GLER-BiGRUnet: A Surface Deformation Prediction Model Fusing Multiscale Features of InSAR Deformation Information and Environmental Factors

Tianbao Huo, Yi He, *Member, IEEE*, Lifeng Zhang, Wang Yang, Jiapeng Tang, Qing Zhang, Jiangang Lu, and Yunhao Zhang

***Abstract*—Accurate surface deformation (SD) predictions are critical for early warning and timely remediation of infrastructure damage. However, the current SD prediction models do not integrate the multiscale features of InSAR SD and environmental factors (EFs), which make their prediction results inaccurate. To address these limitations, we proposed a bidirectional gated recurrent unit (BiGRU) multioutput SD prediction network (GLER-BiGRUnet), which mainly included global–local feature extraction (GLFE), multifactor cross-attention residual (MCAR), and local residual module embedded in self-attention mechanism (RCSA) modules. Specifically, dense and one-dimensional convolutional layers were concatenated in the GLFE module to extract global–local SD features. The long time-series dependence between EFs and SD was learned in the MCAR module using the multihead cross-attention mechanism to obtain the corresponding attention weight feature matrix. The residual connection and self-attention mechanisms were used in the RCSA module to merge the multiscale features and enhance the model fitting ability. We chose four typical regions in the permafrost area of Qinghai–Tibet Railway as the scene for the experiment. The spatial distribution and local profile exhibited relatively small discrepancies between the prediction results of the GLER-BiGRUnet model and the InSAR SD. Meanwhile, the average root-mean-square error of the GLER-BiGRUnet model in the four typical regions was 0.19 mm, and the proposed model had the best evaluation index compared with other SD prediction models. Additionally, the prediction trend of SD of the proposed GLER-BiGRUnet model was consistent with the original InSAR SD, and the prediction results were more stable than those of the other prediction models. The SD prediction model proposed in this article contributes to early warning of SD.**

***Index Terms*—Deformation prediction, GLER-BiGRUnet, multiscale features, Qinghai–Tibet railway (QTR).**

Manuscript received 2 July 2024; revised 8 August 2024; accepted 12 August 2024. Date of publication 16 August 2024; date of current version 5 September 2024. This work was supported in part by the National Natural Scientific Foundation of China under Grant 42201459, in part by the Key and Development Project of Lanzhou Jiao Tong University under Grant LZJTU-ZDYF2301, in part by the Outstanding Youth Fund of Gansu Province under Grant 23JRRA881, and in part by the Science and Technology Project of Gansu Province under Grant 22JR5RA067. *(Corresponding author: Yi He.)*

The authors are with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China, also with the National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China, and also with the Gansu Provincial Key Laboratory of Science and Technology in Surveying and Mapping, Lanzhou 730070, China (e-mail: htb666@126.com; heyi@mail.lzjtu.cn; 119273207@qq.com; yyangwang48@gmail.com; 3153292919@qq.com; 1439112766@qq.com; rs_lujg@qq.com; 596532969@qq.com).

Digital Object Identifier 10.1109/JSTARS.2024.3443833

I. Introduction

As a common geological hazard [1], surface deformation (SD) can destabilize infrastructure [2], threaten people's lives and properties [3], [4], and impede the sustainable development of regional economies [5], [6]. It can be seen that it is of great significance to construct the accurate SD prediction model for early warning of damage caused by SD. Current research on SD prediction can be divided into two categories based on whether or not environmental factors (EFs) are considered. The first one is to construct the SD prediction model only based on GPS, leveling data, or multitemporal interferometric synthetic aperture radar SD monitoring data. For example, Yi et al. [2] constructed a long short-term memory (LSTM) [7] prediction model based on persistent scatterer interferometric synthetic aperture radar [8] SD monitoring data. The results showed that the LSTM had a better performance in short-term prediction. He et al. [4] proposed a CNN-PhLSTM model based on the SD monitoring results obtained by small baseline subset interferometric synthetic aperture radar (SBAS-InSAR) technology [9]. And the results showed that the CNN-PhLSTM prediction model outperforms the support vector regression and CNN-LSTM models in several evaluation indices, and its long time-series prediction results were more in line with the original InSAR SD trend. Peng et al. [10] used independent component analysis to isolate the independent deformation signals from the original InSAR sequences and constructed an LSTM prediction model. The prediction results in both deformation regions were more accurate than LSTM. However, the above studies did not consider the long time-series dependence between the SD monitoring data and EFs, making the prediction results lack credibility and validity. In addition, the SD prediction models constructed without considering the changes of EFs assumed that the EFs [normalized difference vegetation index (NDVI), land surface temperature (LST) and precipitation, etc.) and the geological conditions in the subsidence area were constant when predicting the SD [10], which leads to poor prediction accuracy.

The second type of SD prediction model takes into account the relationship between EFs and SD in the subsidence area. For example, Wang et al. [11] based on the self-attention (SA) mechanism in transformer [12] comprehensively considered the relationship between EFs and SD, and more accurately modeled nonseasonal and seasonal signals to achieve the prediction of SD

around Salt Lake; Yuan et al. [1] used the geographically and temporally weighted regression method to obtain the relative weights of EFs with SD monitoring data. Meanwhile, constructed an SD prediction model based on LSTM and attention mechanism, which can take into account the spatiotemporal nonstationary relationship, and obtained better predictions than the first type of prediction model. However, the above SD prediction models do not consider the multiscale characteristics of InSAR SD and EFs.

Thaw settlement [13], thaw slumping [14], and thermokarst lakes [15] on the Tibet Plateau in recent years have been impacting the safe operation and stability of Qinghai–Tibet Railway (QTR) and other infrastructures, attributed to permafrost degradation [16], [17], [18]. It is also threatening the level of food safety, healthcare, etc., of people along the QTR [19]. Therefore, we chose four typical regions in the permafrost area of QTR as the scene for the experiment.

We proposed a bidirectional gated recurrent unit (BiGRU) multioutput SD prediction network (GLER-BiGRUnet), which mainly included global–local feature extraction (GLFE), multifactor cross-attention residual (MCAR), and local residual module embedded in self-attention mechanism (RCSA) modules. GLFE with the dense and one-dimensional convolutional layer (Conv1D) modules was first introduced to extract the global–local information of SD and EFs. Then, the multihead cross-attention (MHCA) [20] mechanism was introduced to obtain the dependencies between SD and EFs. Meanwhile, the residual connection [21], BiGRU [22], and SA mechanism were introduced to learn multiscale, bidirectional, and critical SD and EFs' features. Finally, the SD trends of four typical regions in the permafrost area of QTR were predicted for the next two years. The objectives of this article are as follows:

1) to construct a GLER-BiGRUnet SD prediction model that can learn the multiscale features of InSAR SD and EFs in permafrost region;
2) to validate the feature learning ability and prediction performance of the GLER-BiGRUnet model by comparing the differences in spatial distributions and local details of the real and predicted data;
3) to predict the SD trend in the next two years of the four typical regions using the GLER-BiGRUnet model and compare the prediction accuracy and the stability of prediction results with other prediction models.

II. Materials

A. Time-Series InSAR Deformation Monitoring Data

As a means of Earth observation, InSAR technology is capable of inverting SD results with high efficiency and accuracy, and can provide reliable time-series data support for SD prediction missions. Therefore, we acquired 88 Sentinel-1A images (Frame: 475 and Path: 77) from the European Space Agency for the Salt Lake to Wuli segment from 2019 to 2022. The images were all in descending orbit, interferometric wide mode, VV polarization, and with incidence and azimuth angles of 34.69° and 90°, respectively. The InSAR deformation data of the Salt Lake to Wuli section [see Fig. 1(b)] were obtained using the SBAS-InSAR technique based on the environment for visualizing images, SARscape 5.6.2 platform. In addition, we used (1) to convert the results of deformation monitoring in the line-of-sight (LOS) direction d_{LOS} to those in the vertical direction d_v [23] and used them for subsequent work. Table I lists some important parameter settings and data information for SBAS-InSAR

$$d_v = \frac{d_{\mathrm{LOS}}}{\cos\theta} \tag{1}$$

where the θ is the incidence angle.

B. Experimental Scene

After analyzing the SD monitoring results and Google Earth images, four typical regions (A, B, C, and D) in the Salt Lake to Wuli segment were selected for SD prediction [see Fig. 1(c)–(f)]. At the same time, the four typical regions selected in this article have different geographical environment characteristics and the number of InSAR deformation characteristics, which can improve the generalization performance of the model on different deformation feature datasets. These typical regions were characterized by severe ground subsidence, large thermokarst lake, and traversed by the QTR. Among them, the lateral thermal erosion induced by thermokarst lake leads to the degradation of peripheral ice-rich permafrost [30], which further extends thermokarst distribution. However, the expansion of the thermokarst lake will increase soil temperature and lower the upper limit of the permafrost. This will trigger uneven ground subsidence on the land surface near the thermokarst lake, which, in turn, will lead to instability of the railway roadbed closer to the thermokarst lake. Additionally, the continuous operation of the QTR and the laying of oil pipelines and power transmission facilities in recent years [19] have disrupted the thermal state of the permafrost layer [31], which seriously threatens the QTRs safe operation [32].

C. Environmental Factors

In MHCA, the input of EFs for the time series is sufficient and necessary to obtain the response relationship between EFs and SD at each time step. Meanwhile, the changes of LST and NDVI as well as continuous precipitation events will affect the degree of permafrost degradation and reduce the stability of slopes [34], [35], [36], thus triggering the occurrence of ground subsidence. Therefore, we chose these three EFs as the input of the model. Detailed information on these EFs is presented in Table II. The pixel size of all EFs is resampled to 15 × 15 m.

