IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 April 2024, accepted 29 April 2024, date of publication 6 May 2024, date of current version 21 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3396913

RESEARCH ARTICLE

TEC Prediction Based on Att-CNN-BiLSTM

HAIJUN LIU[1], HAORAN WANG[1], JING YUAN[2], LIANGCHAO LI[1], AND LILI ZHANG[1]
[1]Institute of Intelligent Emergency Information Processing, Institute of Disaster Prevention, Langfang 065201, China
[2]School of Information Engineering, Institute of Disaster Prevention, Langfang 065201, China

Corresponding author: Jing Yuan (yuanjing@cidp.edu.cn)

This work was supported in part by the Natural Science Foundation of Hebei Province under Grant D2023512004; in part by Langfang City Science and Technology Support Plan Project under Grant 2023011105; in part by the Science and Technology Innovation Program for Postgraduate Students in Institute of Disaster Prevention (IDP) subsidized by the Fundamental Research Funds for Central Universities under Grant ZY20240339; and in part by the Natural Science Foundation of Fujian Province, China, under Grant 2022J01117.

ABSTRACT Prediction of Total Electron Content (TEC) in the ionosphere is vital to improve the accuracy of satellite positioning, navigation and remote sensing systems. Most existing TEC prediction methods ignored the local variation patterns between various positions within the TEC sequence, resulting in limited prediction accuracy. To address this issue, this paper combined attention techniques, convolutional neural networks (CNN) and bidirectional long short-term memory networks (BiLSTM) to propose the Att-CNN-BiLSTM model. In the proposed model, CNN is used to extract positional features, BiLSTM is used to extract bidirectional temporal features, and attention technique is used to adaptively weight the features. This paper selected six locations in China, each with a six-year TEC sequence, including three years of high solar activity and three years of low solar activity. The paper first conducted ablation experiments, and the results showed that adding CNN and Attention can effectively improve prediction performance. Then, the proposed model was compared with LSTM and GRU. The experimental results show that compared with LSTM and GRU, the average RMSE of Att-CNN-BiLSTM in six regions decreased by 10.28% and 16.92% in high solar activity years, and by 11.82% and 8.92% in low solar activity years, respectively. The paper also conducted comparative experiments within one week of the magnetic storms, and the results showed that during the magnetic storm period, the RMSE of the proposed model decreased by 35.50% and 37.35% compared to LSTM and GRU. The R^2 s of the proposed model are also higher than those of the comparison models in all cases.

INDEX TERMS Attention mechanism, CNN-BiLSTM, ionosphere, total electron content, local variation patterns.

I. INTRODUCTION

The ionosphere is an important region of the earth's space. It is coupled with the magnetosphere upward and affected by the lower atmosphere downward, where air molecules are partially ionized by the radiation of solar ultraviolet rays and X-rays, producing negatively charged electrons and positively charged ions.When electromagnetic wave signals transmitted by the Global Navigation Satellite System (GNSS) propagate through the ionosphere, they can incur errors of over ten meters, which is one of the main sources of error in global navigation satellite applications, especially single frequency global navigation satellite applications [1], [2], [3]. The Total Electron Content (TEC) of the ionosphere is a key parameter that describes the characteristics of ionospheric variation, which is widely used for satellite ionospheric delay correction.Precisely predicting ionospheric TEC can enhance the precision of navigation and positioning, holding significant value in mitigating uncertainty within services reliant on the Global Navigation Satellite System(GNSS).Therefore, monitoring and predicting TEC is an important research topic in space weather [4], [5].

Influenced by many factors such as solar activity, geomagnetic activity and low-level atmospheric disturbance, ionospheric TEC has very complex spatiotemporal changes. At present, there is no accurate physical prediction model

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

for TEC [6]. So far, there are three main types of short-term prediction methods for ionospheric TEC: ionospheric empirical models, statistical models, and artificial neural network models [7], [8], [9]. The ionospheric empirical models include the International Reference Ionosphere (IRI) [10], [11], [12], Bent [13], [14], NeQuick [15], [16], Klobuchar [17], [18], [19], etc. These ionospheric empirical models mainly reflect the average changes in the ionosphere and are often used for long-term global TEC predictions. However, their accuracy in short-term TEC predictions in local areas is relatively low [20]. The statistical models for TEC prediction mainly include auto- correlation (AR) analysis [21], auto-regressive moving average (ARMA) [22], [23], autoregressive integrated moving average (ARIMA) [24], [25], etc. These early statistical models were mostly linear, making it difficult to accurately predict the complex nonlinear changes of TEC [26].

Recently, with the development of artificial intelligence technology, machine learning and deep learning techniques have been widely applied in TEC prediction, Table 1 summarizes the current TEC prediction models. Artificial neural network (ANN) has attracted the attention of ionospheric researchers due to its powerful nonlinear representation ability and has been applied to ionospheric TEC prediction with higher accuracy than empirical models [27], [28], [29], [30], [31], [32]. However, ANN only considers the spatial position of the data and cannot characterize the temporal changes in the TEC sequence, resulting in significant errors in TEC prediction [33], [34], [35]. Recurrent neural network (RNN) is a chain-connected neural network that takes sequence data as input and recurses continuously in the evolution direction of the sequence. It is a deep learning model that can characterize both the spatial characteristics and the temporal characteristics of data, and is widely used in TEC prediction [36]. However, when modeling a long time series, RNN has the problem of gradient vanishing, which will cause the model to forget the previous data. Therefore, RNN is unable to remember the characteristics in the long time series, so it is not competent for the prediction task of long time series. Long short-term memory network (LSTM) uses the gating mechanism to remember information in long time series, overcoming the data forgetting problem of RNN [37], which can improve the prediction accuracy of long time series and is widely used in TEC prediction [38], [39], [40]. Some researchers have also attempted to improve the structure of LSTM. For example, Graves and Schmidhuber et al. [41] proposed bidirectional LSTM (BiLSTM) by overlaying two LSTM layers from different directions along the time dimension. One LSTM layer in BiLSTM deals with forward sequences and the other deals with reverse sequences, which has been proved to have better predictive performance than unidirectional LSTM. This idea of bidirectional processing of time series was later used in TEC prediction. For example, Sivakrishna et al. [42] applied BiLSTM for TEC prediction in India, and its prediction accuracy was better than that of LSTM. In recent years, researchers have also combined LSTM with CNN for TEC prediction. For example, Ruwali et al. [43] combined CNN with LSTM and proposed LSTM-CNN. Tang et al. [44] used the CNN-LSTM-Attention model to predict ionospheric TEC at 24 GNSS stations in China. Their experimental results showed that combining LSTM with CNN can improve prediction performance. However, in CNN-LSTM-Attention, they only considered the forward temporal features of TEC and did not utilize the reverse temporal features.

In summary, for TEC prediction problems, BiLSTM outperforms unidirectional LSTM due to its consideration of the bidirectional temporal characteristics of input sequences; CNN has the excellent ability to extract local position features. Therefore, this paper combines CNN and BilSTM to extract bidirectional temporal features and position features respectively. Attention mechanism can adaptively weight features, so we add it to the proposed model. The model is named Att-CNN-BiLSTM.

To test the performance of the proposed model, six locations in China were selected as the research area. At each location, six years of TEC and Disturbance storm time (Dst) (three years from high solar activity and three years from low solar activity) were selected as the experimental data. In the experiment, 7 consecutive days of historical TEC and Dst were used as input to predict the next day's TEC. 4 years of data were used for training and 2 years for testing. Ablation experiments were first conducted both in high and low solar activity years to verify the effectiveness of CNN and attention in improving the model's prediction performance. Experimental results showed that adding CNN and attention can improve the prediction performance. The proposed Att-CNN-BiLSTM was then compared with LSTM and GRU. The experimental results show that the Att-CNN-BiLSTM proposed in this paper is superior to the comparison models in both low solar activity years and high solar activity years. The main contributions of this paper are as follows: (1) For the first time, CNN and BiLSTM were combined together to simultaneously extract bidirectional temporal features and local positional features from TEC sequences. The ablation experiment indicated that the combination of CNN and BiLSTM effectively improved the predictive performance. (2) This paper have added an attention mechanism to the model to adaptively weight the bidirectional temporal features and local positional features. The results of the ablation experiment indicated that adding attention mechanism significantly improved predictive performance.

(3) In six locations in China, Att-CNN-BiLSTM was compared with LSTM and GRU, and the results showed that the proposed model outperformed the comparison models in both high and low solar activity years, as well as during magnetic storms.

The remainder of this paper is structured as follows. Section II introduces the data and data preprocessing. Section III provides a detailed description of the proposed model. Section IV presents evaluation metrics,

TABLE 1. The summary of the neural network models for ionospheric TEC prediction.

