Monthly NDVI Prediction Using Spatial Autocorrelation and Nonlocal Attention Networks

Lei Xu ⓘ, Ruinan Cai ⓘ, Hongchu Yu ⓘ, Wenying Du ⓘ, Zeqiang Chen ⓘ, and Nengcheng Chen ⓘ

Abstract—Accurate prediction of vegetation indices is useful for helping maintain vegetation stability, sustaining food production, and reducing socioeconomic losses. The traditional convolutional long short-term memory (ConvLSTM) model for vegetation prediction ignores the spatial aggregation characteristics of the normalized difference vegetation index (NDVI) itself and the global dependence information in space. In this study, we propose a new NDVI prediction method, namely, the ConvLSTM with spatial autocorrelation and nonlocal attention module (ConvLSTM-SAC-NL), by combining the nonlocal attention module to capture long-range dependence and the spatial autocorrelation modeling based on the local Moran index to learn spatial dependence. The experimental results indicate that the ConvLSTM-SAC-NL model outperforms seven baseline forecasting models, with an R^2 of 0.881 in monthly NDVI prediction in the Huangpi district of Wuhan city, relative to the R^2 values of 0.758, 0.777, 0.741, 0.776, 0.804, 0.829, and 0.815 for random forest, support vector machine regression, long short-term memory, bidirectional long short-term memory, graph convolutional network, predictive recurrent neural network, and ConvLSTM models, respectively. Spatially, the prediction results of the ConvLSTM-SAC-NL model demonstrate improved accuracy over 91.49% of the study area when compared with ConvLTSM. Therefore, the proposed ConvLSTM-SAC-NL model could serve as an effective approach for short-term prediction of vegetation conditions at regional scales.

Index Terms—Convolutional long short-term memory (Conv-LSTM), nonlocal attention module, normalized difference vegetation index (NDVI), spatial autocorrelation, spatiotemporal prediction.

I. INTRODUCTION

VEGETATION is an important part of terrestrial ecosystems and plays an important role in regulating terrestrial energy exchange, water, carbon cycles, and ecological balance [1], [2]. Accurate monthly forecasting of vegetation dynamics is beneficial for early warning of ecosystem instability, preparation of timely crop management, and potential reduction of tree mortality from extreme weather events.

Satellite remote sensing technology is a scientific and efficient tool for large-scale vegetation condition monitoring, which is widely used in the fields of ecological conservation and environmental change [3], [4]. Satellite remote sensing provides a rich data source for extracting and monitoring spatial and temporal dynamics of terrestrial ecosystems [5]. It is convenient, scientific, and feasible to monitor and predict vegetation changes based on satellite-based vegetation indices. Satellite-based vegetation indices are usually calculated by the combinations of spectral bands from remote sensing images and are representations of vegetation growth and health. Normalized difference vegetation index (NDVI) is one of the most commonly used vegetation indices and is derived from the ratio of the difference between red and near-infrared reflectance to its sum. NDVI has been widely used in agricultural production prediction [6], drought monitoring [7], [8], land cover classification [9], [10], and so on. Therefore, accurate NDVI forecasting is useful for agricultural and ecological applications.

Previously, NDVI predictions were mainly based on statistical methods, such as the autoregressive integrated moving average combined with climate data [11] and the triple exponential smoothing method based on the Holt–Winters additive model [12]. However, because of the impact of soil moisture [13], temperature, human activities, and other factors, the underlying mechanisms of vegetation changes are complex. Statistical methods usually assume that the NDVI time series is linear and smooth, while the nonlinear and unsmooth information hidden in the NDVI time series cannot be captured well during the prediction [14].

With the rapid development of artificial intelligence technology, machine learning (ML) methods have been gradually applied in the field of remote sensing. For example, Tong et al. [15] proposed a pixel space gap-filling method for soil moisture estimation in the Qinghai–Tibet Plateau using ML and geostatistics. Compared with traditional statistical prediction methods, ML methods can deal with complex forecasting problems with noisy data by learning the nonlinear relationships between dependent and independent variables through sufficient training [16]. Roy et al. [17] used four ML methods [support vector machine regression (SVR), random forest (RF), linear

and polynomial regression] to predict monthly NDVI data, and found that RF and linear regression methods can obtain high forecasting accuracy, with some effects for the abrupt changes.

Deep learning models have been widely used in remote sensing applications in recent years. The convolutional neural network-long short-term memory (CNN-LSTM) model, which combines the spatial modeling of the convolutional neural network (CNN) and the temporal memory of long short-term memory (LSTM), has been applied to some prediction tasks [18]. Artificial neural networks and cycle-consistent adversarial networks are used for soil moisture time series analysis and estimation, respectively, [19], [20]. In recent years, some scholars proposed new neural network methods to improve NDVI prediction accuracy. Gao et al. [21] combined time series decomposition (TSD), CNN, and LSTM to build a new monthly NDVI prediction model termed TSD-CNN-LSTM, which has the best predictive performance with a root-mean-square error (RMSE), MSE, and mean absolute error (MAE) of 0.0573, 0.9617, and 0.0447, respectively. Cui et al. [22] proposed a new NDVI prediction model called SF-CNN by combining CNN and statistical features of historical data, indicating improved accuracy for the subsequent three months in multiple complex regions.

In spatiotemporal NDVI prediction, most of the existing studies used NDVI time series characteristics and ML to improve the prediction accuracy, while ignoring the spatial dependence and aggregation of vegetation status. Attention may be an effective concept to model the global dependence of vegetation distribution in NDVI prediction. The attention mechanism is based on the characteristics of human vision: when humans find that a scene often has something they want to observe in a certain part, they will focus on that part and concentrate more on the useful part when a similar scene occurs again. The attention mechanism makes the model focus more on a certain part of the data during training and improves the accuracy of perceptual information processing [23]. Attention mechanisms have become an increasingly common component of neural architectures and have been applied to a variety of tasks, such as text classification [24], machine translation [25], motion recognition [26], and voice recognition [27]. Attention mechanisms are also widely applied in remote sensing applications. For instance, Cai et al. [28] employed graph convolution and cross-attention to perform change detection in remote sensing images. Shi et al. [29] proposed DSAMNet based on deep supervised attention measurement, for identifying changes in aerial imagery. Liu et al. [30] created a new deep learning model called RAANet by incorporating attention modules and residual structures into ASPP and improved the classification accuracy of land use and land cover (LULC). In addition, the topographic conditions and climatic factors that influence vegetation growth are generally spatially continuous, and thus the surrounding vegetation condition may be closely correlated with the target vegetation condition. Therefore, the spatial dependence information of neighboring NDVI pixels may be utilized to enhance the NDVI prediction skill. Spatial autocorrelation is a general statistical property of events or phenomena observed across geographic space. Spatial autocorrelation analysis has been used in economic, social, ecological, and other fields. Many geophysical data usually have spatial autocorrelation, such as precipitation, temperature, and vegetation [31], [32], [33]. The existing NDVI prediction simply uses a convolution kernel to consider the spatial autocorrelation of surrounding pixels, which could not identify the geographic variations of local spatial dependence well.

In this article, we propose a refined NDVI forecasting approach, namely ConvLSTM with spatial autocorrelation and nonlocal attention module (ConvLSTM-SAC-NL), by combining the convolutional long short-term memory (ConvLSTM) model, the nonlocal attention module and spatial autocorrelation modeling. The SAC module is added to focus on the local spatial dependence modeling of the whole study area and the NL module is added to capture the long-range dependence of NDVI images. The developed ConvLSTM-SAC-NL model is compared with seven baseline forecasting models to demonstrate its effectiveness and superiority. The key contributions of this study are summarized as follows.

1) The local Moran index is used as a spatial autocorrelation factor to learn the spatial correlations between each pixel and neighboring pixels for NDVI forecasting.
2) The nonlocal attention module is added to the ConvLSTM model to learn global relevance information and capture long-range dependence.
3) A new data-driven spatiotemporal deep learning method ConvLSTM-SAC-NL is proposed for predicting NDVI at a monthly scale and 1 km spatial resolution.

We innovatively introduce the local Moran index to model the spatial autocorrelation phenomenon of vegetation conditions and incorporate it into a deep learning model, which significantly improves the effectiveness of dynamic vegetation prediction and addresses the issue of ignoring spatial correlation relationships in previous studies. By incorporating the nonlocal attention module, the proposed ConvLSTM-SAC-NL model captures long-range dependence in image data. The predictive performance of the ConvLSTM-SAC-NL model surpasses the traditional ConvLSTM model and some popular deep learning models in the NDVI forecasting experiment, which may provide promising insights into data-driven vegetation prediction science.