D. Dataset Making

To fill in the missing Sentinel-1A images on 24 April 2022, we first extracted the raw InSAR SD and EFs images [see Fig. 2(a)] using the InSAR observations [see Fig. 2(b)]. Then, we performed time-series differencing of InSAR SD, LST, and NDVI [10], and monthly filling of precipitation. In this case, the interpolated InSAR SD data contain both surface uplift and subsidence data. The above operation made the EFs and InSAR SD strictly aligned at the time step so that the MHCA mechanism can obtain the dependence between the InSAR SD features and the EFs at each time step. The monthly padding is an operation

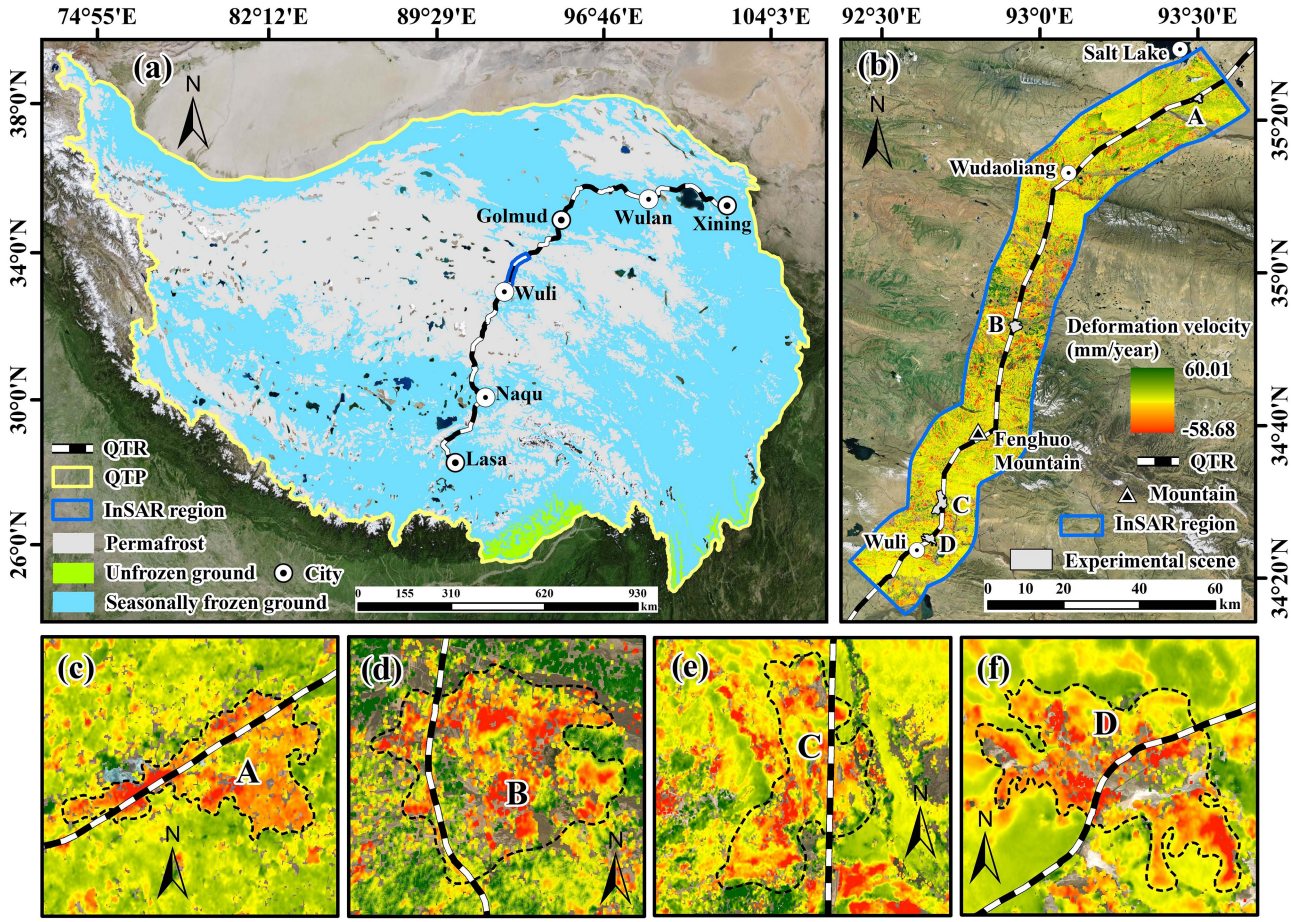

Fig. 1. Geographic location of four typical regions. (a) Data of permafrost distribution comes from Zou et al.'s article [33]. (b) Location of four typical regions. (c), (d), (e), and (f) are the distribution of SD velocity in regions A, B, C, and D, as well as the distribution of thermokarst lake and surface runoff in the region (the missing location of InSAR SD monitoring data in the figure), respectively (geographic map is ESRIWorldImage).

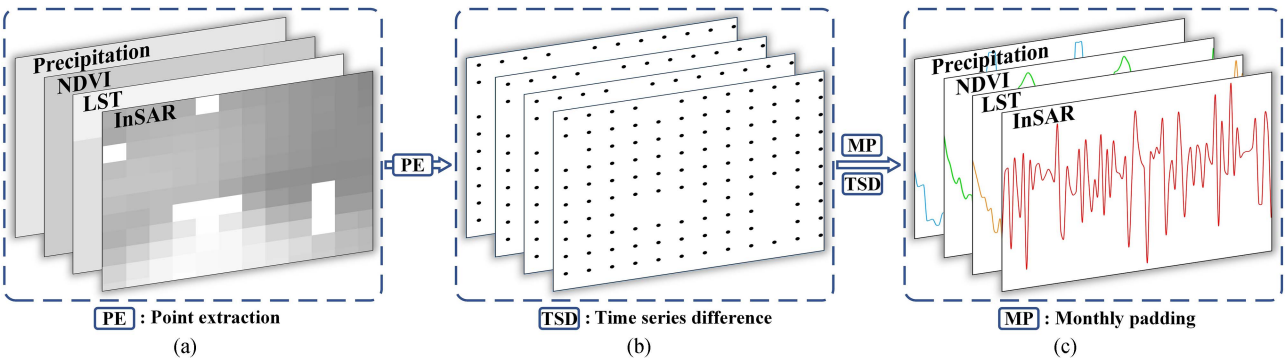

Fig. 2. Dataset production process.

that populates the monthly mean precipitation to the current month's InSAR time step. The time-series difference formula is given as follows:

$$\hat{Y}_i = Y_{i-1} + \frac{(Y_{i+1} - Y_{i-1})}{(T_{i+1} - T_{i-1})}(T_i - T_{i-1}) \tag{2}$$

where T_{i-1} and Y_{i-1} are the previous time point and the corresponding data, respectively, T_{i+1} and Y_{i+1} are the latter time point and the corresponding data, respectively, and T_i is the point in time where the interpolation is required.

Then, we used the standard scaling method to compress the data features of InSAR SD and EFs [see Fig. 2(c)] to have a mean and standard deviation of 0 and 1. Next, the last ten time-step data were taken as the test set. Then, the remaining data were segmented into compression results along the time-step dimension in a multistep to one-step overlapping time-series segmentation pattern [4], [37]. Finally, the remaining compression results were evenly sampled into training and validation set in 7:3.

TABLE I
EFs IN THE EXPERIMENTAL SCENE

Processing flow	Parameters and data	Property	Effect	Source
Connection graph generation	Max Normal Baseline (%) Max Temporal Baseline (days)	10 % 90 days	Generate the connection graph	/
Interferometric process	Multi-look	4：1	Enhance the signal-to-noise ratio [24]	/
	Coherence threshold	0.25	Mitigate the decorrelation	/
	Minimum cost flow	Unwrapping threshold: 0.25	Phase unwrapping	/
	Digital elevation model (DEM)	Shuttle Radar Probe Mission (SRTM, 30 m)	Orbit refining and flattening [25], [26]	http://earthexplorer.usgs.gov/
	Generic Atmospheric Correction Online Service	/	Atmospheric correction and obtain the interferogram [27]	http://www.gacos.net/
	Precise Orbit Ephemerides	Positioning accuracy > 5 m	Correct orbit information	https://s1qc.asf.alaska.edu/aux_poeorb/
SBAS-InSAR inversion	Goldstein filtering algorithm [28]	/	Remove noise phase [29]	/
	Space low-pass and time high-pass filtering	/	Remove the atmospheric phase [4]	/
Geocoding	DEM	SRTM (30 m)	Convert SAR coordinates into geographic coordinates	/
	Rate accuracy threshold	3	Screening of deformation rate results with high accuracy	/
	Elevation accuracy threshold	7	Exclusion of areas with low or unreliable accuracy of elevation data to ensure accuracy and reliability of estimates of deformation results	/

TABLE II
EFs IN THE EXPERIMENTAL SCENE

Name	Data sources	EFs	Scales/Resolution	Time range	Temporal resolution
NDVI	National Aeronautics and Space Administration (https://www.earthdata.nasa.gov/)	NDVI	MOD13Q1 (250 m)	2019.8.29 – 2022.8.13	16 days
LST	National Aeronautics and Space Administration (https://www.earthdata.nasa.gov/)	LST	MOD11A2 (1 km)	2019.9.6 – 2022.8.5	8 days
Precipitation	ERA5-Land (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means)	Precipitation	ERA5-Land (0.1°)	2019.9 – 2022.7	Monthly

III. METHODOLOGY

A. SD Prediction Model

In this article, we constructed a BiGRU multioutput SD prediction network (GLER-BiGRUnet). The network contained an innovative GLFE module, MCAR module, and local residual module embedded in a self-attention mechanism (RCSA). Moreover, the dimension raising module utilized the tandem structure of time-distributed dropout (TD-Dropout) and time-distributed density (TD-Dense) to suppress model overfitting and restore the dimension of RCSA module output results. The details of the main modules in GLER-BiGRUnet are described as follows.