Reference	Prediction method	Prediction content or results
Watthanasangmechai et al. 2012[27]	neural network (NN)	This paper demonstrated a neural network-based model that effectively predicts total electron content (TEC) in the Chumphon of Thailand during low solar activity.
Sahu et al. 2021 [28]	neural network (NN)	This paper proposed a neural network model to accurately predict the vertical total electron content (VTEC) in the equatorial anomaly peak area of Bhopal, India, showing better performance compared with IRI.
Habarulema et al. 2009 [30]	neural network (NN)	This paper used NN for TEC prediction in South Africa and got better performance than IRI-2001.
Mallika et al. 2019 [29]	PCA-NN	This paper combined principal component analysis (PCA) and artificial neural network (ANN) to propose PCA-NN for TEC prediction in Japan. heir research showed that PCA-NN outperformed NN and IRI.
Cesaroni et al. 2020 [31]	nonlinear autoregressive neural network with external input (NARX)	This paper proposed a nonlinear autoregressive neural network (NARX) technology for the 24-hour TEC on 18 locations, with RMSE ranging between 3 and 5 TECU.
Dabbakuti and G, 2019 [32]	singular spectrum analysis-artificial neural network (SSA-ANN)	This paper proposed combined singular spectrum analysis and artificial neural network to propose an hourly TEC prediction model named SSA-ANN. Experiments at Bangalore site during 2009-2017 showed that it outperformed ARMA and IRI-16.
Inyurt and Sekertekin, 2019[33]	artificial neural network (ANN)	This paper used ANN for TEC prediction atTurkey's ANKR GPS station in 2015. Their model was better than artificial neural network-based ionospheric model (ANNIM)
Huang and Yuan, 2014 [34]	radial basis function (RBF) neural network	This paper used the radial basis function (RBF) neural network to predict the TEC at BJFS, WUHN and KUNM. Results shows that their RBF is better than the BP network.
Habarulema et al. 2007 [35]	neural network (NN)	This paper used neural networks to predict the TEC in Springbok, South Africa. and got better performance IRI-2001.
Yuan Tianjiao et al. 2018 [36]	Recurrent Neural Network (RNN)	This paper used RNN to predict the TEC of Beijing Station, and their accuracy is better than that of BP network.
Kaselimi et al. 2020 [38]	Long Short-Term Memory (LSTM)	This paper used LSTM to predict TEC at 6 different sites, and the effect is better than ARMA and AR.
Liu et al. 2020 [39]	long short-term memory (LSTM)	This paper used LSTM to predict global TEC from October 19 to December 31, 2014 and from May 27 to December 31, 2016. And their model is better than IRI-2016 and NeQuick-2.
Reddybattula et al. 2022 [40]	long short-term memory (LSTM)	This paper used LSTM to predict the TEC in 2018 at the Indian low-latitude Bangalore (IISC) location. The prediction effect of this model is significantly better than IRI-2016.
Sivakrishna et al. 2022 [42]	bidirectional long short-term memory (Bi-LSTM)	This paper used BiLSTM to predict TEC of 26 GPS stations in India. The prediction effect of this model is significantly better than that of ANN and LSTM.
Ruwali et al. 2021 [43]	Hybrid Deep Learning Model (LSTM-CNN)	This paper used LSTM-CNN to predict the TEC of Bengaluru, Guntur, and Lucknow GPS stations in 2016 and compared it with NN, GRU, and LSTM. The prediction effect of LSTM-CNN is significantly better than those of the comparison models.
Tang et al. 2022 [44]	CNN-LSTM-Attention	This paper proposed a CNN-LSTM-Attention model to predict the TEC of 24 locations in China in 2018 and compared it with NeQuick, LSTM and CNN-LSTM. The prediction effect of their model is significantly better than those of all the comparison models in different months and under different geomagnetic conditions.

experimental results and analyses. Section V is the discussion. Section 6 concludes the paper.

II. DATA AND DATA PREPROCESSING

A. ABBREVIATIONS AND ACRONYMS

The TEC data used in this paper are provided by the Center for Orbit Determination of Europe (CODE), with a time resolution of 2 hours.[1]

6 locations (numbered $P_1, P_2, P_3, \cdots, P_6$) in China were selected as the research area, as shown in Figure 1, and the specific coordinates are shown in Table 2. In order to study the impact of different solar activity on the prediction performance of the models, six-year TEC data in each location were selected for experiments, including: three years of high solar activity data (January 1, 2000 to December 31, 2002) and three years of low solar activity data (January 1, 2007 to December 31, 2009). 4-year data as training samples (2000-2001,2007-2008), 2-year data as testing samples (2002,2009).

To improve prediction performance, in addition to TEC data, Dst data were also used as an auxiliary input. Dst data were downloaded from the geomagnetic and space magnetic

[1]https://cddis.nasa.gov/archive/gnss/products/ionex/

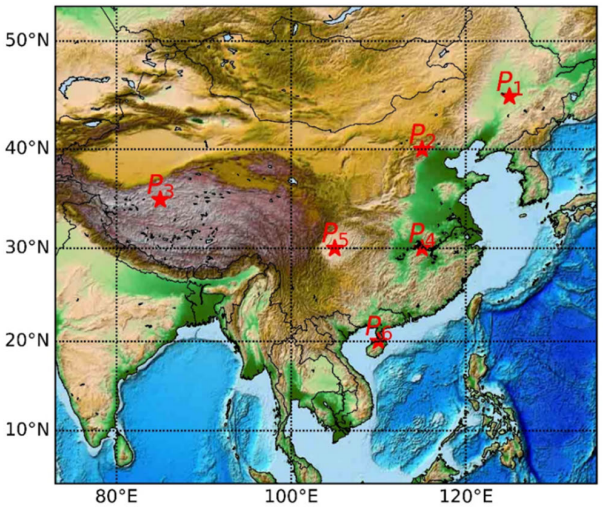


FIGURE 1. Selected geographical locations in this paper.

TABLE 2. Coordinates of the selected regions in this paper.

P_1	P_2	P_3
45°N,125°E	40°N,115°E	35°N,85°E
P_4	P_5	P_6
30°N,115°E	30°N,105°E	20°N,110°E

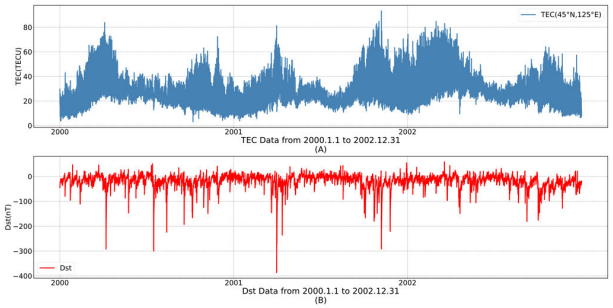


FIGURE 2. Data at P1 from 2000 to 2002: (A) TEC; (B) down-sampled Dst.

data analysis center of Kyoto university.[2] To align with TEC, Dst data were down-sampled every two hours. Figure 2 shows the down-sampled Dst data and TEC data at P1 from January 1, 2000, to December 31, 2002.

B. DATA PREPROCESSING

TEC and Dst data belong to time series data. In time series prediction tasks, only stationary nonrandom sequences can be directly predicted [45]. Therefore, before using TEC and Dst for prediction, stationarity tests and pure randomness tests were first performed on them.

[2]https://wdc.kugi.kyoto-u.ac.jp/dstdir/

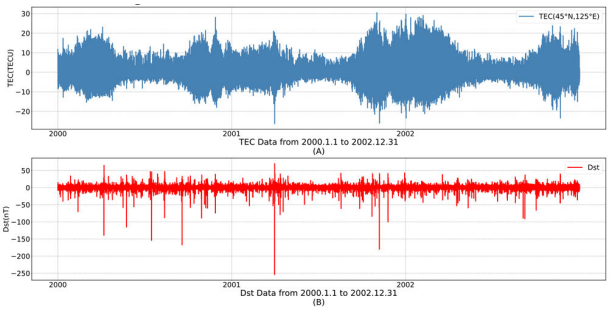


FIGURE 3. Data after first-order differential processing in P_1: (A) TEC; (B) Dst.

1) DATA STATIONARITY TEST

In this paper, the Augmented Dickey Fuller (ADF) method [46] was used to test the stationarity of TEC and Dst. The ADF test results indicated that both the selected TEC and Dst were non-stationary sequences and could not be directly used for TEC prediction. First-order difference was first used to transform them into stationary sequences. The formula for the first-order difference is as follows:

$$\Delta x_t = x_t - x_{t-1} \tag{1}$$

where, Δ is the first-order difference operator, x_t is the observation data (TEC or Dst) at a certain location at time t. Taking the location P_1 as an example, the TEC and Dst data after first-order differential processing are shown in Figure 3.

After first-order differential processing, both TEC and Dst data in the research area passed the ADF test, meaning that the TEC and Dst data after first-order differential processing became stationary.

2) PURE RANDOMNESS TEST

Even if TEC and Dst are processed to be stationary, they cannot be used for prediction if they are pure random sequences. Therefore, pure randomness testing is still needed for the first-order differential TEC and Dst. Tthe LB (Ljung Box) method was used to perform pure randomness tests after the first-order difference processing [47]. The LB test results indicated that they were not purely random data and could be used for prediction.

3) DATA NORMALIZATION

To reduce the adverse effects caused by singular data, Min-Max normalization was used to map the first-order differential TEC and Dst between $[1, 0]$. The calculation formula for Min-Max is as follows:

$$x_{i,t} = \frac{x_{i,t} - \min\limits_{1\le j\le T} x_{i,j}}{\min\limits_{1\le j\le T} x_{i,j} - \max\limits_{1\le j\le T} x_{i,j}} \tag{2}$$

where, $x_{i,t}$ is the TEC or Dst data of a certain position i at time t, $i \in \{P_1, P_2, P_3, P_4, P_5, P_6\}$, $\min\limits_{1\le j\le T} x_{i,j}$ and $\max\limits_{1\le j\le T} x_{i,j}$ represents the minimum and maximum values of all TEC or Dst data at position i, respectively.