The rest of this article is organized as follows. Section II describes the study area and data sources. Section III presents the proposed forecasting model. Section IV presents the experimental results. Section V discusses the experiment. Finally, Section VI concludes this article.

II. STUDY AREA AND DATA SOURCES

A. Study Area

Huangpi district is an area with many vegetation types and famous scenic areas, which is the agglomeration area of natural landscapes in Wuhan city. It is of great significance to effectively predict the vegetation conditions in Huangpi district for the ecological civilization construction of Wuhan city.

Fig. 1 shows the geographical location and topographic view of Huangpi district. Huangpi district is located in the north of Wuhan city and the eastern part of Hubei province. The
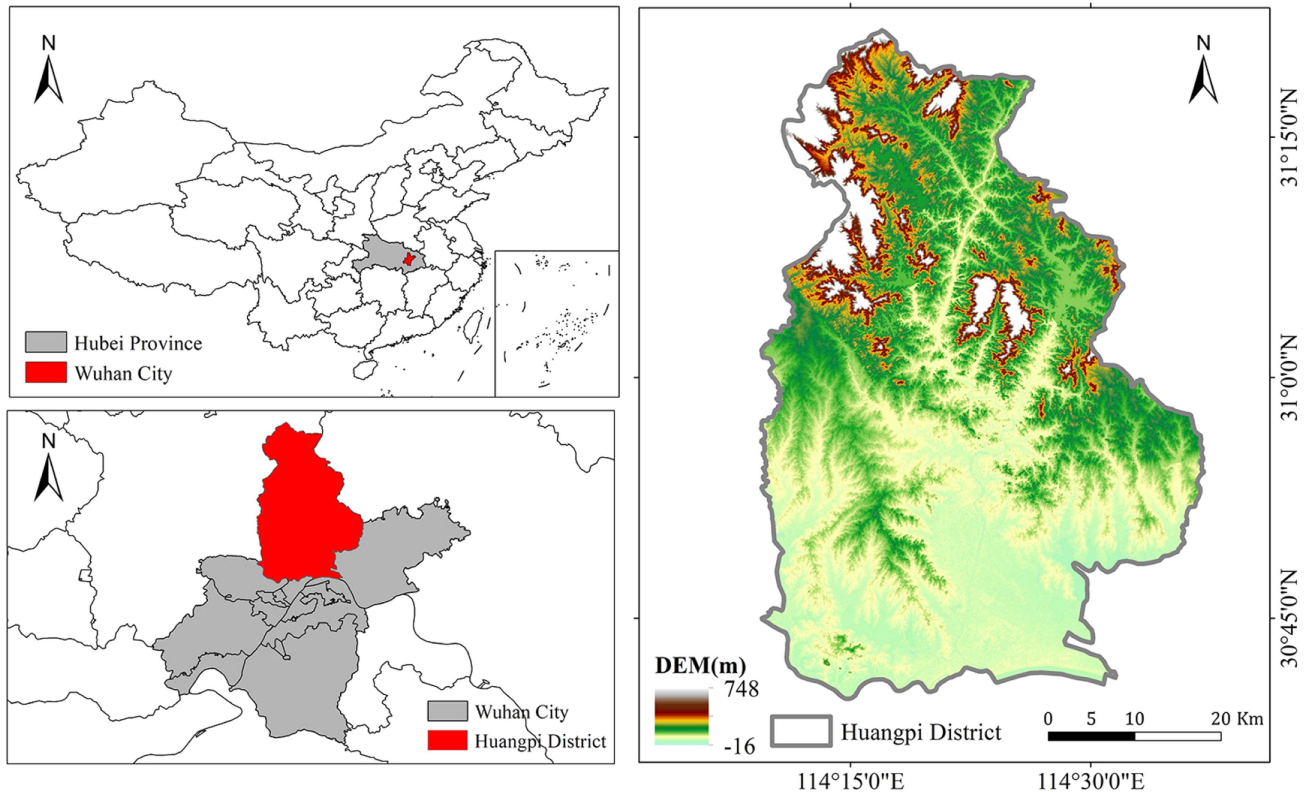
Fig. 1.   Location and DEM map of the studying area.

geographical coordinates are between 114°09′–114°37′ east longitude and 30°40′–31°22′ north latitude, with an area of 2261 km². The northern part of Huangpi district belongs to the western section of the southern foothills of Dabie mountain. The southern part of Huangpi district is located at the eastern section of the north bank of the Yangtze river in the Jianghan plain. The topography is high in the north and low in the south, forming a four-tier ladder of low mountainous areas in the northwest, hilly areas in the northeast, post-like plains in the middle, and lakeside plains in the south.

Huangpi district features a subtropical monsoon climate, with abundant rainfall, four distinct seasons, and an average annual frost-free period of 255 days. The multiyear averaged sunshine hours are 1917.4 h, and the multiyear averaged precipitation is 1202 mm. In general, the climatic conditions in Huangpi district are ideal for vegetation growth.

B. Data and Preprocessing

The moderate resolution imaging spectroradiometer (MODIS) is a key component of NASA's Earth Observing System. It is designed to enable long-term global monitoring of the Earth's land, ocean, and atmospheric properties. MODIS data are free worldwide, with a wide spectral range, simple data reception, and high update frequency. MODIS contains a variety of data products related to the ocean, land, and atmosphere sciences, which are widely used in geophysical applications. The NDVI data, LULC data were derived from MODIS and

downloaded through Google Earth Engine in the study. Table I shows the used data introduction.

The NDVI data were collected from MOD13A2, with a temporal resolution of 16-day (23 scenes/year) and a spatial resolution of 1 km. We synthesized 16-day NDVI data into monthly data using the maximum synthesis method, which simultaneously eliminates the interference of clouds, atmosphere, and solar height. A total of 192 monthly NDVI images were eventually obtained for 16 years from 2003 to 2018, with an image size of 81 rows and 57 columns.

The LULC data were collected from MCD12Q1, with a temporal resolution of 1-year and a spatial resolution of 500 m. We used the nearest neighbor sampling method in ArcGIS software to resample the 500 m LULC images to match the 1 km spatial resolution of NDVI. We finally obtained a total of 16 LULC images for 16 years from 2003 to 2018. In the ConvLSTM-SAC-NL model, one LULC image was shared for each year (12 months) and was used as a feature along with the local Moran index to construct the predictors for NDVI prediction.

III. METHODS

A. Spatial Autocorrelation

The spatial autocorrelation statistic is used to measure the degree of interdependence between data at one location and data at other geographic locations. Due to the influence of spatial interactions and spatial diffusion, most geographic data may not

TABLE I
DATA INTRODUCTION

Data	Product ID	Spatial resolution	Time resolution	Temporal extent
NDVI	MOD13A2	1km	16-day	2000-02-18 to present
LULC	MCD12Q1	500m	1-year	2001-01-01 to present

be independent of each other. According to previous studies, NDVI is spatially autocorrelated. Therefore, here the global Moran index is used to determine whether the study area has spatial aggregation and the local Moran index is used to indicate where the aggregation phenomena and outliers are. Finally, we input the local Moran index as a feature into the model to improve the prediction accuracy.

The global Moran index explains the first law of geography in a mathematical sense: all things are related, but nearby things are more related than distant things. Spatial autocorrelation measures the degree of aggregation of unit attribute values in geographically contiguous units with similar attributes, it is an important method for exploring the spatial differentiation of geographic attributes. The main formulas of the global Moran index are

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}(x_i - \bar{X})(x_j - \overline{X})}{\sum_{i=1}^{n}(x_i - \bar{X})^2} \quad (1)$$

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \quad (2)$$

$$Z_I = \frac{I - E[I]}{\sqrt{V[I]}} \quad (3)$$

$$E[I] = -1/n - 1 \quad (4)$$

$$V[I] = E[I^2] - E[I]^2 \quad (5)$$

where x_i is the attribute value of the pixel, $\bar{X}$ is the mean value of the attribute, n is the total number of pixels. $w_{i,j}$ is the spatial weight between elements i and j, S_0 is the aggregation of all spatial weights, and Z_I is the Z score. The calculation starts by defining a spatial weight matrix to measure the matrix of spatial relationships between different locations. The commonly used spatial weight matrices include the neighborhood matrix, distance matrix, radius matrix, etc. For each location, the correlation between the target's pixel value and the pixel values at neighboring other locations is calculated. Finally, the spatial autocorrelation indices of all locations are summed and normalized to obtain the global Moran index.