1) GLFE Module: Choosing reasonable feature extraction methods is vital for enhancing the efficiency of model feature learning and enhancing the generalization ability during neural network training. In the SD time-series prediction task, existing studies generally directly pass the original feature dimensions through the dense layer for downscaling and global feature extraction [38], [39], [40]. However, the discrepancy between

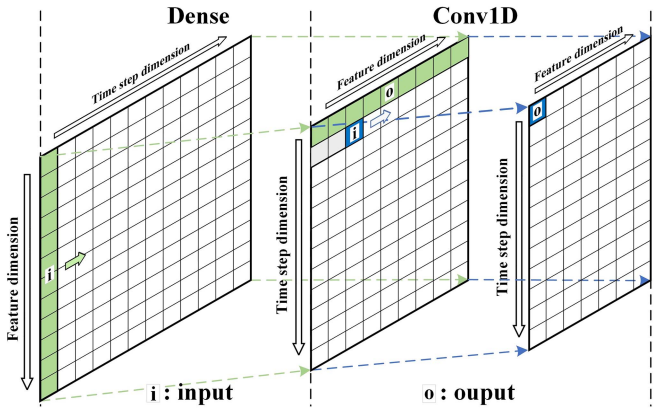

Fig. 3. Principles of feature extraction for dense and Conv1D layers.

the original feature dimension and the input dimension of the RNN layer is too large, which may result in the omission of local deformation features. This leads to insufficient model fitting and affects the prediction accuracy.

Therefore, a GLFE module was constructed by combining a dense-connected layer (Dense) and a Conv1D in this article (see Fig. 3). This module aims to capture both the global and local features of time-series InSAR SD and the EFs [41]. The activation functions used for InSAR SD and LST in the dense and Conv1D layers were TanH, while ReLU was applied for precipitation and NDVI. Additionally, when stacking Conv1D into two layers, the model tended to overfit. This could be because the two Conv1D layers extracted outlier features or noise, resulting in a decrease in the model's generalization ability on the validation set.

This module acts on the deformation feature dimension and the formulae for the module are shown in (3) and (4)

$$\text{output}_{\text{Dense}} = \text{activation}_{\text{Dense}}\ (\text{input}_{\text{Dense}} \cdot W + b) \tag{3}$$

where $\text{input}_{\text{Dense}}$ is the feature vector for the time-series InSAR data and EFs, W is the weighting matrix, and b is the bias vector

$$\text{GLFE} = \text{output}_{\text{Conv1D}}[i] = \text{activation}_{\text{Conv1D}}\left(\sum_{j=0}^{k-1} \text{output}_{\text{Dense}}[i+j] \cdot \text{kernel}[j] + b\right) \tag{4}$$

where i is the corresponding element of the output result of conv1D, j is the corresponding location of Conv1D, k is the size of Conv1D, kernel$[j]$ is the weight of the convolution kernel at position j, $\text{output}_{\text{Dense}}[i : i+k]$ is the window of $\text{output}_{\text{Conv1D}}[i]$, and b is the bias vector. In this article, k is set to 3.

2) MCAR Module:

$$\text{EF}_s = \text{Concat}\left(\text{GLFE}_{\text{LST}}, \text{GLFE}_{\text{NDVI}}, \text{GLFE}_{\text{Precipitation}}\right) \tag{5}$$

where GLFE_{LST}, $\text{GLFE}_{\text{NDVI}}$, and $\text{GLFE}_{\text{Precipitation}}$ represent the outputs of LST, NDVI, and precipitation through the GLFE module, respectively.

In Line 1 of the MCAR module, the SD and EFs data processed by the GLFE module were inputted together into MHCA (see Fig. 5) to learn the relationship between the SD and EFs across different heads of attention [see Fig. 4(c)]. The EFs processed by the GLFE module were input into BiGRU for modeling in Line 2. Among them, MHCA first maps the SDs and EFs into several different representation spaces [12] through the weight matrices (W_i^Q, W_i^K, and W_i^V) of the eight attention heads and obtains the corresponding query (Q_i), key (K_i), and value (V_i) vectors, which are computed as follows:

$$\begin{cases} Q_i = (\text{GLFE}_{\text{SD}}) \cdot W_i^Q \\ K_i = (\text{EF}_S) \cdot W_i^K \\ V_i = (\text{EF}_S) \cdot W_i^V. \end{cases} \tag{6}$$

Next, a highly optimized matrix multiplication based on dot-product cross attention was used for each attention head to calculate the cross-attention score between Q_i and K_i [12]. The cross-attention score was also scaled by the square root of the dimension d_K of the K_i vector to ensure the stability of the gradient [42]. The formula is calculated as follows:

$$\text{Scaled}_{A_i} = \frac{Q_i \cdot K_i^T}{\sqrt{d_K}}. \tag{7}$$

Subsequently, Scaled_{A_i} was converted to cross-attention weights by the softmax function [43], and the output of each attention head was obtained by weighted summation of V_i vectors. The formula is shown as follows:

$$\text{head}_{\boldsymbol{i}} = \text{softmax}\left(\text{Scaled}_{A_i}\right) \cdot V_i. \tag{8}$$

Then, the output of all attention heads was concatenated. And used the weight matrix W^0 to map the connection results, ensuring that the connection results had the same dimensions as the SD. Finally, the attention weight feature matrix (AWFM) ($\text{AWFM}_{\text{MHCA}}$) was obtained. The calculation formula is shown as follows:

$$\text{AWFM}_{\text{MHCA}} = \sum_{i=1}^{8} \text{Concat}\left(\text{head}_i\right) \cdot W^0. \tag{9}$$

Then, the output of the MHCA was passed through the TD-Dense layer and the TD-Dropout layer to prevent overfitting of the model (see Fig. 6). Simultaneously, the number of parameters input to the BiGRU layer was reduced to enhance the model's generalization ability. Subsequently, these results were fed into a three-layer BiGRU to further explore the time-series dependencies within the sequence of InSAR SD from both forward and backward directions [26], and to better extract the InSAR SD features. Finally, the outputs of the BiGRU in Line 1 and Line 2 were each passed through a TD-Dropout layer to suppress model overfitting.

In this article, the input of BiGRU in Line 1 was the AWFM output from the MHCA module, and Line 2 was the output of EFs in the GLFE module [see Fig. 4(c)]. With the forward–backward BiGRUs, the model's ability to interpret and express the attentional weights between different time steps can be further improved, and richer and more accurate feature representations can be extracted. Assuming that the sequence length of the input

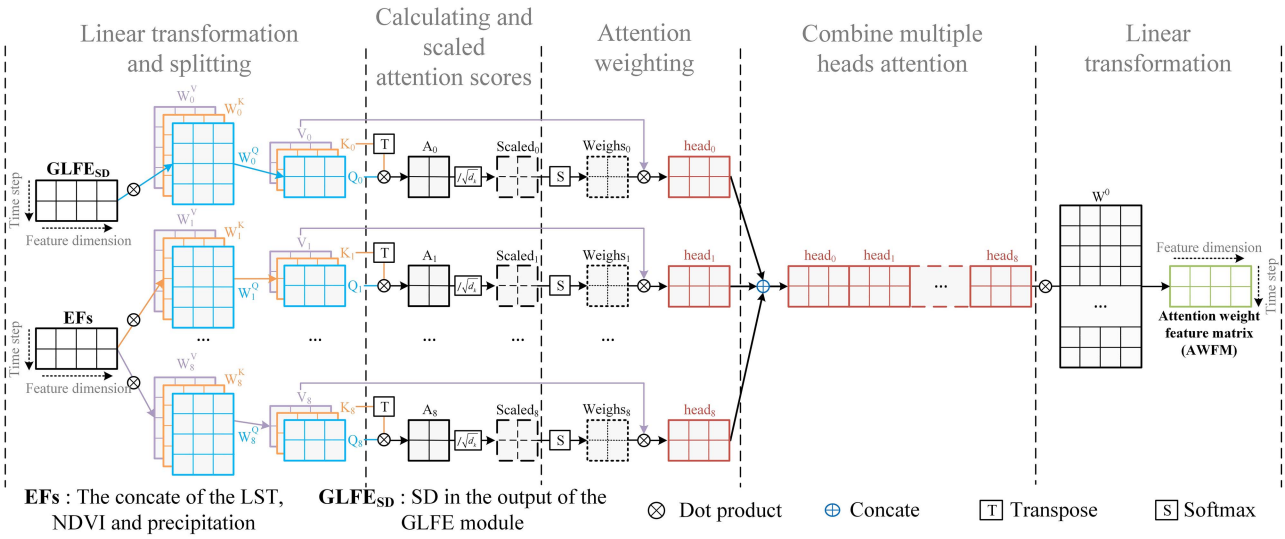

Fig. 5. Structure of the MHCA, and the number of its heads is 8.