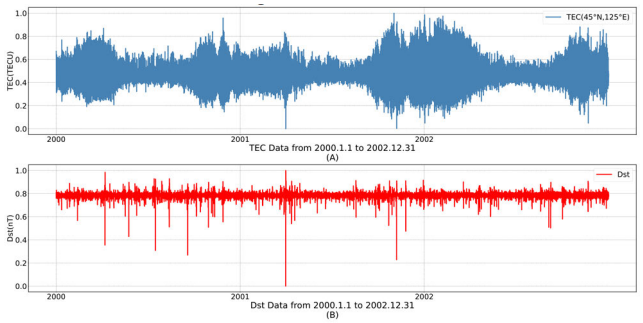


FIGURE 4. Data after Min-Max normalization: (A) TEC; (B) Dst.

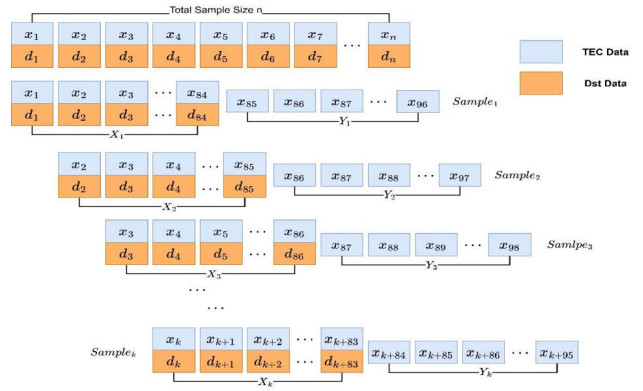


FIGURE 5. Schematic diagram of sample production process.

Taking P1 as an example, after first-order differential processing and Min-Max normalization, the processed TEC and Dst are shown in Figure 4.

4) SAMPLE PRODUCTION PROCESS

Experimental samples were made after completing the stationarity test, first-order difference processing, pure randomness test, and Min-Max normalization. In this paper, an 8-day sliding window method was used to create sample set D, where the TEC and Dst from the first 7 days were used as input X, and the 12 TECs on the 8-th day were used as output Y. Between adjacent samples, a slide of 2 hours was made each time. The composition of the sample set D is shown in formulas (3), (4), and (5).

$$D = \{(X_i, Y_i)|i = (0, 1, 2, \cdots, n)\} \tag{3}$$

$$X_i = (< x_i, d_i >, < x_{i+1}, d_{i+1} >, \cdots, < x_{i+83}, y_{i+83} >) \tag{4}$$

$$Y_i = (y_{i+84}, y_{i+85}, \cdots, y_{i+95}) \tag{5}$$

Among them, (X_i, Y_i) is the i-th sample, where X_i is its input vector and $Y_i$is its output vector; X_i consists of 84 preprocessed TEC (represented by x_i) and Dst (represented by d_i) pairs (7 days, 12 values per day). The sample production process is shown in Figure 5.

The final samples were divided into training and test sets. The training set was used to train the model, and the testing set was used to evaluate the model's predictive performance.

TABLE 3. Sample distribution of training and testing sets in this paper.

Data Set	Training set		Test set	
	high solar activity (2000,2001)	low solar activity (2007,2008)	high solar activity (2002)	low solar activity (2009)
Number of samples	8674	8674	4282	4282
Total	17348		8758	

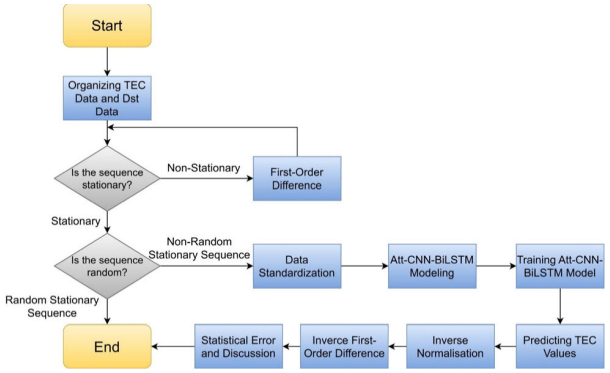


FIGURE 6. Experimental flowchart in this paper.

The distribution of the final training and testing samples is shown in Table 3.

The predicted results obtained by the model required inverse normalization and inverse first-order difference processing to obtain the final predicted value of TEC. The entire experimental process is shown in Figure 6.

III. MODEL

The proposed Att-CNN-BiLSTM integrates three popular technologies in deep learning, namely CNN, BiLSTM and attention mechanism. We will introduce them separately in the following sections.

A. CNN

Convolutional neural network (CNN), which was proposed by Lecun et al. in 1998 [48], has achieved remarkable achievements in image and video recognition [49], [50], emotion recognition [51], natural language processing [52] and other tasks, and can also be effectively applied to time series analysis [53]. A CNN network typically consists of a stack of convolutional layers, pooling layers, and fully connected layers. A convolutional layer often consists of multiple convolutional kernels, each of which extracts a certain local position feature from the sample. Multiple convolutional kernels will extract multiple features to generate feature maps; The pooling layer reduces the spatial size of the feature map through down-sampling, which can reduce the computational complexity of the model and also reduce the risk of overfitting. By stacking multiple convolutional and pooling layers,

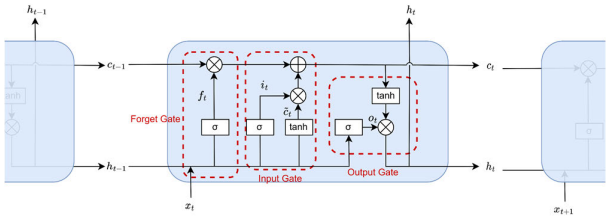


FIGURE 7. Basic internal structure of an LSTM unit.

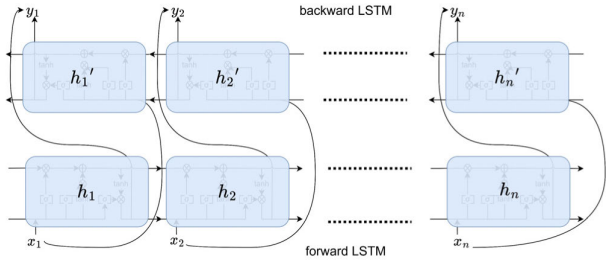


FIGURE 8. Structure of BiLSTM network.

a large number of local location features of different scales can be extracted. In this paper, CNN is used to extract local position features from the input TEC and Dst.

B. LSTM AND BILSTM

LSTM was proposed by Kaselimi et al. in 1997 [38] and is widely used in time series prediction [54]. Each LSTM network is composed of several LSTM units. Each LSTM unit is composed of 3 gate structures (as shown in Figure 7): the input gate i_t, the forget Gate f_t and the output gate o_t. These three gate structures are connected by memory cell unit state c_t to selectively select features in the network. The calculation formulas for each part are as follows:

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{6}$$
$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{7}$$
$$\tilde{C}_t = tanh\left(W_C\left[h_{t-1}, x_t\right] + b_C\right) \tag{8}$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{9}$$
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{10}$$
$$h_t = o_t * tanh\left(c_t\right) \tag{11}$$

where, $*$ represents Hadamard product, σ represents sigmoid activation function, and *tanh* represents hyperbolic tangent activation function, W_i, W_f, W_c, W_o represent the weight matrix of each gate separately, b_i, b_f, b_c, b_o are the bias matrix of each gate. These weight matrices and bias matrices are initialized randomly and then learned through optimization algorithms during the training process of the model. x_t represents the input data of the network at time t, $\tilde{c}_t$ denotes the updated value of memory cell unit state at time t, h_t stands for the LSTM hidden layer state at time t, i_t is the input gate of the LSTM unit at time t, which determines how much feature information in $\tilde{c}_t$ is used to update c_t, f_t is the forget gate, which determines how much information was retained in the memory unit of LSTM at the previous moment, $o_t$is the output gate.

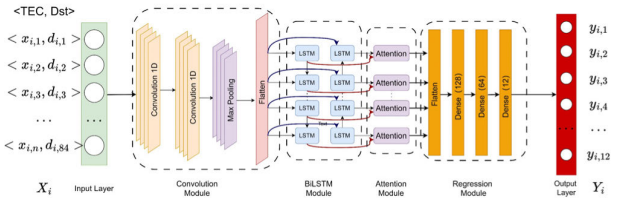


FIGURE 9. Structure of Att-CNN-BiLSTM.

BiLSTM is a variant of LSTM, which includes forward LSTM units $\overrightarrow{LSTM}$ and backward LSTM units $\overleftarrow{LSTM}$, and is able to simultaneously consider both forward and reverse features in the sequence [42]. The internal structure of BiLSTM is shown in Figure 8. The hidden layer state H_t of BiLSTM at time t is concatenated by forward $\overrightarrow{h_t}$ and reverse $\overleftarrow{h_t}$. The calculation formulas for $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are as follows:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}\left(h_{t-1}, x_t, c_{t-1}\right), t \in [1, T] \tag{12}$$
$$\overleftarrow{h_t} = \overleftarrow{LSTM}\left(h_{t+1}, x_t, c_{t+1}\right), t \in [T, 1] \tag{13}$$
$$H_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \tag{14}$$

Among them, [] represents the connection of vectors.

In this paper, the BiLSTM module is placed behind the CNN module to extract bidirectional temporal features from the results of the CNN, and obtain the bidirectional temporal feature vectors $[H_0, H_1, \ldots, H_n]$.