The global Moran index ranges between -1.0 and $+1.0$. When the global Moran index is above 0, the data show a positive spatial correlation, and the larger value indicates a more obvious spatial correlation. On the contrary, when the global Moran index is below 0, the data show a negative spatial correlation, with smaller values showing greater spatial differences. When the global Moran index is 0, it indicates that the data are randomly distributed in space. The confidence level of the spatial autocorrelation analysis is determined by combining the P value and Z score. The P value indicates the probability of spatial autocorrelation, P value is used to judge the significance of

TABLE II
GLOBAL MORAN INDEX CONFIDENCE PARTITION METHOD

Z Score	P value	Confidence degree
<-1.65 or $>+1.65$	<0.10	90%
<-1.96 or $>+1.96$	<0.05	95%
<-2.58 or $>+2.58$	<0.01	99%

the spatial pattern. When the P value is small, the observed spatial pattern is unlikely to arise from a random process (small probability event), then the null hypothesis can be rejected. The Z score refers to the standard deviation multiplier, it reflects the degree of data dispersion. Table II shows the confidence levels of the global Moran index corresponding to the Z score and P value.

Local spatial autocorrelation is used to calculate the degree of spatial correlation between each pixel value and their neighboring pixels, reflecting the spatial heterogeneity and instability in the local region. The main equations of the local Moran index are listed as follows:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^{n} w_{i,j} \left(x_j - \bar{X}\right) \quad (6)$$

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^{n} \left(x_j - \bar{X}\right)^2}{n - 1} \quad (7)$$

$$Z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \quad (8)$$

$$E[I_i] = -\frac{\sum_{j=1, j \neq i}^{n} w_{i,j}}{n - 1} \quad (9)$$

$$V[I_i] = E[I_i^2] - E[I_i]^2. \quad (10)$$

The meaning of each parameter is the same as the global Moran index, except that each parameter here refers to the value of a specific pixel. Unlike the global Moran index, the Moran index value, Z score, and P value of each NDVI image element are obtained after the local Moran index is calculated. Based on the Z score, Moran index value and different spatial weights, each NDVI image can be classified into five types of clusters: High–high, high–low, low–high, low–low, and not significant. The specific division method is shown in Table III.

In this study, the local Moran index of each NDVI image was taken as an input feature into the model to enhance the accuracy of the model. In addition, the study area contains a variety of land use types, so the LULC data were added as an input feature at the same time since vegetation cover changes are not consistent across land use types.
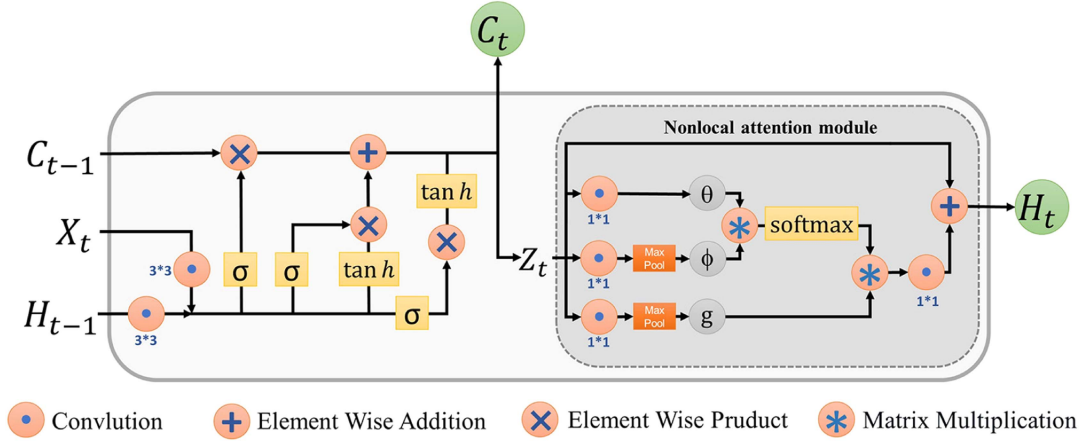
Fig. 2. Cell structure of ConvLSTM-NL.

TABLE III
SPATIAL CLUSTERING TYPE CLASSIFICATION METHOD

Z_i	$\sum\limits_{j \neq i}^{n} w_{i,j} Z_j$	I_i	Meaning
>0	>0	>0	High–high cluster
<0	<0	>0	Low–low cluster
<0	>0	<0	Low–high cluster
>0	<0	<0	High–low cluster

B. ConvLSTM With Nonlocal Attention Module (ConvLSTM-NL)

Nonlocal attention module is a neural network module for image or video processing in computer vision. The main role of the nonlocal attention module is to improve the model's ability to capture long-range dependence in images or videos by considering global contextual information. The nonlocal operations are added into the ConvLSTM model as a generic building block to capture long-range dependence and take the global information into account. The addition of the nonlocal attention module not only preserves the ability of ConvLSTM to extract local information but also learns global information appropriately, thus providing better performance in handling long-range dependence. In this study, the nonlocal attention module is placed after the ConvLSTM structure to effectively handle the hidden states.

The structure of ConvLSTM-NL is shown in Fig. 2. Each structural cell has three inputs: the current input feature X_t, the hidden output H_{t-1}, and the cell state C_{t-1} of the previous time step. Each structural unit consists of two parts: the original ConvLSTM structure and the nonlocal attention module.

The core role of ConvLSTM is to replace the fully connected layer in LSTM with a convolutional layer so that the memory unit and gate structure of LSTM can be combined with the convolutional and pooling layers in CNN to model spatiotemporal information. The spatiotemporal features of the multichannel image sequence are extracted by convolution operations. Then, the information in the sequence is modeled and passed through

LSTM units. The main calculation equations included in the structure of the first part are

$$i_t = \sigma \left(W_{xi} \odot X_t + W_{hi} \odot H_{t-1} + b_i \right) \tag{11}$$

$$f_t = \sigma \left(W_{xf} \odot X_t + W_{hf} \odot H_{t-1} + b_f \right) \tag{12}$$

$$C_t = f_t C_{t-1} + i_t \tanh(W_{xc} \odot X_t + W_{hc} \odot H_{t-1} + b_c) \tag{13}$$

$$Z_t = \tanh \left(C_t \right) \sigma \left(W_{xo} \odot X_t + W_{ho} \odot H_{t-1} + b_o \right) \tag{14}$$

where $\odot$ indicates the convolution operation, σ indicates the sigmoid function. X_t is the input to the neuron at moment t. H_{t-1} is the output of ConvLSTM-NL at moment $t-1$. W_{xi}, W_{xf}, W_{xc}, and W_{xo} are the weights of X_t in different convolution operations, respectively, W_{hi}, W_{hf}, W_{hc}, and W_{ho} are the weights of H_{t-1}. b_i, b_f, b_c, and b_o are the bias values of the convolution operation. i_t controls the input gate. The input gate uses the sigmoid function to select the new input data and the output of the previous neuron to decide the input information of that neuron. f_t controls the forgotten gate. f_t determines how much information from the previous neuron is retained in this neuron. C_t is the cell state of the neuron at moment t. C_t will be output by combining the cell state of the previous neuron and the new data input at moment t of the neuron. Finally, there is an output gate, which is set by gating to output the current cell state proportionally to Z_t.