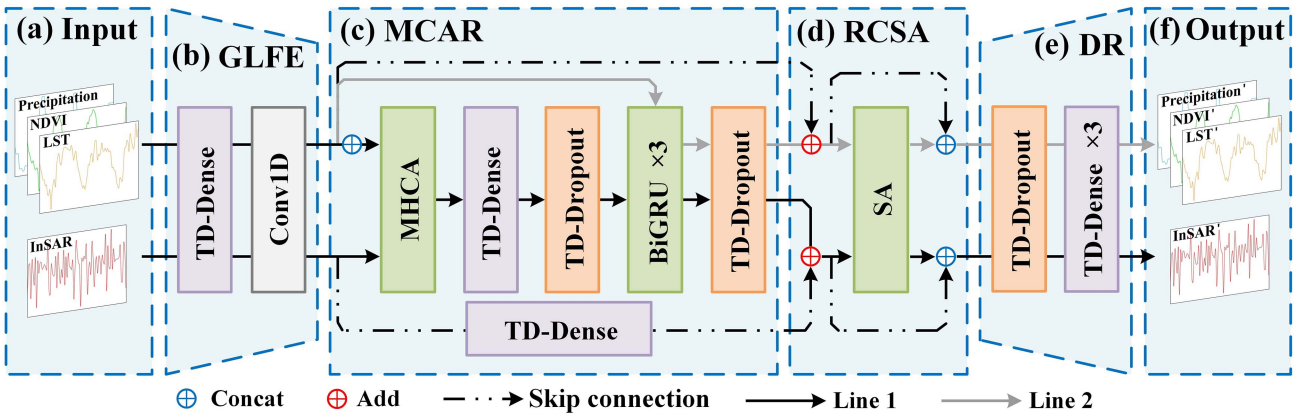

Fig. 4. Flowchart of SD prediction module.

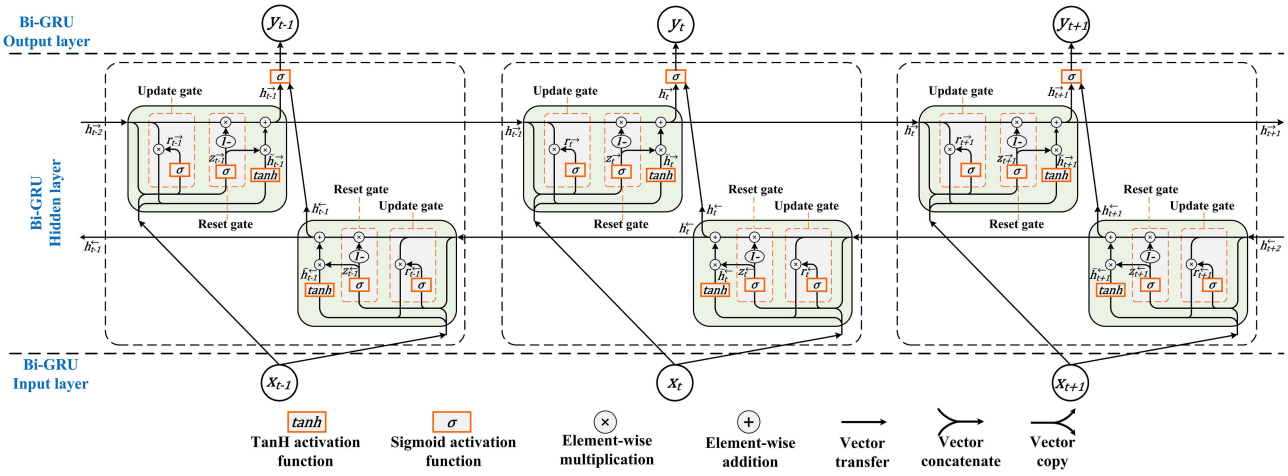

Fig. 6. Structure of the BiGRU.

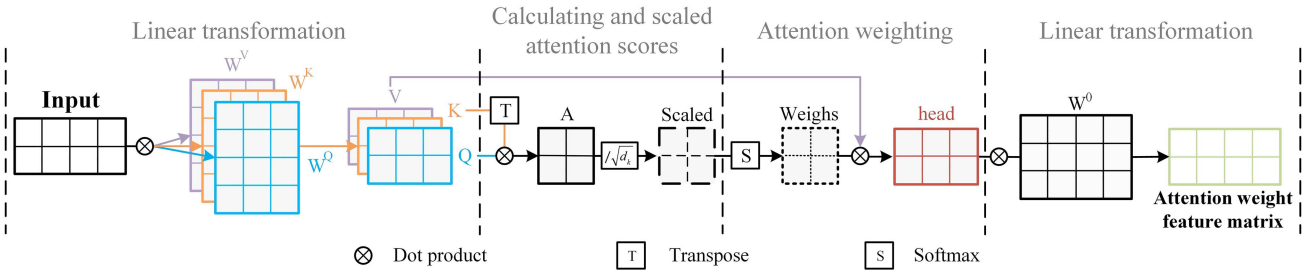

Fig. 7. Structure of the SA mechanism.

BiGRU is the time step T, the input and output of the current time step are x_t and y_t, the hidden states of the forward and backward are $h_t^{\rightarrow}$ and $h_t^{\leftarrow}$, and the formula of the BiGRU is given as follows:

Forward:

$$\begin{cases} z_t^{\rightarrow} = \sigma\left(W_z^{\rightarrow} x_t + U_z^{\rightarrow} h_{t-1}^{\rightarrow} + b_z^{\rightarrow}\right) \\ r_t^{\rightarrow} = \sigma\left(W_r^{\rightarrow} x_t + U_r^{\rightarrow} h_{t-1}^{\rightarrow} + b_r^{\rightarrow}\right) \\ \tilde{h}_t^{\rightarrow} = \tanh\left(W^{\rightarrow} x_t + U^{\rightarrow}\left(r_t^{\rightarrow} \odot h_{t-1}^{\rightarrow}\right) + b^{\rightarrow}\right) \\ h_t^{\rightarrow} = \left(1 - z_t^{\rightarrow}\right) \odot h_{t-1}^{\rightarrow} + z_t^{\rightarrow} \odot \tilde{h}_t^{\rightarrow}. \end{cases} \tag{10}$$

Backward:

$$\begin{cases} z_t^{\leftarrow} = \sigma\left(W_z^{\leftarrow} x_t + U_z^{\leftarrow} h_{t+1}^{\leftarrow} + b_z^{\leftarrow}\right) \\ r_t^{\leftarrow} = \sigma\left(W_r^{\leftarrow} x_t + U_r^{\leftarrow} h_{t+1}^{\leftarrow} + b_r^{\leftarrow}\right) \\ \tilde{h}_t^{\leftarrow} = \tanh\left(W^{\leftarrow} x_t + U^{\leftarrow}\left(r_t^{\leftarrow} \odot h_{t+1}^{\leftarrow}\right) + b^{\leftarrow}\right) \\ h_t^{\leftarrow} = \left(1 - z_t^{\leftarrow}\right) \odot h_{t+1}^{\leftarrow} + z_t^{\leftarrow} \odot \tilde{h}_t^{\leftarrow}. \end{cases} \tag{11}$$

Output of BiGRU:

$$y_t = \text{Concat}\left(h_t^{\rightarrow}, h_t^{\leftarrow}\right) \tag{12}$$

where $\rightarrow$ represents the forward GRU, $\leftarrow$ represents the backward GRU, z_t is the update gate, r_t is the reset gate, $\odot$ represents the elementwise multiplication, W_z, W_r, W and b_z, b_r, b are the weight matrix and bias vector of the current time step, and U_z, U_r, and U are the weight matrices of previous or next time step.

3) RSAM Module: In this module, the output of the GLFE module was input to the RCSA module together with the output of the BiGRU through residual addition [21] [see Fig. 4(d)]. This prevents the gradient from disappearing, preserves the long-range and multiscale SD features, and enhances the model fitting capability [44]. Subsequently, the addition results were fed into the SA mechanism (see Fig. 7), facilitating further feature extraction and association learning to enhance the accuracy and robustness of SD prediction. Finally, the outputs of the SA mechanism and the results of residual addition were concatenated via a constant mapping residual connection [see Fig. 4(d)] to enhance the feature sharing capability, learning efficiency, and parameter sharing of the model and reduce overfitting [45]. The formulae for the residual addition and SA mechanisms are as follows:

$$\text{Input}_{\text{SA}} = \text{Add}\left(\text{GLFE}, \text{MCAR}\right) \tag{13}$$

$$\begin{cases} Q = \left(\text{Input}_{\text{SA}}\right) \cdot W^Q \\ K = \left(\text{Input}_{\text{SA}}\right) \cdot W^K \\ V = \left(\text{Input}_{\text{SA}}\right) \cdot W^V \end{cases} \tag{14}$$

$$\text{SA}_{\text{weights}} = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_K}}\right) \tag{15}$$

$$\text{RSAM} = \text{Concat}\left(\text{Input}_{\text{SA}}, \text{SA}\right) \tag{16}$$

where W^Q, W^K, and W^V represent the weight matrices of Input_{SA}, and d_K is the dimension of the K vector.

B. Model Training and Evaluation Indicators

1) Loss Function and Metrics: The correct selection of the loss function is crucial for the training and performance of SD prediction models. Two loss functions commonly used in regression tasks are mean absolute error (MAE) and mean squared error (MSE) [40], [46]. However, MAE is detrimental to gradient descent and insensitive to outliers during optimization, and MSE is sensitive to outliers but may lead to unstable optimization [47]. In this article, although the SBAS-InSAR SD results have gone through a series of time-series differencing and normalization operations, they are unable to completely avoid the influence of outliers on model training and fitting. In addition, in the process of model training, we often hope that the deformation features are learned completely and, at the same time, high fitting accuracy can be obtained.