C. ATTENTION MECHANISM

An attention module was added after the BiLSTM layer. Research has shown that adding attention mechanisms can enhance the model's ability to focus on key features in long input sequences [55]. In Att-CNN-BiLSTM, the attention module receives bidirectional temporal features $[H_0, H_1, \ldots, H_n]$ output by the BiLSTM module, then calculate the similarity score $e_{t,i}$ between each feature H_i and the regression value y_t using the following attention function:

$$e_{t,i} = score\left(H_i, y_t\right) = V^T tanh(U[H_i; y_t]) \tag{15}$$

Then, the softmax function is used to normalize the similarity score $e_{t,i}$ and obtain the probability distribution $a_{t,i}$, as shown in formula (16), which represents the importance of H_i.

$$a_{t,i} = softmax\left(e_{t,i}\right) = \frac{exp(e_{t,i})}{\sum_{j=0}^{t} exp(e_{t,j})} \tag{16}$$

Finally, take $a_{t,i}$ as the weight of H_i and multiply it with H_i to obtain the weighted feature $\widetilde{H}_t$. The process of weighting features is shown in formula (17).

$$\widetilde{H}_t = \sum_{i=0}^{n} a_{t,i} H_i \tag{17}$$

Among them, U and V are learnable parameters during the training process of the neural network, and the final weighted bidirectional temporal feature vectors are $\left[\widetilde{H_0}, \widetilde{H_1}, \ldots, \widetilde{H_n}\right]$.

D. THE PROPOSED ATT-CNN-BILSTM

The structure of Att-CNN-BiLSTM is shown in Figure 9. The model consists of four modules: convolution module, BiLSTM module, attention module, and regression module.

Convolutional module: This module receives the input TEC and Dst sequences with a length of 84, performs preliminary feature extraction through convolution and pooling operations, and passes the extracted results to the BiLSTM module. This module contains two one-dimensional convolutional layers (convolution 1D), one maximum pooling layer (Max Pooling), and one flattening layer (Flatten). The first convolutional layer contains 128 convolutional kernels of size 9; The second convolutional layer contains 256 convolutional kernels of size 11; The Activation function of these two convolutional layers is relu, and the boundary filling method is valid. Each convolutional kernel is equivalent to a feature detector, which can extract the same features from different positions of input samples by sharing weights, forming a feature map. The 128 convolutional kernels of the first convolutional layer result in 128 feature maps. The second convolutional layer adopts a larger size to improve the perceptual field of feature extraction, resulting in 256 feature maps. In order to reduce memory consumption in calculations, a maximum pooling layer with a step size of 3 was added after two convolutional layers, which can down-sample each feature map and reduce the number of parameters. Then a flattening layer is used to flatten the feature maps extracted from the convolutional layer into one-dimensional feature vectors.

After being processed by the convolution module, the input sample becomes different position features extracted by different feature detectors. These different position features are fused to form a feature vector with a dimension of 5632, which contains a large number of local position features in the sample. This local position feature is passed to the BiLTSTM module for use.

BiLSTM module: Due to the periodic temporal variation of TEC, it is necessary to extract its temporal features. Therefore, a BiLSTM module was added after the convolution module to extract bidirectional temporal changes in local position features. This module contains 512 bidirectional LSTM units, and the activation function is relu. This module ultimately extracts bidirectional temporal features $[H_0, H_1, \ldots, H_n]$ and passes them on to the attention module;

Attention module: This module calculates the weight of each dimension in the high-dimensional feature vector transmitted by the BiLSTM module through formulas (15), (16), and (17), and uses this weight to weight the features transmitted by the BiLSTM layer to obtain a weighted feature vector $\left[\widetilde{H_0}, \widetilde{H_1}, \ldots, \widetilde{H_n}\right]$;

Regression module: This module receives the weighted feature vectors from the attention module and calculates the predicted values corresponding to the input samples. The module is composed of three full connection layers (Dense layer), the number of neurons is 128, 64 and 12 respectively, and the activation function is relu.

The detailed parameter configuration of Att-CNN-BiLSTM is shown in Table 4.

TABLE 4. Detailed parameter configuration of Att-CNN-BiLSTM.

Layer type	parameters
Conv1D	filters=32, kernel_size=2, activation='relu', kernel_regularizer=l2(0.0001)
Conv1D	filters=64, kernel_size=6, activation='relu', kernel_regularizer=l2(0.0001)
MaxPooling	pool_size=2
Flatten	-
BiLSTM	units=256, activation='relu', recurrent_regularizer=l2(0.0001)
Attention	-
Flatten	-
Dense	Units=64, activation='relu'
Dense	Units=32, activation='relu'
Dense	Units=12, activation='relu'

TABLE 5. Hyper-parameters during model training.

hyper-parameter name	settings
optimizer	adam
loss	" MAE "
learning rate	0.001
Batch size	128

IV. RESULTS

A. EVALUATION INDICATORS

In this paper, Root Mean Square Error (RMSE) and correlation coefficient (R^2) are used to evaluate the predictive performance of the model. Their calculation formulas are shown in formulas (18) and (19):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2} \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{m}\left(y_i - \bar{y}\right)^2} \tag{19}$$

TABLE 6. Comparison results of ablation experiments (the optimal values are bolded).

Model	Solar Activity	Evaluation indicator	P_1	P_2	P_3	P_4	P_5	P_6	Mean
BiLSTM	High (2002)	RMSE (TECU)	2.4976	3.0510	2.8416	5.3978	5.2471	5.2209	4.0427
CNN-BiLSTM			2.6109	2.8048	2.8880	5.1664	5.7028	4.8204	3.9989
Att-CNN-BiLSTM			**2.1443**	**2.5312**	**2.4212**	**4.7293**	**4.4112**	**4.5042**	**3.4569**
BiLSTM		R^2	0.9796	0.9747	0.9790	0.9667	0.9697	0.9826	0.9754
CNN-BiLSTM			0.9777	0.9787	0.9783	0.9695	0.9642	0.9853	0.9756
Att-CNN-BiLSTM			**0.9849**	**0.9826**	**0.9848**	**0.9744**	**0.9786**	**0.9872**	**0.9821**
BiLSTM	Low (2009)	RMSE (TECU)	0.8581	0.9708	0.7917	1.4944	1.6466	2.4924	1.3757
CNN-BiLSTM			0.7877	0.9310	0.7691	1.6443	1.7399	2.3606	1.3721
Att-CNN-BiLSTM			**0.6906**	**0.8273**	**0.6638**	**1.4166**	**1.5984**	**2.2145**	**1.2352**
BiLSTM		R^2	0.9077	0.8860	0.9450	0.9235	0.9127	0.9428	0.9196
CNN-BiLSTM			0.9222	0.8951	0.9481	0.9071	0.9025	0.9487	0.9206
Att-CNN-BiLSTM			**0.9402**	**0.9172**	**0.9613**	**0.9310**	**0.9177**	**0.9548**	**0.9370**

TABLE 7. Comparison results with LSTM and GRU (the optimal values are bolded).

Model	Solar activity	Evaluation indicator	P_1	P_2	P_3	P_4	P_5	P_6	Mean
LSTM	High (2002)	RMSE (TECU)	2.7359	2.5721	2.8225	5.0342	4.7827	5.1692	3.8528
GRU			2.8006	3.2247	2.8115	5.3805	5.3739	5.3740	4.1609
Ours			**2.1443**	**2.5312**	**2.4212**	**4.7293**	**4.4112**	**4.5042**	**3.4569**
LSTM		R^2	0.9755	0.9821	0.9793	0.9710	0.9748	0.9831	0.9776
GRU			0.9743	0.9718	0.9794	0.9669	0.9732	0.9817	0.9745
Ours			**0.9849**	**0.9826**	**0.9848**	**0.9744**	**0.9786**	**0.9872**	**0.9821**
LSTM	Low (2009)	RMSE (TECU)	0.9374	0.9687	0.8135	1.6321	1.7041	2.3486	1.4007
GRU			0.8312	0.8934	0.7234	1.6244	1.6800	2.3844	1.3561
Ours			**0.6906**	**0.8273**	**0.6638**	**1.4166**	**1.5984**	**2.2145**	**1.2352**
LSTM		R^2	0.8898	0.8865	0.9419	0.9085	0.9065	0.9492	0.9137
GRU			0.9134	0.9034	0.9541	0.9093	0.9091	0.9476	0.9228
Ours			**0.9402**	**0.9172**	**0.9613**	**0.9310**	**0.9177**	**0.9548**	**0.9370**

TABLE 8. Total parameters, training time and memory usage of each model.

Model	Total parameters	Training time (s/epoch)	Memory usage (MB)
LSTM	331,276	23	103.18
GRU	249,996	21	85.63
Att-CNN-BiLSTM	8,891,112	53	651.75

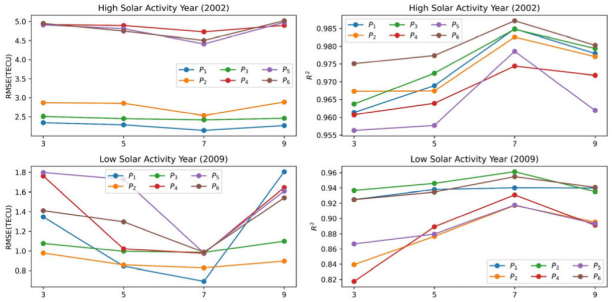


FIGURE 10. Line chart of prediction effect under different input lengths.

Among them, m is the total number of test samples, $\hat{y}_i$ is the predicted value for test sample i, y_i is the true value of test sample i, $\bar{y}$ is the average of the true values of all test samples. RMSE reflects the prediction error of the model, and the smaller the RMSE value of the model, the better its prediction performance; R^2 represents the fit degree between the predicted value of the model and the true value. The closer $R^2$is to 1, the better the prediction performance of the model.