The second part is the nonlocal attention module. In this study, a 1×1 convolution is used to linearly map the output Z_t from the ConvLSTM part and compress the number of channels to obtain the features θ, φ, and g. After combining the dimensions of the above-mentioned three features except the channel dimension by reshape operation, matrix dot product operations are needed to obtain the autocorrelation in class features. Then, the attention coefficients are multiplied correspondingly back into the feature matrix g. Finally, the 1×1 convolution is used again to expand the channel number and the residual operation with the original Z_t is performed to obtain the final output H_t of ConvLSTM-NL. To reduce the amount of computation, the number of channels in θ, φ, and g is set to half the number of channels in Z_t. The
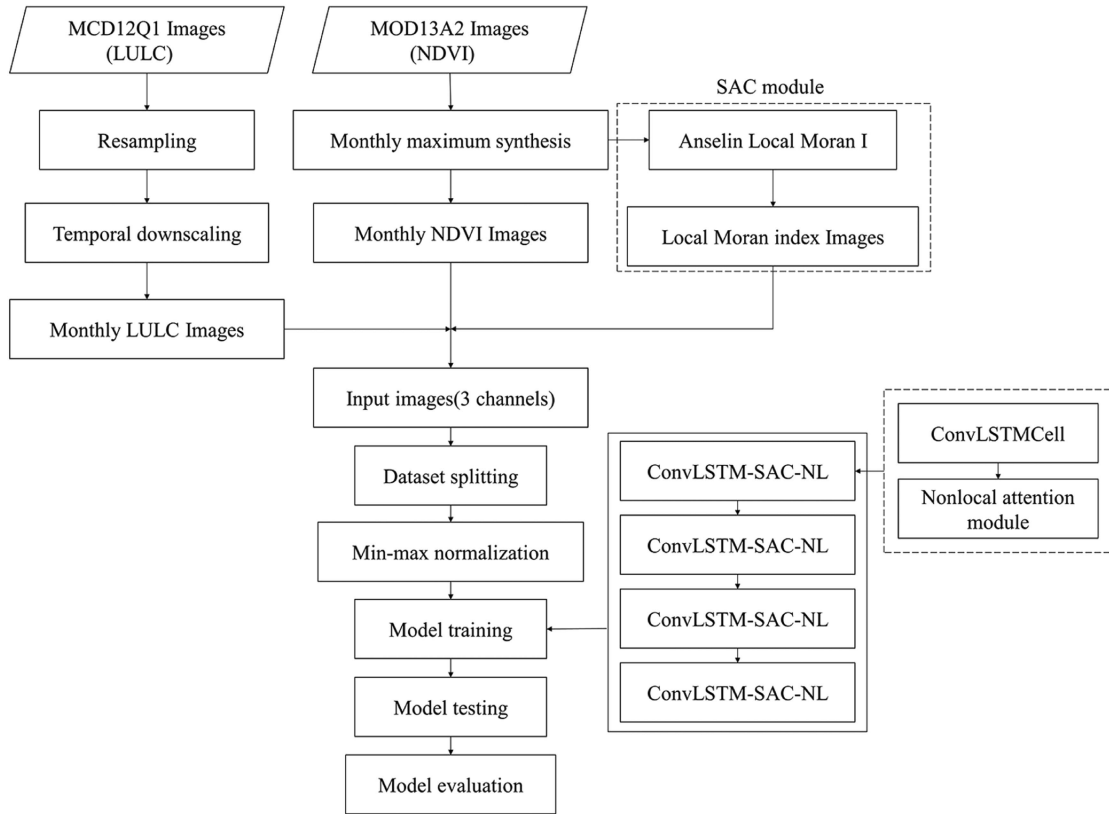
Fig. 3. Proposed ConvLSTM-SAC-NL flowchart.

main calculation formulas used are as follows:

$$y_i = \frac{1}{\mathcal{C}(z)} \sum_{\forall j} f(z_i, z_j) * g(z_j) \qquad (15)$$

$$f(z_i, z_j) = \text{softmax}\left(\theta(z_i)^T * \phi(z_j)\right) \qquad (16)$$

$$H_t = W_i \odot y_i + Z_t \qquad (17)$$

where $*$ indicates matrix multiplication. i indicates the index of the output location, such as space, time, or spacetime, whose response should be enumerated for j and then computed. f functional equation calculates the similarity of the pixel values at position i and j, then obtains the value of the weight with the value range [0, 1] by the softmax function, which is the desired attention factor. $\mathcal{C}(z) = N$, N is the number of pixels, which is used for normalization. W_i is a 1×1 convolution of y_i to expand the channel number, so that the channel number is consistent with the input data. H_t is the final output.

C. Model Training and Evaluation

In this study, seven models, i.e., RF, SVR, LSTM, bidirectional long short-term memory (BiLSTM), graph convolutional network (GCN), predictive recurrent neural network (PredRNN), and ConvLSTM were used as the baseline models for our experiments. The baseline experiments used NDVI as the input data directly.

The flow of ConvLSTM-SAC-NL is shown in Fig. 3. We used NDVI, local Moran index, and LULC data as input features. The 192 NDVI data were divided into the training set, validation set, and test set according to the ratio of 6:1:1. The min–max normalization operation was conducted on the LULC and local Moran index before data input to eliminate the prediction error caused by the gauge inconsistency problem. The normalized formula is

$$X_i' = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \qquad (18)$$

where X_i indicates the NDVI value of pixel i, $X_{\min}$ and $X_{\max}$ indicate the minimum and the maximum value of the NDVI images in the training set, respectively.

Both the ConvLSTM-SAC-NL model and the ConvLSTM model had a four-layer model structure and used the same hyperparameters, where the batch size, learning rate, epoch, and the number of convolution kernels were set to 3, 0.0005, 100, and 16, respectively. Through experiments, the loss curve converged after 100 iterations and the rest of the hyperparameter settings were optimal for both single- and multiple-feature prediction.

To evaluate model accuracy, four metrics are chosen to judge the prediction results of each model, including MAE, RMSE, coefficient of determination (R^2), and Pearson correlation coefficient to test the prediction results of each model. MAE can be used to measure the average degree of difference between the predicted pixel values and the true pixel values. RMSE is used to measure the variance between the predicted pixel values and the

true pixel values. R^2 and correlation coefficient can measure the squared and the linear correlation between the predicted pixel values and the true pixel values, respectively. The relevant formulas for the calculation of evaluation indicators are

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

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{20}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{21}$$

$$\text{Correlation} = \frac{\sum_{i=1}^{n} (y_i - \bar{y}) (\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2}} \tag{22}$$

where n is the number of the predicted pixels, y_i and $\hat{y}_i$ indicate the true value and the predicted value, respectively, and $\bar{y}$ and $\bar{\hat{y}}$ indicate the mean of the true values and the predicted value, respectively.

In addition, this study used the floating-point operations per second (FLOPS) and parameter count (Params) to measure the computational complexity of each deep learning model. The FLOPS is a commonly used unit for measuring computational complexity, which can represent the number of floating-point operations a computer can perform in one second. A higher FLOPS value indicates that the system or algorithm can execute computational tasks faster, thereby improving computational efficiency and speed. The Params represents the number of learnable parameters in the model. The size of the Params directly affects the complexity and storage requirements of the neural network. A smaller Params indicates lower computational demands and storage costs for the model.

IV. EXPERIMENTAL RESULT

A. Spatial Autocorrelation Analysis

After calculation, the global Moran index of all the NDVI images is above 0.9, and the Z scores are all greater than 2.58 with P values less than 0.01, which indicates the significance of the Moran index. These results are sufficient to demonstrate that the NDVI data have obvious spatial aggregation characteristics in Huangpi district.

Fig. 4 illustrates the spatial clustering types of the mean NDVI images in Huangpi district in 2018, with five clustering types, mainly high–high cluster and low–low cluster. The high–high cluster is mainly concentrated in the northern part of Huangpi district, while the low–low cluster is mainly distributed in the southern part. The not significant, high–low cluster, and low–high cluster types are scattered and account for a small fraction of the area.

The spatial distributions of NDVI clustering types are closely related to the elevation and land use in the Huangpi district. The northern part of Huangpi district has lower population density, higher elevation, and more lush vegetation compared to the southern part. Therefore, the high-value clustering prevailed in
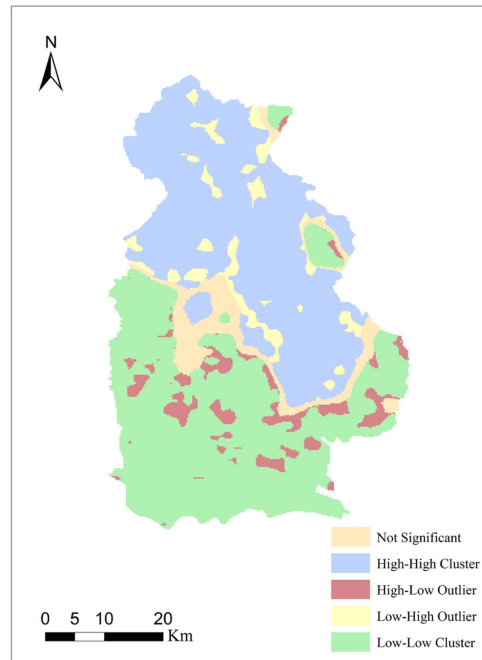

Fig. 4. Spatial aggregation of monthly mean NDVI in year 2018 for the Huangpi district.

TABLE IV
COMPARISON OF THE OVERALL ACCURACY OF EACH MODEL

Models	RMSE	MAE	R^2	Correlation
SVR	0.084	0.064	0.777	0.889
RF	0.088	0.064	0.758	0.887
LSTM	0.091	0.069	0.741	0.864
BiLSTM	0.084	0.062	0.776	0.885
GCN	0.077	0.060	0.804	0.899
PredRNN	0.074	0.056	0.829	0.912
ConvLSTM	0.076	0.058	0.815	0.925
ConvLSTM-SAC-NL	**0.061**	**0.045**	**0.881**	**0.941**

the northern part due to dense, while the southern part has more low-value clustering.

