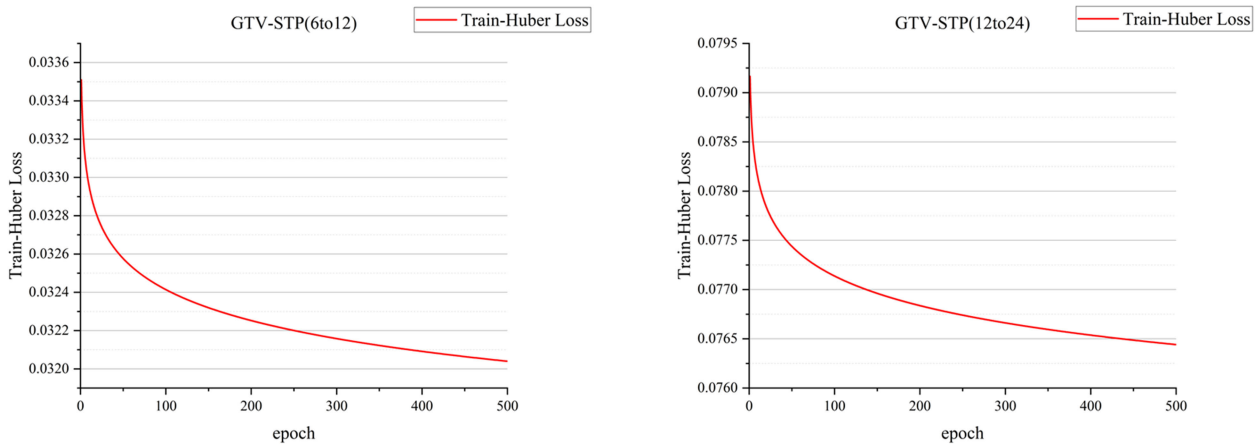

FIGURE 10. Change curves of the huber loss with different epochs.

TABLE 4. GTV-STP train errors.

Dataset	Variables	Epoch=95, train=28		
		Train-MAE	Train-RMSE	Train-Huber Loss
6to12	WT	0.980793	0.801943	
	pH	1.004877	0.822724	
	DO	1.004399	0.779118	
	COD_{MN}	0.993349	0.576285	
	NH_3-N	0.939310	0.426148	
	P	1.014218	0.702301	
				0.032185
Dataset	Variables	Epoch=144, Train=17		
		Train-MAE	Train-RMSE	Train-Huber Loss
12to24	WT	0.965986	0.794441	
	pH	1.002669	0.818754	
	DO	0.996550	0.772304	
	COD_{MN}	0.995476	0.576301	
	NH_3-N	0.965699	0.426127	
	P	1.005028	0.707151	
				0.076821

TABLE 5. GTV-STP verification errors.

Dataset	Variables	Epoch=95, train=28		
		Val-MAE	Val-RMSE	Val-Huber Loss
6to12	WT	0.856020	0.766158	
	pH	0.985278	0.785930	
	DO	0.963421	0.745162	
	COD_{MN}	0.946406	0.622119	
	NH_3-N	0.990082	0.736130	
	P	0.918676	0.613817	
				0.033561
Dataset	Variables	Epoch=144, Train=17		
		Val-MAE	Val-RMSE	Val-Huber Loss
12to24	WT	0.830978	0.740471	
	pH	0.975974	0.777429	
	DO	0.943826	0.730165	
	COD_{MN}	0.963535	0.631536	
	NH_3-N	0.980354	0.732465	
	P	0.930738	0.608695	
				0.079183

as the final number of training rounds to be used for the 6to12 dataset. In the same way, the train and validation errors using epoch=144 and train=17 as the final number of training rounds for the 12to24 dataset, are shown in Table 4 and Table 5.