B. OPTIMAL PARAMETERS OF THE MODEL

When training the model, it is necessary to specify training hyper-parameters. There are four important hyper-parameters during the model training process, namely learning rate, optimizer, loss function, and batch size. These hyper-parameters

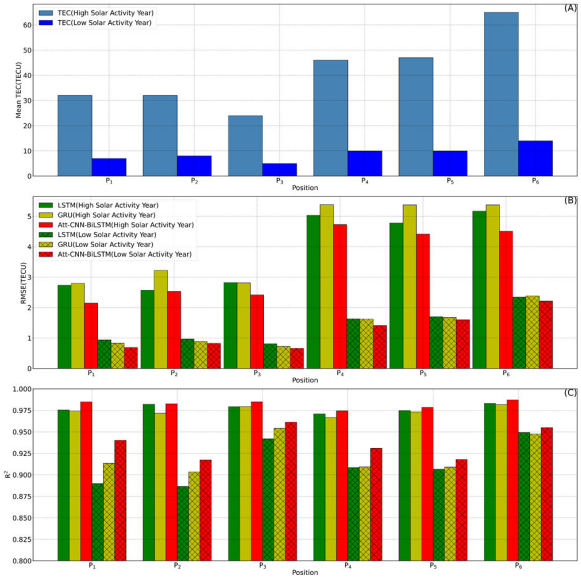


FIGURE 11. Comparison of TEC mean and prediction performance of various models: (A) TEC mean values;(B) RMSE;(C)R^2.

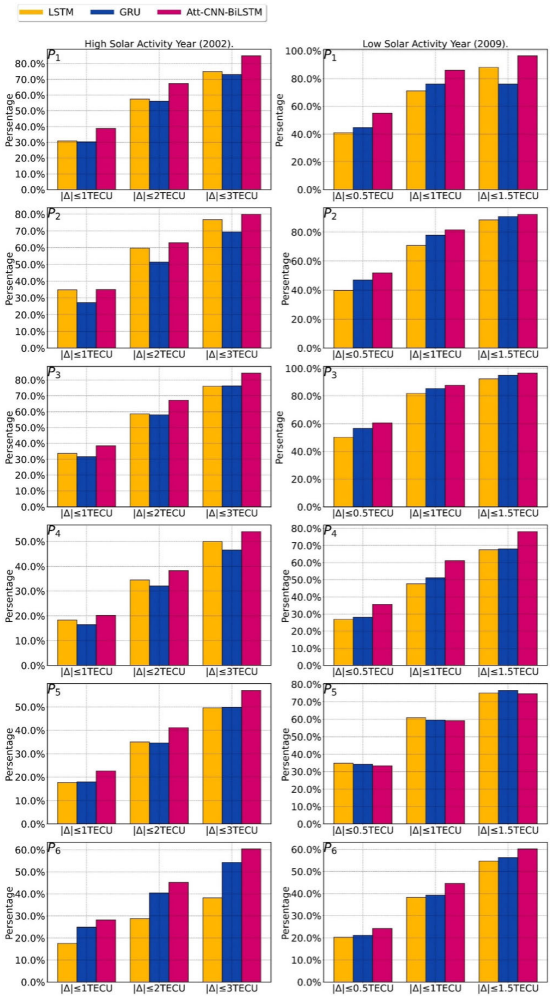


FIGURE 12. Absolute error distribution statistics: Left for high solar activity year; Right for low solar activity year.

are obtained by grid search. The final values of the hyper-parameters are shown in Table 5.

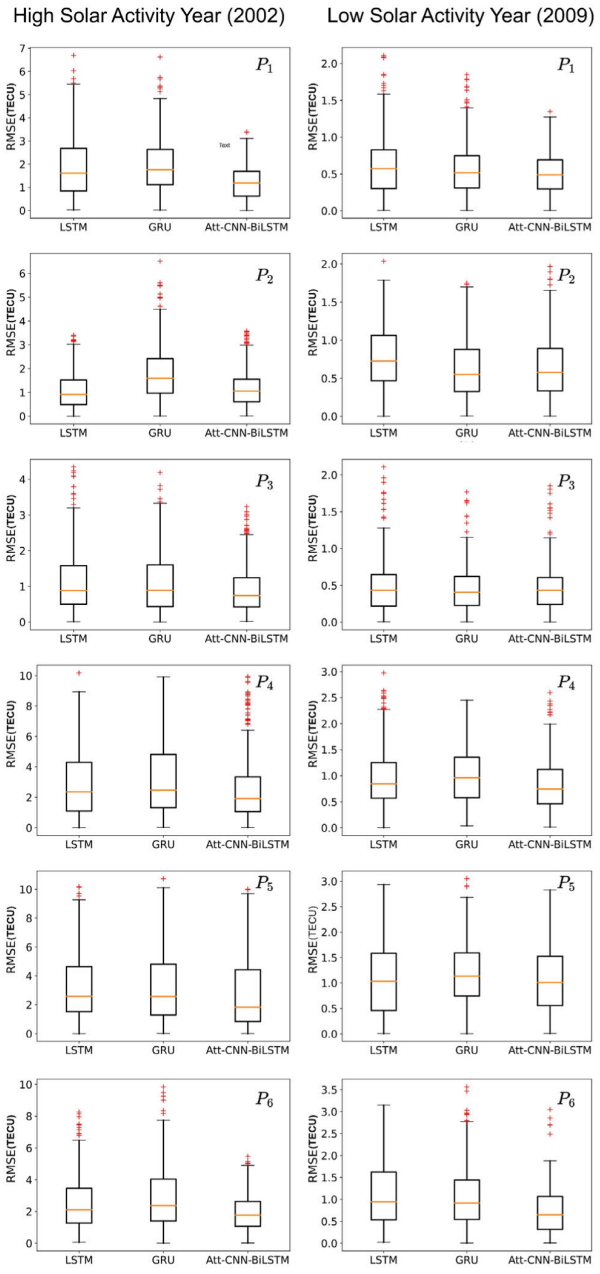


FIGURE 13. Box plot comparisons of daily RMSE statistical results.

C. THE IMPACT OF DIFFERENT INPUT LENGTHS ON PREDICTION PERFORMANCE

Figure 10 shows the impact of different model lengths (Length) on the RMSE and R^2. Each row represents a different prediction point (P_1 to P_6). As can be seen from the figure, the prediction effect is best when the input length is 7 days. Therefore, this article uses an input length of 7 days for prediction.

D. ABLATION EXPERIMENT

To verify the effectiveness of attention and CNN in the TEC prediction model, ablation experiments were conducted. The results are shown in Table 6. It can be seen that, whether

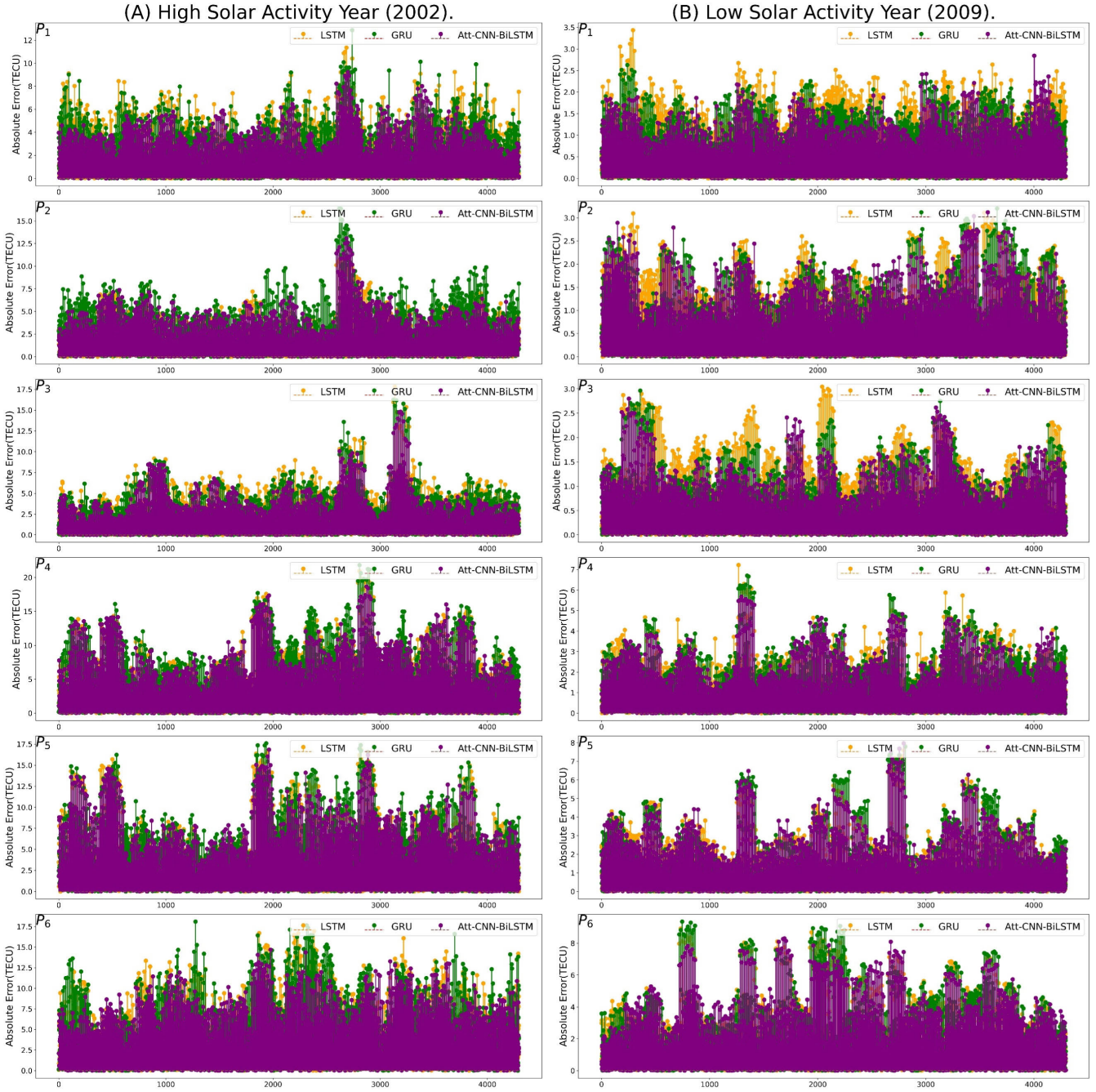


FIGURE 14. Absolute error comparisons of 3 models at 6 locations.