B. Overall Accuracy Evaluation

The overall NDVI forecasting accuracy is shown in Table IV. The SVR model outperforms the RF model in one-month lead NDVI forecasting in terms of four evaluation metrics. The LSTM model has lower predictive accuracy relative to RF and SVR, which is possibly related to insufficient data as the training of LSTM requires a sufficient number of time series (144 images used in this study). In addition, LSTM suffers from gradient disappearance and explosion problems, and thus the unsatisfactory results are expected versus RF and SVR methods. The BiLSTM model learns information flow from both past and future directions and generates relatively good results. The PredRNN model is capable of effectively capturing long-term dependence in spatiotemporal sequence data, demonstrating better forecasting accuracy. The GCN model could utilize the connections between nodes for information propagation and feature learning. However, relying solely on local node information transmission

TABLE V
COMPARISON OF THE ACCURACY OF ABLATION EXPERIMENTS

Models	Inputs	RMSE	MAE	R^2	Correlation
ConvLSTM	NDVI	0.076	0.058	0.815	0.925
ConvLSTM-NL	NDVI	0.063	0.048	0.873	0.935
ConvLSTM-SAC	NDVI, LULC, local Moran index	0.066	0.050	0.860	0.931
ConvLSTM-SAC-NL	**NDVI, LULC, local Moran index**	**0.061**	**0.045**	**0.881**	**0.941**

may overlook the capture of global graph information, leading to limited predictive performance. The ConvLSTM model could capture both temporal and spatial information by LSTM and convolutional kernel window, respectively. However, the spatial dependence is uniformly considered in ConvLSTM by convolutional kernels, ignoring the spatial variations of dependence structure for different locations and variables.

The R^2 of ConvLSTM-SAC-NL is 0.881, higher than that of ConvLSTM (0.815), indicating the superiority of the proposed model by adding nonlocal attention and spatial autocorrelation modules. A two-part ablation experiment is conducted to verify the effect of each part on the prediction accuracy in ConvLSTM-SAC-NL. Table V shows the inclusion of the nonlocal attention module and the effect of considering spatial autocorrelation on the prediction results. After adding the nonlocal attention module, the ConvLSTM-NL model with NDVI as input demonstrates improved accuracy relative to ConvLSTM, with an increase of R^2 by 0.058 and a decrease of RMSE by more than 17%. Moreover, the R^2 metric improves by 0.021 for ConvLSTM-SAC-NL versus ConvLSTM-SAC when the inputs are NDVI, LULC, and local Moran index. After adding LULC and the local Moran index as input data, the R^2 value for the ConvLSTM-SAC model improves by 0.045 relative to the ConvLSTM model with NDVI as input.has a great impact on the results because it not only considers the convolutional kernel window size pixel values but also extracts the global information. Even if the input data are only NDVI, the R^2 is still improved by 0.058, and both RMSE and MAE are reduced by more than 17%. Moreover, R^2 is also improved by 0.021 when the input data are NDVI, LULC, and the local Moran index, which may be a limited increase in accuracy because the model has already learned the spatial aggregation characteristics of NDVI to some extent after increasing the input variables. After adding LULC and the local Moran index as input data, the prediction with the ConvLSTM model also improved R^2 by 0.045. Therefore, both the spatial autocorrelation and the nonlocal attention module play positive roles in enhancing prediction accuracy, and the combination of them demonstrates further added values than either one.

Fig. 5 shows the scatter plots of NDVI prediction results for each model. The SVR model exhibits some dispersions between predictions and observations for low-value NDVI samples, and the RF model indicates better prediction results over high-value NDVI samples than SVR. The LSTM model predicts poorly for some NDVI pixels with a range of 0.4–0.9, as the scatter plots exhibit obvious dispersion in that range. Overall, the ConvLSTM-SAC-NL model demonstrates better consistency between observations and predictions for both high and low NDVI samples. Fig. 6 shows the spatial comparison of prediction

results between ConvLSTM-SAC-NL and ConvLSTM. It can be seen that 91.49% of the pixels improved prediction accuracy for ConvLSTM-SAC-NL versus ConvLSTM in the study area. The unenhanced regions are sparsely distributed over the study area, indicating little improvement in predictive accuracy for ConvLSTM-SAC-NL versus ConvLSTM for these pixels. These unenhanced regions are probably related to a variety of factors, such as strong spatial heterogeneity and active human activities. To compare the single-point accuracy of each model more specifically, Fig. 7 shows the box plot of the prediction accuracy for all pixels using eight models. It can be seen that ConvLSTM-SAC-NL shows great superiority in NDVI predictions in terms of the mean, median, and maximum–minimum values of the evaluation metrics.

C. Spatial Prediction Accuracy Evaluation

Fig. 8 shows the observations and spatial prediction results of all models in each season (average value of three months). The RF and SVR models only perform sequential prediction for each point without considering the relationship between surrounding pixels. Thus, the predictions of the RF and SVR models for each season have an obvious blurring phenomenon compared to the observations. The prediction performance of the LSTM model in summer and autumn is significantly better than that in spring and winter, which indicates that the prediction of the LSTM model for high-value NDVI samples is better than that for low-value NDVI samples. Overall, the BiLSTM model performs better in NDVI prediction across all seasons versus LSTM because of the consideration of the bidirectional information. The PredRNN model shows poor prediction performance during spring and summer, especially in high-value areas in the northern region. The GCN model exhibits unsatisfactory predictive performance during summer. Since the ConvLSTM model focuses on spatial features, it ignores the transmission of temporal information. Although the training is sufficient, the ConvLSTM model still cannot guarantee the prediction performance in different time series data of NDVI.

The ConvLSTM-SAC-NL model effectively takes into account the long-range dependence of NDVI data and considers the spatial autocorrelation through the local Moran index and LULC data. The spatial prediction results of the ConvLSTM-SAC-NL model for all seasons are closer to the observations than other models, which is better than other models. The ConvLSTM-SAC-NL method has a better prediction performance for the regions that the rest of the models cannot accurately predict. However, there are still very small areas where the predictions
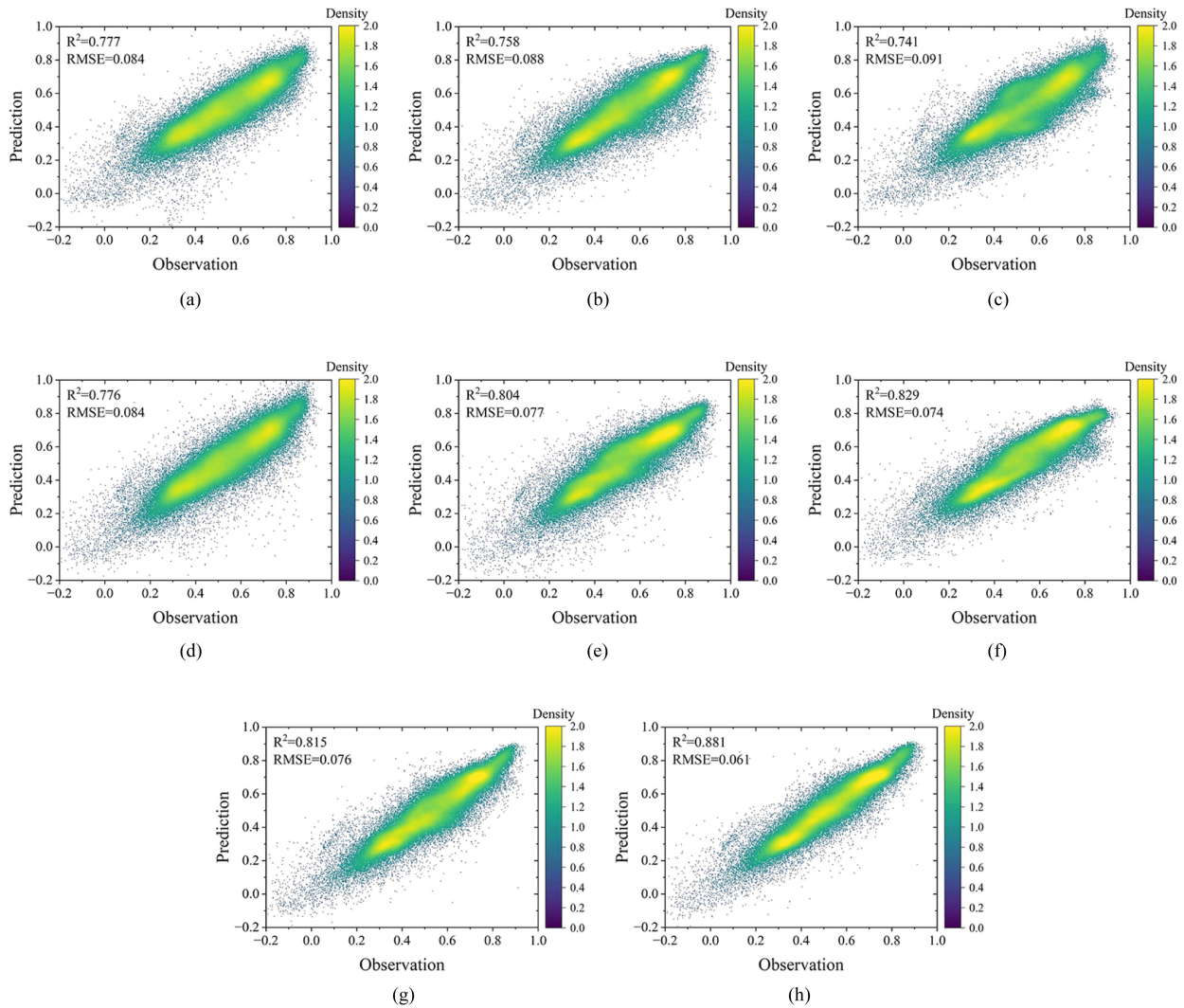
Fig. 5. Scatter plots of each models' predictive accuracy. (a)–(h) show the NDVI prediction accuracy for eight models, where the horizontal and vertical axes denote the observed and predicted values, respectively. (a) SVR. (b) RF. (c) LSTM. (d) BiLSTM. (e) GCN. (f) PredRNN. (g) ConvLSTM. (h) ConvLSTM-SAC-NL.

are not close to the observations, and the reasons for this can be further studied in the future.