The Huber loss of the GTV-STP for the 6to12 dataset is smaller than that for the 12to24 dataset, which demonstrates the higher accuracy of the GTV-STP framework for small sample predictions. The introduction of the Attention mechanism to the GTV-STP framework enables the the framework to make more accurate predictions for time series problems. Meanwhile, the spatial graph embedding enables the GTV-STP framework to distinguish different sites, to learn the water quality sequential information from different sites and to show a its better applicability. The above reasons will make the GTV-STP framework more suitable for water quality information predictions.

The following baseline models are established and trained with the same two datasets of 6to12 and 12to24 as the GTV-STP used.

1) LSTM-CNN-ATT
2) LSTMNet
3) MLP
4) GTV-STP ablation study (remove the spatial and temporal information)

Comparisons of the prediction accuracy between the GTV-STP and baseline models are shown in Table 6 and Table 7. The GTV-STP ablation study (remove the spatial and temporal information) is shown in Table 8.

As can be seen from these tables, for the small-scale 6to12 dataset, the GTV-STP framework is superior to the baseline models in water quality indicator predictions with a val-Huber Loss of 0.033561, while the MLP model give the worst performance with a val-Huber Loss of 3.438210. It is shown that the GTV-STP framework is highly accurate in smaller-scale predictions. For the larger-scale12to24 dataset, the val-Huber Loss of the GTV-STP framework is 0.079183, whereas that of the LSTM-CNN-ATT is 1.176752 and the other baselines give the similar performances as the LSTM-CNN-ATT. So the GTV-STP framework is also suitable for the prediction of larger-scale samples.

TABLE 6. 6to12 dataset.

model /frame	Variables	6to12		
		Val-MAE	Val-RMSE	Val-Huber Loss
GTV-STP	WT	0.856020	0.766158	
	pH	0.985278	0.785930	
	DO	0.963421	0.745162	
	COD_{MN}	0.946406	0.622119	
	NH_3-N	0.990082	0.736130	
	P	0.918676	0.613817	
				0.033561
LSTM-CNN-ATT	WT	0.365130	0.245334	
	pH	0.445684	0.291896	
	DO	0.604705	0.458726	
	COD_{MN}	0.773135	0.401403	
	NH_3-N	0.789529	0.559803	
	P	0.664184	0.301269	
				0.059511
LSTMNet	WT	0.886552	0.803087	
	pH	0.982476	0.780333	
	DO	0.981273	0.768341	
	COD_{MN}	0.946468	0.623462	
	NH_3-N	0.996851	0.735546	
	P	0.911845	0.630233	
				0.416798
MLP	WT	0.379271	0.232264	
	pH	0.440109	0.276347	
	DO	0.449416	0.299779	
	COD_{MN}	0.769022	0.394672	
	NH_3-N	0.579433	0.330514	
	P	0.807309	0.313057	
				3.438210

TABLE 7. 12to24 dataset.

model /frame	Variables	12to24		
		Val-MAE	Val-RMSE	Val-Huber Loss
GTV-STP	WT	0.830978	0.740471	
	pH	0.975974	0.777429	
	DO	0.943826	0.730165	
	COD_{MN}	0.963535	0.631536	
	NH_3-N	0.980354	0.732465	
	P	0.930738	0.608695	
				0.079183
LSTM-CNN-ATT	WT	0.485550	0.363846	
	pH	0.518976	0.358358	
	DO	0.643964	0.496044	
	COD_{MN}	0.902777	0.549390	
	NH_3-N	0.651680	0.413603	
	P	0.738813	0.373433	
				1.176752
LSTMNet	WT	0.462271	0.338698	
	pH	0.513487	0.357650	
	DO	0.612991	0.469757	
	COD_{MN}	0.831951	0.484942	
	NH_3-N	0.781999	0.549883	
	P	0.757468	0.382938	
				0.996955
MLP	WT	0.440629	0.309935	
	pH	0.490103	0.320186	
	DO	0.494859	0.344597	
	COD_{MN}	0.803937	0.438088	
	NH_3-N	0.574602	0.333156	
	P	0.762651	0.340970	
				0.319198