in high or low solar activity years, adding only CNN did not improve the predictive performance of BiLSTM, but adding both CNN and attention significantly improved the model's predictive performance. For example, in high solar activity year (2002), compared with BiLSTM, the RMSE of Att-CNN-BiLSTM decreased by 14.15%, 17.04%, 14.79% 12.38%, 15.93% and 13.73% in P_1-P_6, respectively. In low solar activity year (2009), compared with BiLSTM, the RMSE of Att-CNN-BiLSTM decreased by 19.52%, 14.78%, 16.16%, 5.21%, 2.93% and 11.15% in P_1-P_6, respectively.

E. COMPARISON WITH OTHER TEC PREDICTION MODELS

In this section, Att-CNN-BiLSTM was compared with LSTM and GRU in the high and low solar activity years, respectively. Experimental results are shown in Table 7, from which it can be seen that the mean RMSE of Att-CNN-BiLSTM in six regions is 16.92% lower than GRU's and 10.28% lower than LSTM's in high solar activity year. In low solar activity year, the mean RMSE of Att-CNN-BiLSTM is 11.82% lower than LSTM's and 8.92% lower than GRU's. The R^2 s of Att-CNN-BiLSTM in six locations are all higher than those of LSTM

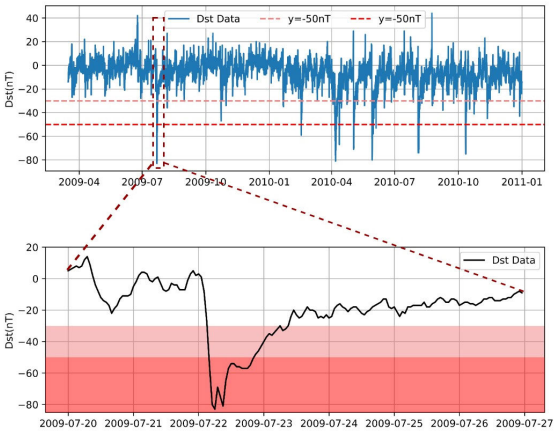


FIGURE 15. Geomagnetic storm Dst index data.

and GRU, indicating that Att-CNN-BiLSTM has a higher prediction accuracy than the comparison models.

In addition, this paper compared the number of parameters, time consumed per iteration, and memory usage of each model. Results are shown in Table 8. It can be seen that the number of parameters, time required for each iteration, and memory usage of Att-CNN-BiLSTM are much higher than those of the comparative models. This indicates that the performance improvement of Att-CNN-BiLSTM comes at the cost of sacrificing computational time and **memory usage**, which is the limitation of our model in this paper.

This paper further analyzed the impact of TEC mean values on the predictive performance of each model, as shown in Figure 11. It can be seen that at P_4, P_5, P_6, the TEC mean is relatively high, and the RMSE of each model in these three regions is also relatively large. In the other three regions, the TEC mean is relatively low, and the RMSE of each model is also relatively small, indicating that RMSE is greatly influenced by the mean TEC of a location. From Figure 11, it can also be seen that R^2 is less affected by the TEC mean and is relatively stable.

To quantitatively analyze the predictive performance of each model more accurately, the distribution percentages of the absolute error $|\Delta|$ ($|\Delta| = |y_i - \hat{y}_i|$) between the predicted and true values of various models in 2002 and 2009 were further compared, as shown in Figure 12. It can be seen that at all positions, in high solar activity year (2002), the proportions of $|\Delta| \leq 1TECU, |\Delta| \leq 2TECU, |\Delta| \leq 3TECU$ of the model proposed in this paper are obviously higher than those of other comparative models. In low solar activity year (2009), the percentages of $|\Delta| \leq 0.5TECU, |\Delta| \leq 1TECU$ *and* $|\Delta| \leq 1.5TECU$ of Att-CNN-BiLSTM outperform LSTM and GRU except for P_5.

Furthermore, for a more detailed comparison, the daily RMSE of three models in each region was calculated, and their mean, median, and minimum values were summarized. Figure 13 shows the box plots of daily RMSE statistical results of each model. From Figure 13, it can be observed that the median error of Att-CNN-BiLSTM model is the

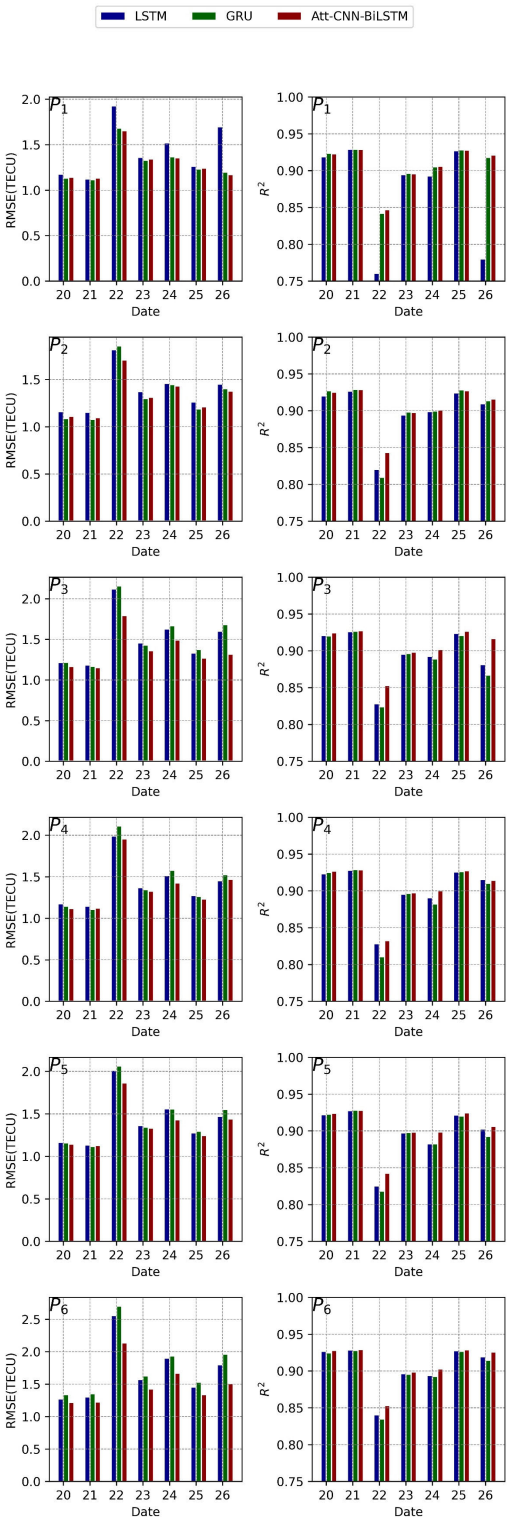


FIGURE 16. Comparison results of different models in magnetic storm time (from July 20, 2009 to July 26, 2009).

lowest at all six locations in high solar activity year. In low solar activity year, at P_2 and P_3, Att-CNN-BiLSTM model's median error is slightly higher than GRU, at other 4 locations, Att-CNN-BiLSTM model's median errors are the smallest

in 3 models. Compared to LSTM and GRU, Att-CNN-BiLSTM model has a more concentrated RMSE distribution. These indicate that Att-CNN-BiLSTM proposed model outperforms LSTM and GRU in the vast majority of cases.

Furthermore, the absolute prediction errors of the three models are presented in Figure 14. It can be seen that in the vast majority of cases, the absolute error of Att-CNN-BiLSTM model is significantly lower than those of the comparative models.

Based on the above comparison results, it can be concluded that the prediction performance of the model proposed in this paper is significantly superior to LSTM and GRU.

F. IMPACT OF MAGNETIC STORMS ON PREDICTION PERFORMANCE

To further verify the predictive performance of the proposed model in extreme situations, the week from July 20, 2009, to July 26, 2009, was selected according to Dst as magnetic storm data. Figure 15 shows the Dst curve between 2009 and 2011. It can be seen that there are significant fluctuations in the Dst curve from July 20, 2009 to July 26, 2009, indicating the occurrence of a magnetic explosion during that period. Figure 16 shows the prediction results of the model proposed in this paper and the comparison models during this storm period.

As can be seen from Figure 16, after the magnetic storm occurred, the prediction effect of Att-CNN-BiLSTM model was far better than other models. On the day of the geomagnetic storm (July 22, 2009), the overall RMSE of each model in the six regions were 0.5876, 0.6049, 0.3790, and $\boldsymbol{R^2}$ were 0.9531, 0.9595, 0.9800 respectively. Compared with the comparative model, the RMSE of Att-CNN-BiLSTM model was respectively It decreased by 35.50% and 37.35%, and $\boldsymbol{R^2}$ increased by 2.74% and 2.09% respectively. This shows that Att-CNN-BiLSTM model predicts best when magnetic storms occur at various selected points.