D. Temporal Prediction Accuracy Evaluation

Fig. 9 shows the mean value time series line chart of each prediction image for 21 months of each model. The worst performance of temporal prediction is the LSTM model, although the LSTM model focuses on learning temporal features. The predicted mean value of LSTM is much larger than that of observations, especially for the period from November 2017 to March 2018. The temporal prediction line of the BiLSTM model is much better compared to LSTM, without any abrupt changes or anomalies. The PredRNN model performs slightly better, but it also exhibits anomalies in some months, such as May 2017. The GCN model has an overall lower mean prediction. The ConvLSTM model processes the surrounding pixels with convolutional kernels, its temporal change trend of the prediction results is better than that of the RF and SVR models. However,

the ConvLSTM model frequently has low-value NDVI prediction results. Because of the addition of the nonlocal attention module and the spatial autocorrelation, the predicted mean value of the ConvLSTM-SAC-NL model in each month is the closest to observations. For example, from July 2018 to October 2018, the mean change line of the ConvLSTM-SAC-NL model is closest to the observation line among all the models.

E. Computational Complexity Analysis

The FLOPS and Params are calculated for six deep learning models to measure the computational complexity of each prediction model. Table VI shows that the ConvLSTM model has relatively high FLOPS. Compared to the ConvLSTM model, the ConvLSTM-SAC-NL model only increases FLOPS by 0.15G and Params by 0.01M. This indicates that ConvLSTM-SAC-NL does not add excessive computational complexity, so the proposed model is highly feasible. Furthermore, it can be observed that GCN has the highest Params among six deep learning models due to the involvement of numerous matrix operations.
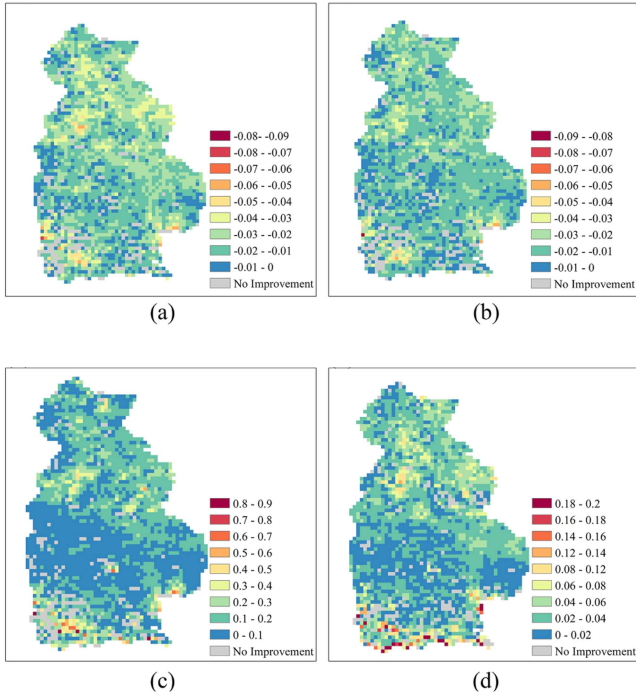
Fig. 6. Spatial maps of the differenced accuracy between ConvLSTM-SAC-NL and ConvLSTM. The difference is obtained by subtracting the evaluation metric for ConvLSTM-SAC-NL from that of ConvLSTM. (a)–(d) are the differenced results for RMSE, MAE, R^2, and correlation coefficient, respectively. (a) RMSE. (b) MAE. (c) R^2. (d) Correlation coefficient.
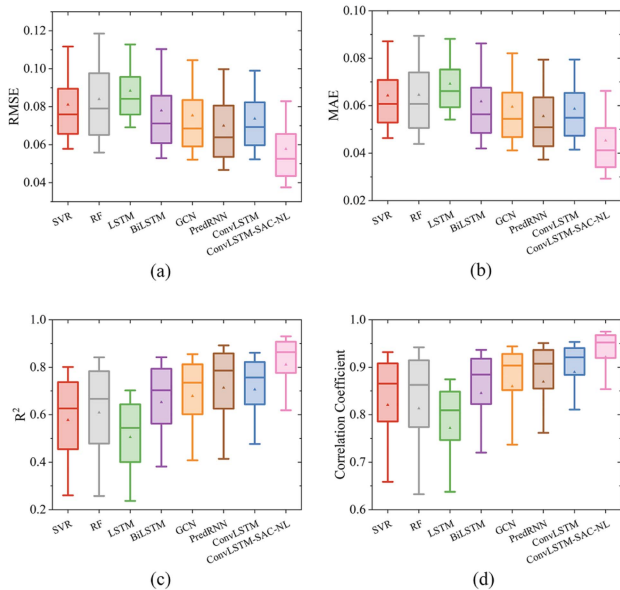

Fig. 7. Box plots of all pixel prediction accuracies of each model. The whiskers of the box plots show the single-point precision values with confidence intervals from 10% to 90%, the upper and lower quartiles of the boxes indicate the upper and lower quartiles, the horizontal line in the center of the box indicates the median of the single-point precision, and the triangle indicates the mean of the single-point precision. (a)–(d) are the box plots of the evaluation metrics RMSE, MAE, R^2, and correlation coefficient for each pixel of each model, respectively. (a) RMSE. (b) MAE. (c) R^2. (d) Correlation coefficient.

TABLE VI
COMPARISON OF THE COMPUTATIONAL COMPLEXITY

Model	FLOPS	Params	Runtime(h)
LSTM	122.21K	12.96K	7.70
BiLSTM	318.14K	34.11K	12.83
GCN	2.13G	469.01M	0.73
PredRNN	2.48G	59.65K	0.14
ConvLSTM	2.75G	19.12M	0.11
ConvLSTM-SAC-NL	2.9G	19.13M	0.19

The PredRNN model reduces the input and output dimensions through downsampling and upsampling strategies for handling time series data, resulting in a reduction in Params. Since the LSTM and the BiLSTM models perform pointwise predictions, each point requires individual training, so their runtimes are significantly higher compared to the other models.

V. DISCUSSION

This study evaluated the predictive capabilities of different models for NDVI. The data used in this study were the NDVI and LULC data of Huangpi district from MODIS, spanning from 2003 to 2018. We improved the traditional ConvLSTM model by adding the nonlocal attention module to capture the long-range dependence of NDVI. Considering the spatial autocorrelation of vegetation, the local Moran index, and LULC data were incorporated as input variables into the ConvLSTM-SAC-NL model for NDVI prediction. The study designed and trained seven contrast models that were previously applied to vegetation prediction, including LSTM, BiLSTM, PredRNN, GCN, and ConvLSTM.

Among these models, the ConvLSTM-SAC-NL model achieves the highest prediction accuracy using four metrics (MAE, RMSE, R^2, and correlation coefficient). Compared to the prediction results of the traditional ConvLSTM model, the ConvLSTM-SAC-NL model shows an improvement of 8.10% in R^2, and a reduction of 19.74% and 22.41% in MAE and RMSE, respectively. The ablation experiments confirm that the addition of both the nonlocal attention module and the local Moran index enhances the predictive capability of the ConvLSTM-SAC-NL model.