From the comparison between Table 8 and Table 5, it is evident that the GTV-STP ablation study, which involves removing the spatial and temporal information, results in significantly higher Val-Huber Loss on both the 6to12 and 12to24 datasets compared to GTV-STP. The reason for the substantial increase in Val-Huber Loss in the validation set, following the removal of time and space information in the ablation experiment, is that temporal information is crucial for capturing the dynamic characteristics of data as it changes over time, while spatial information provides the model with context for understanding the internal structure and patterns of the data. The absence of these pieces of information restricts the model's ability to learn complex patterns from the data, especially when these patterns depend on the continuity of time and spatial structures. The interaction between temporal and spatial information is particularly critical for understanding certain datasets, such as water quality information. By removing these crucial pieces of information, the model is unable to effectively learn the true distribution of the data, severely impacting its generalization capability and prediction accuracy. This is ultimately reflected in the higher Val-Huber Loss on the datasets, with the Val-Huber Loss reaching 1.228409 on the 6to12 dataset and 6.526153 on the 12to24 dataset.

TABLE 8. GTV-STP ablation study (remove the spatial and temporal information) verification errors.

Dataset	Variables	Val-MAE	Val-RMSE	Val-Huber Loss
6to12	WT	0.421228	0.315583	
	pH	0.568560	0.420447	
	DO	0.646620	0.497306	
	COD_{MN}	0.855414	0.521825	
	NH_3-N	0.808117	0.567801	
	P	0.724479	0.409261	
				1.228409
Dataset	Variables	Val-MAE	Val-RMSE	Val-Huber Loss
12to24	WT	0.487886	0.372674	
	pH	0.539690	0.394461	
	DO	0.617094	0.472320	
	COD_{MN}	0.877071	0.523048	
	NH_3-N	0.799139	0.580523	
	P	0.735286	0.364807	
				6.526153

IX. CONCLUSION

In conclusion of this paper, the GTV-STP framework is proposed to solve the water quality predictions using the spatial embedding method of introducing the spatial hydrological features into the the water quality information. The predictions are performed on two datasets of 6to12 in smaller-scale and 12to24 in larger-scale. Results show that the GTV-STP framework has the highest accuracy with the Val-Huber Loss of 0.033561 on 6to12 and 0.079183 on 6to12, whereas the MLP among the baseline models has the worst accuracy with Val-Huber Loss of 3.438210 on 6to12. In future, water quality predictions using other more precise frameworks containing spatial and time series information under deep learning will be one of the research directions.

REFERENCES

[1] Z. Li, Q. Mei, Y. Li, P. Wang, Y. Yang, and W. Hu, "Prediction and analysis of ship traffic flow based on a space-time graph traffic computing framework," presented at the IEEE 20th Int. Conf. Embedded Ubiquitous Comput. (EUC), 2022.