V. CONCLUSION

In this paper, we propose a TEC prediction model Att-CNN-BiLSTM that can simultaneously extract time series features and local position features and can adaptively weight the features. We selected 6 locations in China and used 7 consecutive days of TEC and Dst data as inputs to predict TEC for the next 24 hours. We first conducted ablation experiments, and the results showed that adding CNN and attention simultaneously can significantly improve the predictive performance of the model. We then compared our model with LSTM and GRU in high and low solar activity year. Results showed that the mean RMSE of our model in six regions is 16.92% lower than GRU's and 10.28% lower than LSTM's in high solar activity year. In low solar activity year, the mean RMSE of our model is 11.82% lower than LSTM's and 8.92% lower than GRU's.

The research work in this paper shows that combining temporal features with local position features and adding attention mechanisms to adaptively weight features can significantly improve TEC prediction performance. However, the model proposed in this paper also has shortcomings. For example, the boxplot of Att-CNN-BiLSTM shows a wider interquartile range, indicating that its prediction results are more volatile, especially in 2002 (high solar activity). While in 2009 (low solar activity year), the RMSE distribution of Att-CNN-BiLSTM is relatively more concentrated and consistent, but its performance is slightly worse than the GRU model on such as P_3 and P_6, which shows that under different solar activity conditions, its robustness may have room for improvement.

In this paper, our preprocessing of TEC and Dst data only includes differentiation. In the future, we will do more preprocessing on the data, such as a new empirical mode decomposition method based on Akima spline interpolation technology [56], to further improve the performance of the model. The model in this paper sacrifices memory footprint for better prediction performance, and in the future, we will investigate lightweight models that reduce computational requirements without compromising accuracy. And, a wider range will be selected. In the future, we will try to study the performance of the model in the real world.

REFERENCES

[1] C. Shi, S. Gu, Y. Lou, and M. Ge, "An improved approach to model ionospheric delays for single-frequency precise point positioning," *Adv. Space Res.*, vol. 49, no. 12, pp. 1698–1708, Jun. 2012, doi: 10.1016/j.asr.2012.03.016.

[2] E. P. Macalalad, L.-C. Tsai, J. Wu, and C.-H. Liu, "Application of the Taiwan ionospheric model to single-frequency ionospheric delay corrections for GPS positioning," *GPS Solutions*, vol. 17, no. 3, pp. 337–346, Jul. 2013.

[3] O. Øvstedal, "Absolute positioning with single-frequency GPS receivers," *GPS Solutions*, vol. 5, no. 4, pp. 33–44, Apr. 2002, doi: 10.1007/pl00012910.

[4] G. Sharma, P. K. Champati Ray, S. Mohanty, and S. Kannaujiya, "Ionospheric TEC modelling for earthquakes precursors from GNSS data," *Quaternary Int.*, vol. 462, pp. 65–74, Dec. 2017, doi: 10.1016/j.quaint.2017.05.007.

[5] T. Liu, B. Zhang, Y. Yuan, and M. Li, "Real-time precise point positioning (RTPPP) with raw observations and its application in real-time regional ionospheric VTEC modeling," *J. Geodesy*, vol. 92, no. 11, pp. 1267–1283, Nov. 2018, doi: 10.1007/s00190-018-1118-2.

[6] L. Li, H. Liu, H. Le, J. Yuan, W. Shan, Y. Han, G. Yuan, C. Cui, and J. Wang, "Spatiotemporal prediction of ionospheric total electron content based on ED-ConvLSTM," *Remote Sens.*, vol. 15, no. 12, p. 3064, Jun. 2023, doi: 10.3390/rs15123064.

[7] J. Tang, S. Zhang, X. Huo, and X. Wu, "Ionospheric assimilation of GNSS TEC into IRI model using a local ensemble Kalman filter," *Remote Sens.*, vol. 14, no. 14, p. 3267, Jul. 2022, doi: 10.3390/rs14143267.

[8] J. Qiao, Y. Liu, Z. Fan, Q. Tang, X. Li, F. Zhang, Y. Song, F. He, C. Zhou, H. Qing, and Z. Li, "Ionospheric TEC data assimilation based on Gauss–Markov Kalman filter," *Adv. Space Res.*, vol. 68, no. 10, pp. 4189–4204, Nov. 2021, doi: 10.1016/j.asr.2021.08.004.

[9] Z. Huang, Q. B. Li, and H. Yuan, "Forecasting of ionospheric vertical TEC 1-h ahead using a genetic algorithm and neural network," *Adv. Space Res.*, vol. 55, no. 7, pp. 1775–1783, Apr. 2015, doi: 10.1016/j.asr.2015.01.026.

[10] D. Bilitza, "International reference ionosphere: Recent developments," *Radio Sci.*, vol. 21, no. 3, pp. 343–346, May 1986, doi: 10.1029/RS021i003p00343.

[11] D. Bilitza, "International reference ionosphere 2000," *Radio Sci.*, vol. 36, no. 2, pp. 261–275, Mar. 2001, doi: 10.1029/2000RS002432.

[12] D. Bilitza, D. Altadill, V. Truhlik, V. Shubin, I. Galkin, B. Reinisch, and X. Huang, "International reference ionosphere 2016: From ionospheric climate to real-time weather predictions," *Space Weather*, vol. 15, no. 2, pp. 418–429, Feb. 2017, doi: 10.1002/2016sw001593.

[13] A. B. Makar, K. E. Mcmartin, M. Palese, and T. R. Tephly, "Formate assay in body fluids: Application in methanol poisoning," *Biochem. Med.*, vol. 13, no. 2, pp. 117–126, Jun. 1975.

[14] C. M. Ho, B. D. Wilson, A. J. Mannucci, U. J. Lindqwister, and D. N. Yuan, "A comparative study of ionospheric total electron content measurements using global ionospheric maps of GPS, TOPEX radar, and the bent model," *Radio Sci.*, vol. 32, no. 4, pp. 1499–1512, Jul. 1997, doi: 10.1029/97RS00580.

[15] G. Hochegger, B. Nava, S. Radicella, and R. Leitinger, "A family of ionospheric models for different uses," *Phys. Chem. Earth, Part C, Sol., Terr. Planet. Sci.*, vol. 25, no. 4, pp. 307–310, Jan. 2000.

[16] B. Nava, P. Coïsson, and S. M. Radicella, "A new version of the NeQuick ionosphere electron density model," *J. Atmos. Solar-Terrestrial Phys.*, vol. 70, no. 15, pp. 1856–1862, Dec. 2008, doi: 10.1016/j.jastp.2008.01.015.

[17] J. A. Klobuchar, "Ionospheric time-delay algorithm for single-frequency GPS users," *IEEE Trans. Aerosp. Electron. Syst.*, vols. AES-23, no. 3, pp. 325–331, May 1987, doi: 10.1109/TAES.1987.310829.

[18] N. Wang, Y. Yuan, Z. Li, and X. Huo, "Improvement of Klobuchar model for GNSS single-frequency ionospheric delay corrections," *Adv. Space Res.*, vol. 57, no. 7, pp. 1555–1569, Apr. 2016, doi: 10.1016/j.asr.2016.01.010.

[19] E. R. Ørskov and C. Fraser, "The effects of processing of barley-based supplements on rumen pH, rate of digestion and voluntary intake of dried grass in sheep," *Brit. J. Nutrition*, vol. 34, no. 3, pp. 493–500, Nov. 1975, doi: 10.1017/s0007114575000530.

[20] X. Lin, H. Wang, Q. Zhang, C. Yao, C. Chen, L. Cheng, and Z. Li, "A spatiotemporal network model for global ionospheric TEC forecasting," *Remote Sens.*, vol. 14, no. 7, p. 1717, Apr. 2022, doi: 10.3390/rs14071717.

[21] P. Muhtarov and I. Kutiev, "Autocorrelation method for temporal interpolation and short-term prediction of ionospheric data," *Radio Sci.*, vol. 34, no. 2, pp. 459–464, Mar. 1999, doi: 10.1029/1998RS900020.

[22] A. Krankowski, W. Kosek, L. W. Baran, and W. Popinski, "Wavelet analysis and forecasting of VTEC obtained with GPS observations over European latitudes," *J. Atmos. Solar-Terrestrial Phys.*, vol. 67, no. 12, pp. 1147–1156, Aug. 2005, doi: 10.1016/j.jastp.2005.03.004.

[23] D. V. Ratnam, Y. Otsuka, G. Sivavaraprasad, and J. R. K. K. Dabbakuti, "Development of multivariate ionospheric TEC forecasting algorithm using linear time series model and ARMA over low-latitude GNSS station," *Adv. Space Res.*, vol. 63, no. 9, pp. 2848–2856, May 2019, doi: 10.1016/j.asr.2018.03.024.

[24] R. K. Vankadara, S. Sasmal, A. K. Maurya, and S. K. Panda, "An autoregressive integrated moving average (ARIMA) based forecasting of ionospheric total electron content at a low latitude Indian location," in *Proc. URSI Regional Conf. Radio Sci. (USRI-RCRS)*. India: IEEE, Dec. 2022, pp. 1–4, doi: 10.23919/URSI-RCRS56822.2022.10118532.

[25] M. Saqib, E. şentürk, S. A. Sahu, and M. A. Adil, "Ionospheric anomalies detection using autoregressive integrated moving average (ARIMA) model as an earthquake precursor," *Acta Geophysica*, vol. 69, no. 4, pp. 1493–1507, Aug. 2021, doi: 10.1007/s11600-021-00616-3.