Therefore, in this article, a logarithmic hyperbolic cosine function (LogCosh) [4] was chosen as the base loss function. This function made the model more stable during training (smooth and quadratically differentiable at the closed-form solution) and had a certain degree of robustness to outliers [47], [48]. Considering that this article was a multioutput model, we customized the multiobjective average LogCosh (MOA-LogCosh) and multiobjective average coefficient of determination (MOA-R^2) to train the model. The mathematical equation for both is given as follows:

$$\begin{aligned} &\text{MOA} - \text{LogCosh}\ (y, \hat{y}) \\ &= \frac{1}{4}\left(\sum_{i=1}^{4}\left(\frac{1}{n}\sum_{j=1}^{n}\log\left(\cosh\left(y_i - \hat{y}_i\right)\right)\right)_i\right) \end{aligned} \tag{17}$$

$$\text{MOA} - R^2\ (y, \hat{y}) = \frac{1}{4}\left(\sum_{i=1}^{4}\left(1 - \frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}{\sum_{j=1}^{n}\left(y_j - \bar{y}_j\right)^2}\right)_i\right) \tag{18}$$

TABLE III
TRAINING HYPERPARAMETERS FOR GLER-BIGRUNET IN TYPICAL REGIONS

Hyperparameter Type	Parameter name	Parameter values
Datasets	Sampling Interval	1
	Data repetition (training set repetition coefficient)	2^6
	Data repetition (Validation set repetition coefficient)	2^4
	Random shuffling (buffer size)	Len(train_x) * 2^6
Model training	Time step (Input)	15 (A); 16 (B); 17 (C); 17 (D)
	Time step (Output)	15 (A); 16 (B); 17 (C); 17 (D)
	Epochs	128
	Batch size	64
	Learning rate	1e-4
	Optimizer	Nadam
	Loss	MOA-LogCosh
	Metrics	MOA-R^2
Callback function (reduce learning rate on plateau)	Monitor	Validation loss
	Factor	0.2 (A); 0.16 (B); 0.15 (C); 0.2 (D)
	Patience	4
	Cooldown	4
	Min learning rate	1.00E-15

where n is the number of samples, y_j and $\hat{y}_j$ are the true and predicted values of the ith observation, and $\bar{y}_j$ is the mean value of the observation y_j.

2) Hyperparameter Setting: Choosing the correct hyperparameters can effectively improve the model performance and convergence speed, avoid overfitting and underfitting, and also improve the stability of the model. In this article, two data enhancement techniques, random shuffling, and data repetition were used to increase the diversity and number of samples (see Table III), thus enhancing the generalization ability of the model and reducing overfitting and underfitting. At the same time, we also introduced the Callback function to reduce the learning rate when verifying performance stagnation and help the model converge better, thus avoiding overfitting and underfitting. Moreover, the Nadam optimizer was selected to improve model accuracy and reduce root-mean-square error (RMSE) [49]. After several trials, we determined the overall hyperparameter settings of the model by considering the computational efficiency and hardware conditions, as shown in Table III.

In addition, since the four typical regions are located in the permafrost zone, they are characterized by frost heave and thaw subsidence. Therefore, to ensure that the time-series overlapping segmentation results have a certain periodicity [50] and sufficient model training samples, the fast Fourier transform [51] was used to calculate the time steps. Smaller frequencies are generally chosen as the time step, and it should be ensured that the time step chosen corresponds to the frequency with the smallest absolute value of the phase spectrum. In addition, the empirical value of the time step is generally around 15 [4], [39], [52].

3) Evaluation Indicators: The RMSE, MAE, R^2, and symmetric mean absolute percentage error (SMAPE) were utilized to evaluate the GLER-BiGRUnet model's performance constructed in this article. The detailed formulae of each evaluation index are shown in Fig. 8.

Fig. 8. Evaluation index.

C. Long Time-Series Prediction Method

The multioutput GLER-BiGRUnet constructed in this article was able to take into account the variation of EFs and achieve long time-series prediction. Based on the well-trained SD prediction model, we devised a cyclic prediction method considering EFs, as illustrated in Fig. 9. The specific steps include the following:

1) utilizing GLER-BiGRUnet to simultaneously perform one-step predictions on InSAR SD, LST, NDVI, and precipitation [see Fig. 9(a)], and concatenating the prediction results with the input sequences [see Fig. 9(b)];
2) inputting the concatenated InSAR SD and EFs into GLER-BiGRUnet for one-step prediction [see Fig. 9(b)] and concatenating the prediction results with the input sequences;
3) iteratively applying the aforementioned cyclic prediction mode [2], [37], [52] and performing inverse time-series difference and antinormalization on the final SD [see Fig. 9(d)] [4].

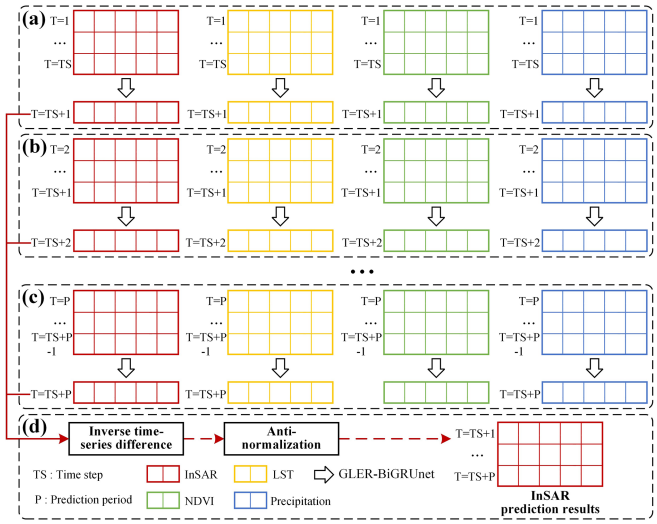

Fig. 9. Prediction flow chart. (a) represents the first step of prediction; (b) represents the second step of prediction; (c) represents the P-th step of prediction; and (d) represents the processing and concatenation of the results from each prediction step.

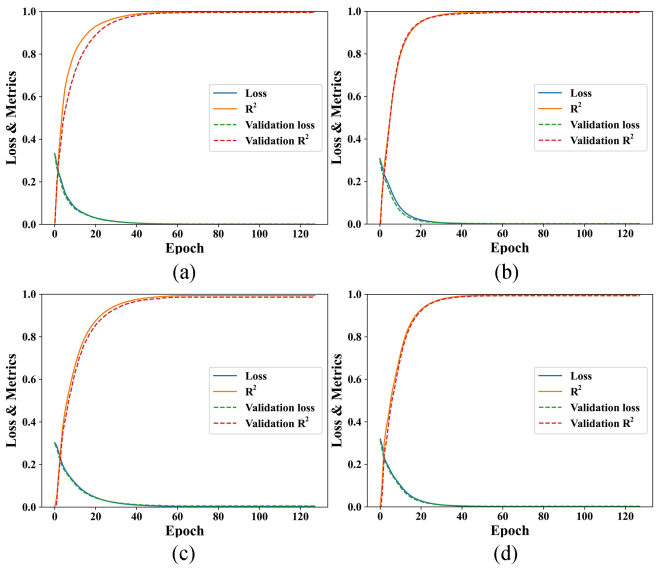

Fig. 10. Variation curves of Loss and R^2 for regions A, B, C, and D are represented in (a), (b), (c), and (d), respectively.

Since the revisit period of the Sentinel-1A satellite is 12 days, and the long time-series prediction in this article starts from the test set, the two-year SD prediction in this article requires 72 iterations, i.e., $P = 72$ in Fig. 9(d).

IV. Results and Analysis

A. Prediction Performance of the Proposed Model

Based on the hyperparameters set in Table III, the proposed GLER-BiGRUnet model in this article converges normally to 0 and 1 for both Loss and R^2 on the four typical regions (see Fig. 10). In addition, when the epoch was located in [4, 40], the model overfitting in region A was serious and then eases, and finally R^2 gradually converges to 1 when epoch $>$ 40. In this process, the adjustment of the learning rate played a key role in suppressing the overfitting, which indicates that the hyperparameters in Table III were set reasonably. Meanwhile, both Loss and R^2 finally converge after 40 epochs of iteration in regions B and D. Moreover, since the number of SD features in region C (50 596) was two to three times that of regions A (19 578), B (24 036), and C (24 726), the Loss and R^2 finally converge after the iteration of epoch 55.

The SD results for four typical regions on 23 June 2022 for comparative analysis with the predicted results from the GLER-BiGRUnet model are shown in Fig. 11. From the first and second columns in Fig. 11, the spatial distributions of the original SD and the predicted SD in regions A–D were consistent, meaning that the proposed prediction model had good predictive performance for different types of SDs (surface uplift and subsidence). In the third column, 50% of the residuals of the original and predicted SD of the four typical regions were located in the ranges of [−1.20, 0.81], [−1.13, 1.66], [−1.41, 0.89], and [−1.45, 1.71], and the mean values of the residuals were better than ± 0.5 mm. In addition, the residual results of regions A, B, and D in the third column were more in line with the positive distribution, while the residual results of region C were concentrated in the interval [−2.00, 2.00]. In the fourth column, the fits between the true SD and the predicted SD of the four profiles were better, and the residuals fluctuated around 0. The above results indicated that the proposed GLER-BiGRUnet model in this article was accurate in predicting the spatial distribution as well as the local profiles in different regions.