In this study, the LSTM model performs the worst in NDVI prediction because of the relatively limited training data, which can result in good performance of the training data but poor generalization on the testing data. The BiLSTM model is capable of learning feature representations from both directions in a sequence. The forward LSTM learns the historical information of the sequence, while the backward LSTM learns the future information. By combining the features from two directions, the BiLSTM model can provide a richer and comprehensive feature representation, reducing information loss and enabling more accurate NDVI predictions. The PredRNN model adopts a stacked recurrent layer structure, which effectively captures long-term dependence in time series. By introducing recursive hidden states to retain and update past information, along with
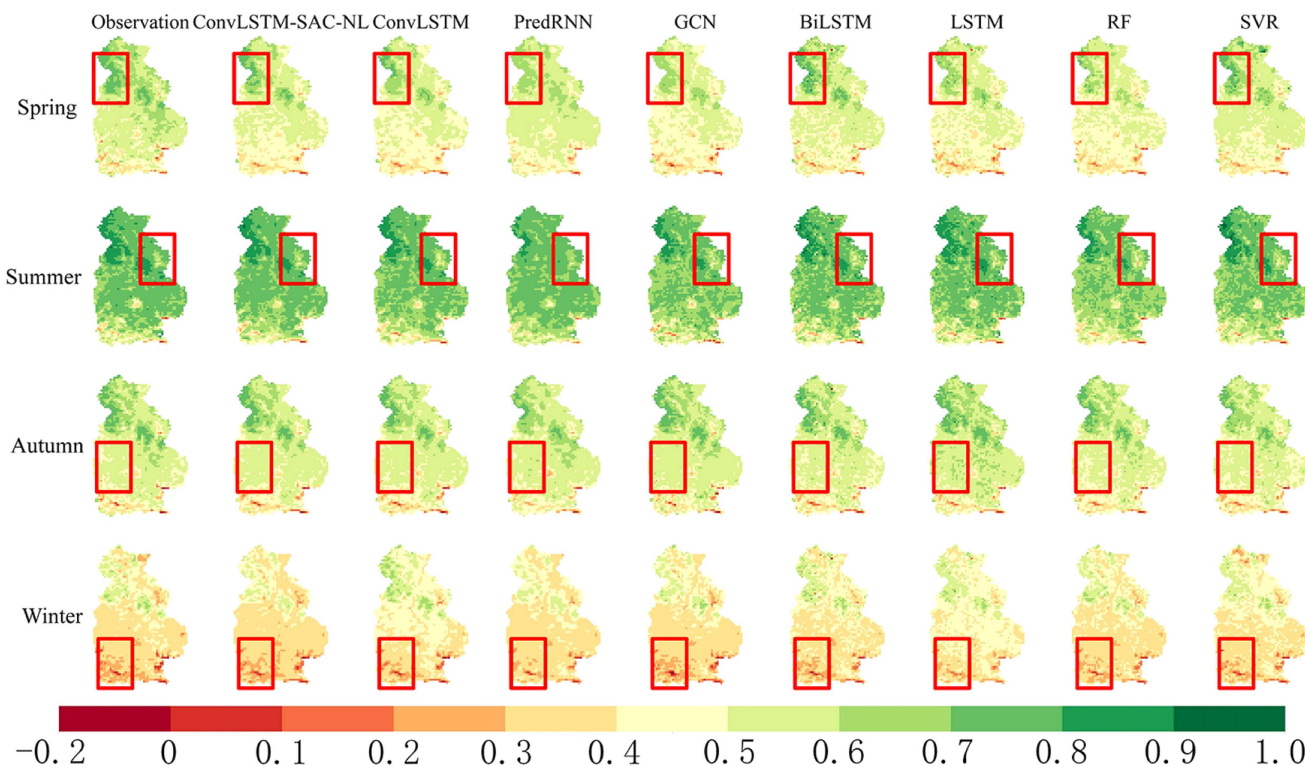
Fig. 8. Comparison between the predictions and the observations in each season, the red rectangle box is the area where the prediction effect has obvious difference.
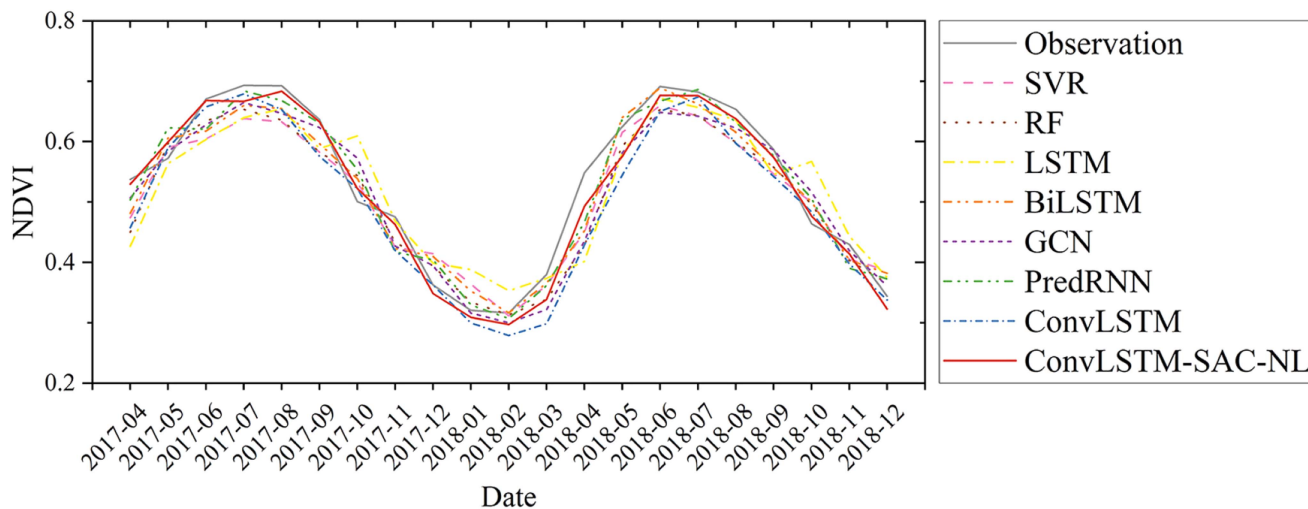

Fig. 9. Time series lines of the mean values of predictions of 21 months for each model.

the stacking of multiple recurrent layers, PredRNN layers can extract features at different time scales of NDVI.

The GCN, ConvLSTM, and ConvLSTM-SAC-NL models all consider spatial information. The GCN model utilizes the neighboring information of nodes for convolutional operations and updates node representations by aggregating features from nodes and their neighbors. This neighbor aggregation gives GCN a better understanding of semantics and context at a global level by obtaining information from surrounding nodes. However, the

GCN model may struggle to fully capture global information for NDVI data with long-range dependency relationships. The ConvLSTM model combines the structures of the CNN and the LSTM models, which makes the model capture dynamic changes and spatial correlations of input data in the spatiotemporal sequence data more effectively.

Compared with other existing models, the proposed ConvLSTM-SAC-NL model has two components contributing to improving NDVI prediction accuracy. First, the nonlocal

attention module considers the global information of the study area. The idea of nonlocal processing allows the model to use several different regions to integrate spectral and spatial information. The max-pooling operation in the nonlocal attention module reduces the computational effort of the model and improves the training efficiency. Second, the ConvLSTM-SAC-NL model learns the spatial aggregation characteristics of NDVI by adding the local Moran index and LULC data as features. This will provide a new direction for future monthly scale NDVI prediction research.

However, the ConvLSTM-SAC-NL model also has certain limitations. For example, the ConvLSTM-SAC-NL model could not improve the prediction accuracy of each pixel versus the traditional ConvLSTM model. The reason may be the inadequacy of the training data and the poor generalization ability of the model. In the future, it is important to consider how to learn the temporal trends of NDVI better.

VI. Conclusion

This study proposed a new NDVI prediction method called ConvLSTM-SAC-NL that combines the ConvLSTM model, the nonlocal attention module, and the spatial autocorrelation of NDVI. We used controlled experiment and ablation experiment to verify the advantages of the ConvLSTM-SAC-NL prediction method, by overall accuracy, single-point accuracy, temporal accuracy, spatial accuracy, and computational complexity. The main conclusions obtained are as follows. 1) The proposed ConvLSTM-SAC-NL model availably handles the problem of ignoring spatial autocorrelation in NDVI prediction research by incorporating the local Moran index into the ConvLSTM model. 2) The ConvLSTM-SAC-NL model extracts the long-range dependence to overcome the limitation of convolutional kernels with the nonlocal attention module.

Compared with the seven classical models, the NDVI prediction accuracy of the ConvLSTM-SAC-NL model is the best. The proposed ConvLSTM-SAC-NL model can be widely used in monthly NDVI spatiotemporal prediction. At the same time, the spatial autocorrelation and the nonlocal attention module provide a new idea for the spatiotemporal prediction of vegetation conditions.

References

[1] W. Gao et al., "NDVI-based vegetation dynamics and their responses to climate change and human activities from 1982 to 2020: A case study in the MU US sandy land, China," *Ecological Indicators*, vol. 137, 2022, Art. no. 108745.