[2] L. Heesung, A. Hyunuk, K. Haedo, and L. Jeaju, "Prediction of pollution loads in the Geum River upstream using the recurrent neural network algorithm," *Korean J. Agricult. Sci.*, vol. 46, pp. 67–78, Mar. 2019, doi: 10.7744/kjoas.20180085.
[3] F. J. Chang, P. A. Chen, L. C. Chang, and Y. H. Tsai, "Estimating spatio-temporal dynamics of stream total phosphate concentration by soft computing techniques," *Sci. Total Environ.*, vol. 562, pp. 228–236, Aug. 2016, doi: 10.1016/j.scitotenv.2016.03.219.
[4] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1–9.
[5] Y. Chen, L. Song, Y. Liu, L. Yang, and D. Li, "A review of the artificial neural network models for water quality prediction," *Appl. Sci.*, vol. 10, no. 17, p. 5776, Aug. 2020, doi: 10.3390/app10175776.
[6] M. A. Antar, I. Elassiouti, and M. N. Allam, "Rainfall-runoff modelling using artificial neural networks technique: A blue Nile catchment case study," *Hydrol. Processes*, vol. 20, no. 5, pp. 1201–1216, 2006, doi: 10.1002/hyp.5932.
[7] S. Khani and T. Rajaee, "Modeling of dissolved oxygen concentration and its hysteresis behavior in rivers using wavelet transform-based hybrid models," *CLEAN–Soil, Air, Water*, vol. 45, no. 2, 2016, doi: 10.1002/clen.201500395.
[8] D. Antanasijević, V. Pocajt, D. Povrenović, A. Perić-Grujić, and M. Ristić, "Modelling of dissolved oxygen content using artificial neural networks: Danube River, North Serbia, case study," *Environ. Sci. Pollut. Res.*, vol. 20, no. 12, pp. 9006–9013, Dec. 2013, doi: 10.1007/s11356-013-1876-6.
[9] W. Tian, Z. Liao, and X. Wang, "Transfer learning for neural network model in chlorophyll-a dynamics prediction," *Environ. Sci. Pollut. Res.*, vol. 26, no. 29, pp. 29857–29871, Oct. 2019, doi: 10.1007/s11356-019-06156-0.
[10] H. Yuan, G. Li, Z. Bao, and L. Feng, "Effective travel time estimation: When historical trajectories over road networks matter," presented at the ACM SIGMOD Int. Conf. Manag. Data, 2020.
[11] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. KDD*, Aug. 2016, pp. 855–864, doi: 10.1145/2939672.2939754.
[12] W. Du, Y. Wang, and Y. Qiao, "Recurrent spatial–temporal attention network for action recognition in videos," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1347–1360, Mar. 2018, doi: 10.1109/TIP.2017.2778563.
[13] J. Fan, K. Zhang, Y. Huang, Y. Zhu, and B. Chen, "Parallel spatio-temporal attention-based TCN for multivariate time series prediction," *Neural Comput. Appl.*, vol. 35, no. 18, pp. 13109–13118, 2021, doi: 10.1007/s00521-021-05958-z.
[14] X. Wei, T. Zhang, Y. Li, Y. Zhang, and F. Wu, "Multi-modality cross attention network for image and sentence matching," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020.
[15] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.
[16] Y. He and J. Zhao, "Temporal convolutional networks for anomaly detection in time series," presented at the ICAACE/J. Phys., Conf., 2019.
[17] C. Saranya and G. Manikandan, "A study on normalization techniques for privacy preserving data mining," *Int. J. Eng. Technol.*, vol. 5, no. 3, pp. 2701–2704, Jun./Jul. 2013.
[18] NetworkX Community. (2023). *NetworkX. [Software]*. Accessed: Mar. 10, 2024. [Online]. Available: https://networkx.org/

YUBO LEI was born in Xi'an, Shaanxi, China, in 1993. He received the B.S. degree in civil engineering informatization from Changchun University of Architecture and Civil Engineering, in 2017. He is currently pursuing the master's degree in electronics and information engineering with Nanchang Institute of Technology. His main research interests include water environmental protection and water resources informatization.

BO HU was born in Benxi, Liaoning, China, in 1995. He received the B.S. degree in civil engineering informatization from China University of Mining and Technology-Beijing, in 2018. He is currently pursuing the master's degree in electronics and information engineering with Nanchang Institute of Technology. His main research interests include natural hazard forecast and civil engineering informatization.

HONGYUAN HUANG was born in Linwu, Hunan, China, in 1979. She received the bachelor's, M.S., and Ph.D. degrees in hydraulic engineering from Chongqing Jiaotong University (CJU), Chongqing, in 2002, 2005, and 2021, respectively. She is currently a Lecturer with the School of Civil and Architecture Engineering, Nanchang Institute of Technology. Her research interests include composite material structure design, structure online monitoring, and geotechnical engineering.

YOUPING LIU was born in Zhuzhou, Hunan, China, in 1979. He received the Ph.D. degree from the School of Resource and Environmental Engineering, Beijing University of Science and Technology, in 2015. He is currently a Professor with the School of Civil Engineering and Architecture, Nanchang Institute of Technology. His research interests include fiber optic monitoring technology, civil engineering informatization, and rock and soil disaster prevention and treatment.