B. Comparison With Other Time-Series Prediction Network Models

To validate the prediction accuracy of the GLER-BiGRUnet model, Conv1D-MHCA-BiGRU, MHCA-BiGRU, MHCA-GRU, MHSA-BiGRU, MHSA-GRU [53], SA-BiGRU, SA-GRU, BiGRU [54], and GRU [55] were selected (see Table IV). In the model considering only InSAR SD, SA-GRU scored worse than BiGRU for all evaluation metrics in regions A, B, and D, while it outperformed BiGRU in region C. This may be due to overfitting in regions with fewer SD features as the model complexity increases. Conversely, in regions with more SD features, the increase in model complexity helps the model to learn the SD features better. In addition, a similar situation exists in regions A and C for MHSA-GRU and SA-GRU.

Among the models considering the multiscale features of SD and EFs, our proposed GLER-BiGRUnet model scores higher on all evaluation metrics. At the same time, it does not appear that models with higher model complexity have lower evaluation metrics instead. In addition, the RMSE score of the GLER-BiGRUnet model was 68% better than the MHCA-GRU model (only considered EFs) in region A. In region C, the GLER-BiGRUnet improved the RMSE score by 24% over the Conv1D-MHCA-BiGRU, which did not consider multiscale SD features. Moreover, the comparison among Conv1D-MHCA-BiGRU, MHCA-BiGRU, and MHCA-GRU two-by-two showed

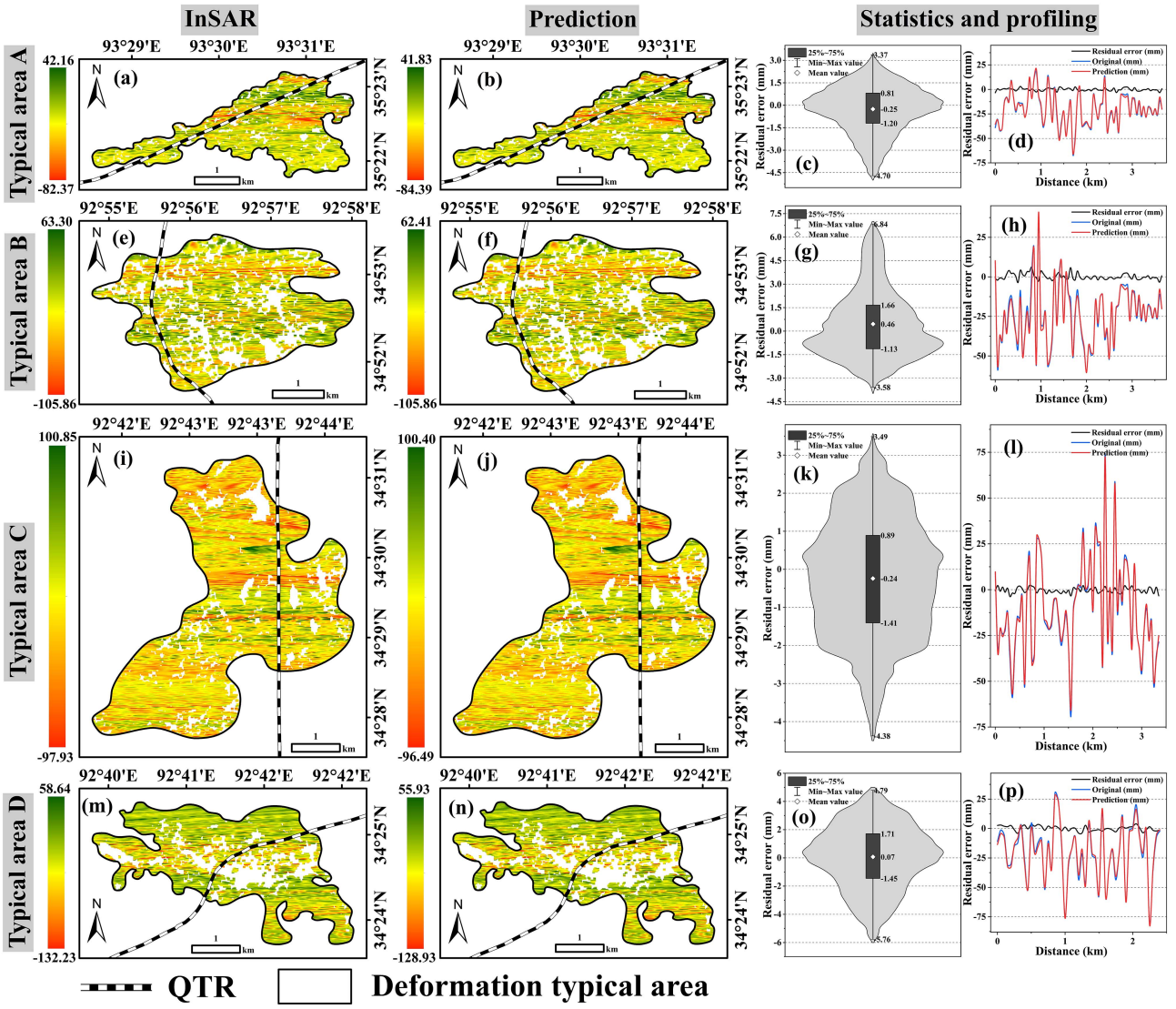

Fig. 11. Comparison of typical region SD prediction results on spatial distribution and profile. The first and second columns are the original and predicted SD, the third column is the mathematical statistics result of the difference between the data in the first and second columns, and the fourth column is the comparison of the predicted results on the profile.

that considering local or bidirectional SD features enhances prediction accuracy. Overall, the proposed GLER-BiGRUnet had the highest prediction accuracy in different deformation regions.

C. Prediction Results of SD

We used the proposed GLER-BiGRUnet model to predict the SD trend of four typical regions along the QTR in the next two years, and compared the prediction results with other models (see Fig. 12). On the test set, the prediction results of models, such as MHSA-BiGRU and GRU, have large discrepancies with the InSAR SD, and the prediction trends were unstable [see Fig. 12(b) and (f)]. Models, such as the MHCA-BiGRU, have a significant lack of feature capturing ability at some locations where the trend direction of the InSAR SD changes [see Fig. 12(b), (d), (f), and (h)]. On the contrary, the proposed GLER-BiGRUnet model was able to better fit the InSAR SD by considering the local details as well as the multiscale features of SD.

The prediction results of the model without considering the EFs changes were unstable in the long time-series prediction for the next two years. The predictions of the model without considering the multiscale features were significantly more stable, but the predictions did not accurately reflect the future trends of SD (see Fig. 12). Especially in region A, there was a large discrepancy between the predictions of different prediction models. However, compared with other prediction models, the proposed GLER-BiGRUnet was able to learn the multiscale deformation characteristics of region A with steeply increasing or decreasing SD trends on 11 January 2021, 28 February 2021, and 29 April 2021 [see Fig. 12(a)]. This resulted in the highest agreement between the GLER-BiGRUnet model predictions and the InSAR SD trends. In addition, in regions B, C, and D, the InSAR SD trend was smooth, and different prediction model's prediction results were more concentrated. Although the prediction results

TABLE IV
COMPARISON OF THE GLER-BIGRUNET MODEL WITH OTHER PREDICTIVE MODELS ON DIFFERENT EVALUATION METRICS

Typical region	Evaluation index	Consider EFs models				Only consider InSAR SD models					
		GLER-BiGRUnet	Conv1D-MHCA-BiGRU	MHCA-BiGRU	MHCA-GRU	MHSA-BiGRU	MHSA-GRU	SA-BiGRU	SA-GRU	BiGRU	GRU
A	RMSE	**0.163**	0.275	0.405	0.513	0.573	0.708	0.681	0.867	0.766	1.057
	MAE	**0.147**	0.265	0.400	0.492	0.558	0.704	0.656	0.856	0.707	1.052
	R^2	**0.988**	0.967	0.928	0.884	0.856	0.78	0.796	0.669	0.742	0.508
	SMAPE	**0.793**	1.447	2.168	2.664	3.036	3.75	3.485	4.516	3.749	5.530
B	RMSE	**0.205**	0.357	0.462	0.578	0.681	0.744	0.855	1.053	0.943	1.188
	MAE	**0.199**	0.352	0.456	0.571	0.675	0.739	0.851	1.048	0.939	1.185
	R^2	**0.990**	0.970	0.950	0.922	0.892	0.871	0.829	0.742	0.792	0.671
	SMAPE	**0.855**	1.510	1.950	2.426	2.854	3.097	3.604	4.404	3.973	4.970
C	RMSE	**0.212**	0.279	0.356	0.458	0.643	0.742	0.837	0.942	1.083	1.287
	MAE	**0.195**	0.271	0.351	0.448	0.642	0.741	0.832	0.937	1.078	1.283
	R^2	**0.983**	0.972	0.954	0.923	0.849	0.799	0.743	0.675	0.571	0.394
	SMAPE	**1.380**	1.912	2.476	3.156	4.495	5.140	5.848	6.469	7.311	8.645
D	RMSE	**0.166**	0.241	0.337	0.392	0.553	0.620	0.718	0.954	0.82	1.140
	MAE	**0.157**	0.236	0.336	0.389	0.551	0.617	0.716	0.953	0.819	1.140
	R^2	**0.995**	0.988	0.977	0.969	0.939	0.924	0.898	0.819	0.866	0.742
	SMAPE	**0.780**	1.210	1.713	1.982	2.801	3.194	3.718	4.858	4.198	5.873

Note: Bold entities/values represent the highest evaluation metric scores.