[26] M. Kaselimi, A. Voulodimos, N. Doulamis, A. Doulamis, and D. Delikaraoglou, "Deep recurrent neural networks for ionospheric variations estimation using GNSS measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022, Art. no. 5800715, doi: 10.1109/TGRS.2021.3090856.

[27] K. Watthanasangmechai, P. Supnithi, S. Lerkvaranyu, T. Tsugawa, T. Nagatsuma, and T. Maruyama, "TEC prediction with neural network for equatorial latitude station in Thailand," *Earth, Planets Space*, vol. 64, no. 6, pp. 473–483, Jun. 2012.

[28] S. Sahu, R. Trivedi, R. K. Choudhary, A. Jain, and S. Jain, "Prediction of total electron content (TEC) using neural network over anomaly crest region bhopal," *Adv. Space Res.*, vol. 68, no. 7, pp. 2919–2929, Oct. 2021, doi: 10.1016/j.asr.2021.05.027.

[29] I. L. Mallika, D. V. Ratnam, Y. Ostuka, G. Sivavaraprasad, and S. Raman, "Implementation of hybrid ionospheric TEC forecasting algorithm using PCA-NN method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 371–381, Jan. 2019, doi: 10.1109/JSTARS.2018.2877445.

[30] J. B. Habarulema, L.-A. McKinnell, P. J. Cilliers, and B. D. L. Opperman, "Application of neural networks to south African GPS TEC modelling," *Adv. Space Res.*, vol. 43, no. 11, pp. 1711–1720, Jun. 2009, doi: 10.1016/j.asr.2008.08.020.

[31] C. Cesaroni, L. Spogli, A. Aragon-Angel, M. Fiocca, V. Dear, G. De Franceschi, and V. Romano, "Neural network based model for global total electron content forecasting," *J. Space Weather Space Climate*, vol. 10, p. 11, 2020, doi: 10.1051/swsc/2020013.

[32] J. R. K. K. Dabbakuti, "Application of singular spectrum analysis using artificial neural networks in TEC predictions for ionospheric space weather," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5101–5107, Dec. 2019, doi: 10.1109/JSTARS.2019.2956968.

[33] S. Inyurt and A. Sekertekin, "Modeling and predicting seasonal ionospheric variations in Turkey using artificial neural network (ANN)," *Astrophys. Space Sci.*, vol. 364, no. 4, p. 62, Apr. 2019, doi: 10.1007/s10509-019-3545-9.

[34] Z. Huang and H. Yuan, "Ionospheric single-station TEC short-term forecast using RBF neural network," *Radio Sci.*, vol. 49, no. 4, pp. 283–292, Apr. 2014, doi: 10.1002/2013rs005247.

[35] J. B. Habarulema, L.-A. McKinnell, and P. J. Cilliers, "Prediction of global positioning system total electron content using neural networks over south Africa," *J. Atmos. Solar-Terrestrial Phys.*, vol. 69, no. 15, pp. 1842–1850, Nov. 2007, doi: 10.1016/j.jastp.2007.09.002.

[36] T. Yuan, Y. Chen, S. Liu, and J. Gong, "Prediction model for ionospheric total electron content based on deep learning recurrent neural networkormalsize," *Chin. J. Space Sci.*, vol. 38, no. 1, p. 48, 2018, doi: 10.11728/cjss2018.01.048.

[37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[38] M. Kaselimi, A. Voulodimos, N. Doulamis, A. Doulamis, and D. Delikaraoglou, "A causal long short-term memory sequence to sequence model for TEC prediction using GNSS observations," *Remote Sens.*, vol. 12, no. 9, p. 1354, Apr. 2020, doi: 10.3390/rs12091354.

[39] L. Liu, S. Zou, Y. Yao, and Z. Wang, "Forecasting global ionospheric TEC using deep learning approach," *Space Weather*, vol. 18, no. 11, pp. 1–12, Nov. 2020, doi: 10.1029/2020sw002501.

[40] K. D. Reddybattula, L. S. Nelapudi, M. Moses, V. R. Devanaboyina, M. A. Ali, P. Jamjareegulgarn, and S. K. Panda, "Ionospheric TEC forecasting over an Indian low latitude location using long short-term memory (LSTM) deep learning network," *Universe*, vol. 8, no. 11, p. 562, Oct. 2022, doi: 10.3390/universe8110562.

[41] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.

[42] K. Sivakrishna, D. Venkata Ratnam, and G. Sivavaraprasad, "A bidirectional deep-learning algorithm to forecast regional ionospheric TEC maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4531–4543, 2022, doi: 10.1109/JSTARS.2022.3180940.

[43] A. Ruwali, A. J. S. Kumar, K. B. Prakash, G. Sivavaraprasad, and D. V. Ratnam, "Implementation of hybrid deep learning model (LSTM-CNN) for ionospheric TEC forecasting using GPS data," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1004–1008, Jun. 2021, doi: 10.1109/LGRS.2020.2992633.

[44] J. Tang, Y. Li, M. Ding, H. Liu, D. Yang, and X. Wu, "An ionospheric TEC forecasting model based on a CNN-LSTM-attention mechanism neural network," *Remote Sens.*, vol. 14, no. 10, p. 2433, May 2022, doi: 10.3390/rs14102433.

[45] N. K. Ahmed, A. F. Atiya, N. E. Gayar, and H. El-Shishiny, "An empirical comparison of machine learning models for time series forecasting," *Econ. Rev.*, vol. 29, nos. 5–6, pp. 594–621, Aug. 2010, doi: 10.1080/07474938.2010.481556.

[46] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Stat. Assoc.*, vol. 74, no. 366a, pp. 427–431, Jun. 1979, doi: 10.1080/01621459.1979.10482531.

[47] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, Aug. 1978, doi: 10.1093/biomet/65.2.297.

[48] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

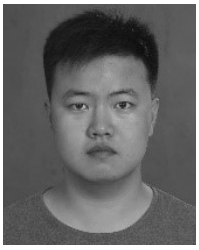
[49] E. Akleman, "Deep learning," *Computer*, vol. 53, no. 9, pp. 1–17, Sep. 2020, doi: 10.1109/MC.2020.3004171.

[50] B. SravyaPranati, D. Suma, C. ManjuLatha, and S. Putheti, "Large-scale video classification with convolutional neural networks," in *Information and Communication Technology for Intelligent Systems*. Cham, Switzerland: Springer, 2021, pp. 689–695.

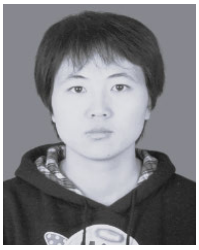
[51] B. Chakravarthi, S.-C. Ng, M. R. Ezilarasan, and M.-F. Leung, "EEG-based emotion recognition using hybrid CNN and LSTM classification," *Frontiers Comput. Neurosci.*, vol. 16, Oct. 2022, Art. no. 1019776, doi: 10.3389/fncom.2022.1019776.
[52] A. Severyn and A. Moschitti, "Learning to rank short text pairs with convolutional deep neural networks," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.* Chile: ACM, Aug. 2015, pp. 373–382, doi: 10.1145/2766462.2767738.
[53] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2412–2424, Jun. 2021, doi: 10.1109/TKDE.2019.2954510.
[54] M. Ali, D. M. Khan, H. M. Alshanbari, and A. A.-A.-H. El-Bagoury, "Prediction of complex stock market data using an improved hybrid EMD-LSTM model," *Appl. Sci.*, vol. 13, no. 3, p. 1429, Jan. 2023, doi: 10.3390/app13031429.
[55] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
[56] M. Ali, D. M. Khan, I. Saeed, and H. M. Alshanbari, "A new approach to empirical mode decomposition based on Akima spline interpolation technique," *IEEE Access*, vol. 11, pp. 67370–67384, 2023, doi: 10.1109/ACCESS.2023.3253279.

HAIJUN LIU received the Ph.D. degree from China University of Geosciences (Beijing).

She is currently an Associate Professor. She completed postdoctoral research with Nanjing University. Her main research areas are image recognition and data mining. She has presided more than three projects of Hebei Provincial Department of Education, one project of China Seismological Administration Teacher Fund, and three projects of the central universities. She has published more than ten articles as the first author and corresponding author. Her training directions include disaster information processing technology, disaster risk assessment, and disaster prevention and reduction planning. Her research interests include machine learning, artificial intelligence, image recognition, and intelligent emergency technology.



HAORAN WANG was born in 1999. He received the bachelor's degree from Shanghai Normal University. He is currently pursuing the degree with the Graduate School, Institute of Disaster Prevention. His main research interests include deep learning and time series forecasting.



JING YUAN graduated from Tsinghua University, under the supervision of Professor Zhang Yujin, an expert in computer vision, is mainly engaged in the research of computer vision and machine learning related aspects. She has guided students to participate in the American College Mathematics Competition and won the second prize once and the third prize twice; she has guided students to participate in the Tianchi Big Data Competition and won the second prize twice and the third prize twice. She has published a total of 20 SCI papers, EI papers and core papers in recent years; presided over and participated in, one major special project of the Ministry of Science and Technology, one provincial project, one Earthquake Spark Plan project, and one Natural Science Foundation project.

LIANGCHAO LI is currently pursuing the degree with the Graduate School, Institute of Disaster Prevention. His main research interests include deep learning and time series forecasting.

LILI ZHANG received the Ph.D. degree in 2011. She is currently an Associate Professor with the Institute of Disaster Prevention. Her research interests include hydrogeology engineering geology and disaster information processing technology.