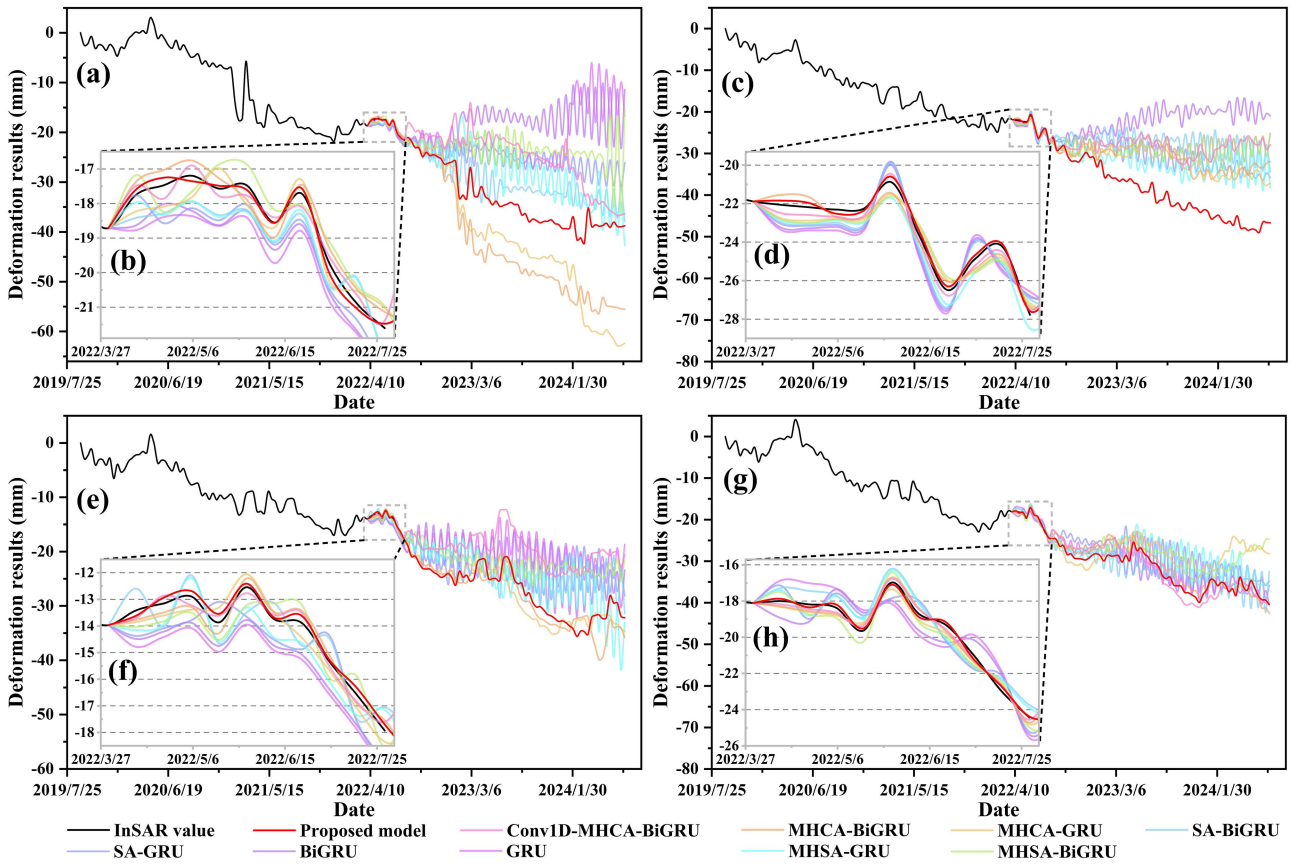

Fig. 12. SD predictions from different models. (a), (c), (e), and (g) are the changes in the mean values of the original SD and the predicted SD for regions A, B, C, and D, respectively. (b), (d), (f), and (h) are the changes in the mean values of the original and predicted SD on the test set for the four regions, respectively.

of Conv1D-MHCA-BiGRU, MHCA-BiGRU, and MHCA-GRU models were roughly the same as the deformation trend of InSAR, they still cannot predict the SD at some detailed locations. On the contrary, the proposed GLER-BiGRUnet model had the most stable prediction results in different regions and can best represent the future development trend of SD in typical regions.

V. Conclusion

This article proposed a BiGRU multioutput SD prediction network called GLER-BiGRUnet, which addresses the challenge of accurately long time-series prediction while preserving multiscale SD and EFs by fully considering both InSAR SD and EFs features. Specifically, the proposed GLER-BiGRUnet model mainly consists of GLFE, MCAR, and RCSA module. The GLFE module extracts the global–local features of SD and EFs through the concatenation of Dense and Conv1D. The MCAR module obtains the long time-series dependency between SD and EFs by introducing MHCA and BiGRU, and extracts the forward and backward feature information. The RCSA module fuses the long-distance, multiscale SD and EFs feature information through the residual concatenation and SA mechanism. Based on the four typical regions in the permafrost area of the QTR InSAR SD and EFs dataset, we verify the effectiveness of the proposed GLER-BiGRUnet. The proposed GLER-BiGRUnet model can predict long time-series SD, and the prediction results have higher stability. In future work, we encompass the generalization of the proposed GLER-BiGRUnet to widen the applicability and effectiveness of GLER-BiGRUnet across various scenarios.

References

[1] Y. Yuan et al., "Land subsidence prediction in Zhengzhou's main urban area using the GTWR and LSTM models combined with the attention mechanism," *Sci. Total Environ.*, vol. 907, 2024, Art. no. 167482.

[2] C. Yi et al., "Surface deformation prediction based on TS-InSAR technology and long short-term memory networks," *Nat. Remote Sens. Bull.*, vol. 26, no. 7, pp. 1326–1341, 2022.

[3] Q. Liu, Y. Zhang, M. Deng, H. Wu, Y. Kong, and J. Wei, "A deep learning prediction method for large-scale surface settlement time series," *J. Surv. Mapping*, vol. 50, pp. 396–404, 2021.

[4] Y. He et al., "Time-series analysis and prediction of surface deformation in the Jinchuan mining area, Gansu Province, by using InSAR and CNN–PhLSTM network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6732–6751, Aug. 2022.

[5] F. Cigna and D. Tapete, "Present-day land subsidence rates, surface faulting hazard and risk in Mexico City with 2014–2020 Sentinel-1 IW InSAR," *Remote Sens. Environ.*, vol. 253, 2021, Art. no. 112161.

[6] M. Peng et al., "Mapping land subsidence and aquifer system properties of the Willcox Basin, Arizona, from InSAR observations and independent component analysis," *Remote Sens. Environ.*, vol. 271, 2022, Art. no. 112894.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatterers in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 8–20, Jan. 2001.

[9] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2375–2383, Nov. 2002.

[10] M. Peng et al., "Characterization and prediction of InSAR-derived ground motion with ICA-assisted LSTM model," *Remote Sens. Environ.*, vol. 301, 2024, Art. no. 113923.

[11] J. Wang et al., "InSAR time-series deformation forecasting surrounding Salt Lake using deep transformer models," *Sci. Total Environ.*, vol. 858, 2023, Art. no. 159744.

[12] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6000–6010.

[13] W. Ma, Z. Wen, Y. Sheng, Q. Wu, D. Wang, and W. Feng, "Remedying embankment thaw settlement in a warm permafrost region with thermosyphons and crushed rock revetment," *Can. Geotech. J.*, vol. 49, no. 9, pp. 1005–1014, 2012.

[14] J. Luo, F. Niu, Z. Lin, M. Liu, G. Yin, and Z. Gao, "Inventory and frequency of retrogressive thaw slumps in permafrost region of the Qinghai–Tibet plateau," *Geophysical Res. Lett.*, vol. 49, no. 23, 2022, Art. no. e2022GL099829.

[15] G. Yang et al., "Characteristics of methane emissions from alpine thermokarst lakes on the Tibetan plateau," *Nature Commun.*, vol. 14, no. 1, 2023, Art. no. 3121.

[16] E. A. G. Schuur and M. C. Mack, "Ecological response to permafrost thaw and consequences for local and global ecosystem services," *Annu. Rev. Ecol., Evol., Systematics*, vol. 49, no. 1, pp. 279–301, 2018.

[17] J. Hjort, D. Streletskiy, G. Doré, Q. Wu, K. Bjella, and M. Luoto, "Impacts of permafrost degradation on infrastructure," *Nature Rev. Earth Environ.*, vol. 3, no. 1, pp. 24–38, 2022.

[18] R. Li et al., "Impact of climate warming on permafrost changes in the Qinghai-Tibet plateau," *Cold Regions Sci. Technol.*, vol. 205, 2023, Art. no. 103692.

[19] Y. Ran et al., "Permafrost degradation increases risk and large future costs of infrastructure on the third pole," *Commun. Earth Environ.*, vol. 3, no. 1, 2022, Art. no. 238.

[20] X. Zhang, S. Cheng, L. Wang, and H. Li, "Asymmetric cross-attention hierarchical network based on CNN and transformer for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 2000415.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[22] A. Gao et al., "A vector-based coastline shape classification approach using sequential deep learning model," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 129, 2024, Art. no. 103810.

[23] Y. He, W. Wang, H. Yan, L. Zhang, Y. Chen, and S. Yang, "Characteristics of surface deformation in Lanzhou with Sentinel-1A TOPS," *Geosciences*, vol. 10, no. 3, 2020, Art. no. 99.

[24] Y. He, Y. Chen, W. Wang, H. Yan, L. Zhang, and T. Liu, "TS-InSAR analysis for monitoring ground deformation in Lanzhou new district, the loess plateau of China, from 2017 to 2019," *Adv. Space Res.*, vol. 67, no. 4, pp. 1267–1283, 2021.

[25] Y. He et al., "An identification method of potential landslide zones using InSAR data and landslide susceptibility," *Geomatics, Natural Hazards Risk*, vol. 14, no. 1, 2023, Art. no. 2185120.

[26] Y. He et al., "An integrated neural network method for landslide susceptibility assessment based on time-series InSAR deformation dynamic features," *Int. J. Digit. Earth*, vol. 17, no. 1, 2024, Art. no. 2295408.

[27] C. Yu, Z. Li, N. T. Penna, and P. Crippa, "Generic atmospheric correction model for interferometric synthetic aperture radar observations," *J. Geophysical Res., Solid Earth*, vol. 123, no. 10, pp. 9202–9222, 2018.

[28] H. A. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 950–959, Sep. 1992.

[29] Y. He et al., "Thaw slump susceptibility mapping based on sample optimization and ensemble learning techniques in Qinghai-Tibet railway corridor," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5443–5459, Feb. 2024.

[30] J. Luo, F. Niu, Z. Lin, M. Liu, and G. Yin, "Thermokarst lake changes between 1969 and 2010 in the Beilu river basin, Qinghai–Tibet plateau, China," *Sci. Bull.*, vol. 60, no. 5, pp. 556–564, 2015.

[31] Z. Zhang, M. Wang, Z. Wu, and X. Liu, "Permafrost deformation monitoring along the Qinghai-Tibet plateau engineering corridor using InSAR observations with multi-sensor SAR datasets from 1997–2018," *Sensors*, vol. 19, no. 23, 2019, Art. no. 5306.

[32] G. Zhou et al., "Accelerating thermokarst lake changes on the Qinghai–Tibetan plateau," *Sci. Rep.*, vol. 14, no. 1, 2024, Art. no. 2985.

[33] D. Zou et al., "A new map of permafrost distribution on the Tibetan plateau," *Cryosphere*, vol. 11, no. 6, pp. 2527–2542, 2017.

[34] S. L. Smith, H. B. O'Neill, K. Isaksen, J. Noetzli, and V. E. Romanovsky, "The changing thermal state of permafrost," *Nature Rev. Earth Environ.*, vol. 3, no. 1, pp. 10–23, 2022.

[35] M. M. Heijmans et al., "Tundra vegetation change and impacts on permafrost," *Nature Rev. Earth Environ.*, vol. 3, no. 1, pp. 68–84, 2022.
[36] J. Hao et al., "Numerical analysis of the impacts of rainfall on permafrost-related slope stability on the Qinghai–Tibet plateau," *J. Hydrol., Regional Stud.*, vol. 47, 2023, Art. no. 101439.
[37] S. Yao et al., "A convLSTM neural network model for spatiotemporal prediction of mining area surface deformation based on SBAS-InSAR monitoring data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jan. 2023, Art. no. 5201722.
[38] Y. Wang et al., "Ground deformation analysis using InSAR and backpropagation prediction with influencing factors in Erhai Region, China," *Sustainability*, vol. 11, no. 10, 2019, Art. no. 2853.
[39] Y. Chen et al., "Prediction of InSAR deformation time-series using a long short-term memory neural network," *Int. J. Remote Sens.*, vol. 42, no. 18, pp. 6919–6942, 2021.
[40] S.-H. Xiong, Z.-P. Wang, G. Li, M. J. Skibniewski, and Z.-S. Chen, "Prediction of airport runway settlement using an integrated SBAS-InSAR and BP-EnKF approach," *Inf. Sci.*, vol. 665, 2024, Art. no. 120376.
[41] F. Yuan, Z. Zhang, and Z. Fang, "An effective CNN and transformer complementary network for medical image segmentation," *Pattern Recognit.*, vol. 136, 2023, Art. no. 109228.
[42] N. M. Foumani, L. Miller, C. W. Tan, G. I. Webb, G. Forestier, and M. Salehi, "Deep learning for time series classification and extrinsic regression: A current survey," *ACM Comput. Surv.*, vol. 56, no. 9, 2024, Art. no. 217.
[43] Y. Liu and J. Zhang, "Integrating SBAS-InSAR and AT-LSTM for time-series analysis and prediction method of ground subsidence in mining areas," *Remote Sens.*, vol. 15, no. 13, 2023, Art. no. 3409.
[44] Q. Yu, W. Wei, Z. Pan, J. He, S. Wang, and D. Hong, "GPF-net: Graph-polarized fusion network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Aug. 2023, Art. no. 5519622.
[45] L. Li et al., "Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111584.
[46] Q. Ding, Z. Shao, X. Huang, O. Altan, Q. Zhuang, and B. Hu, "Monitoring, analyzing and predicting urban surface subsidence: A case study of Wuhan City, China," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 102, 2021, Art. no. 102422.
[47] A. Jadon, A. Patil, and S. Jadon, "A comprehensive survey of regression-based loss functions for time series forecasting," in *Proc. Int. Conf. Data Manage. Analytics Innov.*, 2024, pp. 117–147.
[48] R. A. Saleh and A. Saleh, "Statistical properties of the log-cosh loss function used in machine learning," 2022, *arXiv:2208.04564*.
[49] J. Sharma et al., "A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: A case study of India," *Energy Sci. Eng.*, vol. 10, no. 8, pp. 2909–2929, 2022.
[50] P. Hill, J. Biggs, V. Ponce-López, and D. Bull, "Time-series prediction approaches to forecasting deformation in Sentinel-1 INSAR data," *J. Geophysical Res., Solid Earth*, vol. 126, no. 3, 2021, Art. no. e2020JB020176.
[51] X. Liu, Y. Wang, W. Zhan, and T. Yu, "Improving MODIS precipitable water vapour in mainland China based on the LSF model," *Adv. Space Res.*, vol. 72, no. 8, pp. 3133–3149, 2023.
[52] Y. He, S. Yao, Y. Chen, H. Yan, and L. Zhang, "Spatio-temporal prediction of time-series InSAR land subsidence based on ConvLSTM neural network," Geomatics Inf. Sci. Wuhan Univ., pp. 1–21, 2023.
[53] Y. Duan, P. Liu, and Y. Lu, "MhSa-GRU: Combining user's dynamic preferences and items' correlation to augment sequence recommendation," *J. Intell. Inf. Syst.*, vol. 61, no. 1, pp. 225–248, 2023.
[54] C. Zhang et al., "Temporal data-driven failure prognostics using BiGRU for optical networks," *J. Opt. Commun. Netw.*, vol. 12, no. 8, pp. 277–287, 2020.
[55] S. Kumar, D. Kumar, P. K. Donta, and T. Amgoth, "Land subsidence prediction using recurrent neural networks," *Stochastic Environ. Res. Risk Assessment*, vol. 36, no. 2, pp. 373–388, 2022.

Tianbao Huo received the B.E. degree in surveying and mapping engineering in 2022 from Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in mapping science and technology with the Faculty of Geomatics.

His research interests include land subsidence monitoring and time-series InSAR prediction based on neural networks.

Yi He (Member, IEEE) received the B.S. degree in geographic information system from Lanzhou Jiaotong University, Lanzhou, China, in 2011, and the Ph.D. degree in Earth system science from Lanzhou University, Lanzhou, China, in 2016.

He has been a Postdoctoral Researcher with the School of Environment and Municipal Engineering, Lanzhou Jiaotong University, where he is currently a Professor with the Faculty of Geomatics. His research interests include disaster remote sensing, ecological remote sensing, image processing, and time-series InSAR prediction based on deep learning.

Lifeng Zhang received the M.S. degree in cartography and geographic information system and the Ph.D. degree in environmental science and engineering from Lanzhou Jiaotong University, Lanzhou, China, in 2010 and 2017, respectively.

He is currently a Professor with the Faculty of Geomatics, Lanzhou Jiaotong University. His research interests include ecological environment remote sensing monitoring, visualized analysis, and analysis of land use change.

Wang Yang received the M.S. degree in surveying and mapping in 2023 from Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the Ph.D. degree in surveying and mapping with the Faculty of Geomatics.

His research interests include InSAR data processing technology, machine learning, and remote sensing image information extraction.

Jiapeng Tang received the B.E. degree in remote sensing science and technology in 2021 from Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in mapping science and technology with the Faculty of Geomatics.

His research interests include InSAR data processing technology and geological hazard identification of highway corridor in loess area.

Qing Zhang received the B.E. degree in surveying and mapping engineering in 2022 from Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in mapping science and technology with the Faculty of Geomatics.

His research interests include InSAR data processing technology, machine learning, and remote sensing image processing.

Jiangang Lu received the B.E. degree in remote sensing science and technology from the Shandong University of Science and Technology, Qingdao, China, in 2020. He is currently working toward the M.S. degree in mapping science and technology with the Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, China.

His research interests include landslide hazard risk assessment, machine learning, and remote sensing image information extraction.

Yunhao Zhang received the B.E. degree in surveying and mapping engineering in 2023 from Lanzhou Jiaotong University, Lanzhou, China, where he is currently working toward the M.S. degree in resources and environmental engineering with the Faculty of Geomatics.

His research interests include landslide displacement mechanism analysis and trend prediction.