[2] W. Liang, Q. Quan, B. Wu, and S. Mo, "Response of vegetation dynamics in the three-north region of China to climate and human activities from 1982 to 2018," *Sustainability*, vol. 15, no. 4, 2023, Art. no. 3073.

[3] X. Tong et al., "Quantifying the effectiveness of ecological restoration projects on long-term vegetation dynamics in the karst regions of Southwest China," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 54, pp. 105–113, 2017.

[4] N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth, "Using the satellite-derived NDVI to assess ecological responses to environmental change," *Trends Ecol. Evol.*, vol. 20, no. 9, pp. 503–510, 2005.

[5] A. Ferchichi, A. B. Abbes, V. Barra, and I. R. Farah, "Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review," *Ecological Informat.*, vol. 68, 2022, Art. no. 101552.

[6] K. Choudhary, W. Shi, M. S. Boori, and S. Corgne, "Agriculture phenology monitoring using NDVI time series based on remote sensing satellites: A case study of Guangdong, China," *Opt. Memory Neural Netw.*, vol. 28, pp. 204–214, 2019.

[7] X. Hu et al., "Agricultural drought monitoring using european space agency sentinel 3A land surface temperature and normalized difference vegetation index imageries," *Agricultural Forest Meteorol.*, vol. 279, 2019, Art. no. 107707.

[8] M. H. Afshar, A. Al-Yaari, and M. T. Yilmaz, "Comparative evaluation of microwave l-band VOD and optical NDVI for agriculture drought detection over central Europe," *Remote Sens.*, vol. 13, no. 7, 2021, Art. no. 1251.

[9] S. Baeza and J. M. Paruelo, "Land use/land cover change (2000–2014) in the Rio de la Plata grasslands: An analysis based on modis NDVI time series," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 381.

[10] Y. Ju and G. Bohrer, "Classification of wetland vegetation based on NDVI time series from the HLS dataset," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 2107.

[11] A. Fernández-Manso, C. Quintano, and O. Fernández-Manso, "Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale," *Int. J. Remote Sens.*, vol. 32, no. 6, pp. 1595–1617, 2011.

[12] M. S. Omar and H. Kawamukai, "Prediction of NDVI using the holt-winters model in high and low vegetation regions: A case study of East Africa," *Sci. Afr.*, vol. 14, 2021, Art. no. e01020.

[13] L. Xu, N. Chen, X. Zhang, H. Moradkhani, C. Zhang, and C. Hu, "In-situ and triple-collocation based evaluations of eight global root zone soil moisture products," *Remote Sens. Environ.*, vol. 254, 2021, Art. no. 112248.

[14] S. Huang, J. Chang, Q. Huang, and Y. Chen, "Monthly streamflow prediction using modified EMD-based support vector machine," *J. Hydrol.*, vol. 511, pp. 764–775, 2014.

[15] C. Tong, H. Wang, R. Magagi, K. Goïta, and K. Wang, "Spatial gap-filling of SMAP soil moisture pixels over tibetan plateau via machine learning versus geostatistics," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9899–9912, 2021.

[16] L. Xu, N. Chen, Z. Chen, C. Zhang, and H. Yu, "Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions," *Earth- Sci. Rev.*, vol. 222, 2021, Art. no. 103828.

[17] B. Roy, "Optimum machine learning algorithm selection for forecasting vegetation indices: Modis NDVI & EVI," *Remote Sens. Appl.: Soc., Environ.*, vol. 23, 2021, Art. no. 100582.

[18] S. Zhou, L. Xu, and N. Chen, "Rice yield prediction in Hubei province based on deep learning and the effect of spatial heterogeneity," *Remote Sens.*, vol. 15, no. 5, 2023, Art. no. 1361.

[19] X. Geng et al., "Potential of ANN for prolonging remote sensing-based soil moisture products for long-term time series analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2503205.

[20] N. Efremova, M. E. A. Seddik, and E. Erten, "Soil moisture estimation using sentinel-1/-2 imagery coupled with cycleGAN for time-series gap filing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4705111.

[21] P. Gao, W. Du, Q. Lei, J. Li, S. Zhang, and N. Li, "NDVI forecasting model based on the combination of time series decomposition and CNN–LSTM," *Water Resour. Manage.*, vol. 37, no. 4, pp. 1481–1497, 2023.

[22] C. Cui, W. Zhang, Z. Hong, and L. Meng, "Forecasting NDVI in multiple complex areas using neural network techniques combined feature engineering," *Int. J. Digit. Earth*, vol. 13, no. 12, pp. 1733–1749, 2020.

[23] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021.

[24] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

[25] W. He, Y. Wu, X. Li, and Q. Le, "Attention mechanism for neural machine translation: A surve," in *Proc. IEEE 5th Inf. Technol., Netw., Electron. Automat. Control Conf.*, Xi'an, China, 2021, pp. 1485–1489, doi: 10.1109/ITNEC52019.2021.9586824.

[26] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, pp. 4263–4270.

[27] T. Yoshimura, T. Hayashi, K. Takeda, and S. Watanabe, "End-to-end automatic speech recognition integrated with ctc-based voice activity detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Barcelona, Spain, 2020, pp. 6999–7003, doi: 10.1109/ICASSP40776.2020.9054358.

[28] W. Cai and Z. Wei, "Remote sensing image classification based on a cross-attention mechanism and graph convolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8002005.

[29] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5604816.

[30] R. Liu et al., "RAANet: A residual ASPP with attention framework for semantic segmentation of high-resolution remote sensing images," *Remote Sens.*, vol. 14, no. 13, 2022, Art. no. 3109.

[31] M. Beyer, R. Ahmad, B. Yang, and P. Rodríguez-Bocca, "Deep spatial-temporal graph modeling for efficient NDVI forecasting," *Smart Agricultural Technol.*, vol. 4, 2023, Art. no. 100172.

[32] W. Yu et al., "Spatial–temporal prediction of vegetation index with deep recurrent neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2501105.

[33] B. Huang, Z. Wang, J. Shang, G. Chen, and M. Radenkovic, "A spectral sequence-based nonlocal long short-term memory network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3041–3051, 2022.

Wenying Du received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017.

She is currently an Associate Professor with the National Engineering Research Center of Geographic Information System, China University of Geosciences (Wuhan), Wuhan, China. She has authored or coauthored many papers in related journals include *Building and Environment*, *Computers and Electronics in Agriculture*, *GIScience and Remote Sensing*, etc. Her research interests include flood process monitoring, flood process analysis, and flood knowledge map construction.

Lei Xu received the B.Eng. degree in surveying and mapping engineering from the School of Geodesy and Geomatics, Wuhan University, Wuhan, China, in 2016, and the Ph.D. degree in cartography and geographic information system from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, in 2021.

He is currently the Distinguished Professor of the National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan, China. He has authored or coauthored many papers in related journals include *Earth-Science Reviews*, *Remote Sensing of Environment*, *Geophysical Research Letters*, *Water Resources Research*, *Journal of Geophysical Research: Atmospheres*, etc. His research interests include geographic spatiotemporal prediction, spatiotemporal big data analysis, uncertainty modeling, remote sensing big data, and hydrometeorological remote sensing.

Zeqiang Chen received the Ph.D. degree in geographic information science from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, in 2012.

He is currently a Professor with the National Engineering Research Centre for Geographic Information Systems, China University of Geosciences (Wuhan), Wuhan, China. He has authored or coauthored many papers in related journals include *Remote Sensing of Environment*, *Atmospheric research*, *Water Resources Research*, *Earth System Science Data*, etc. His research interests include sensor networks research, spatiotemporal big data intelligence and application in smart cities.

Ruinan Cai received the B.Eng. degree in remote sensing science and technology from the School of Surveying and Mapping, Henan Polytechnic University, Jiaozuo, China, in 2022. She is currently working toward the master's degree in resources and environment from the National Engineering Research Center of Geographic Information System, China University of Geosciences (Wuhan), Wuhan, China.

Her research interests include environmental remote sensing, geotemporal prediction, and deep learning.

Nengcheng Chen received the B.Eng. degree in surveying and geomatics engineering and the M.Eng. degree in geospatial information engineering from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1997 and 2000, respectively, and the Ph.D. degree in geospatial information engineering from Wuhan University, Wuhan, China, in 2003.

From 2004 to 2008, he was an Associate Professor, and from 2008 to 2021, he was a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. Since 2021, he has been the Director of the National Engineering Research Center of Geographic Information System and the Vice President of the School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan, China. He applies for more than 60 invention patents and software, and authored or coauthored six monograph, four international standard, and more than 200 academic papers. His research interests include digital twins, Earth observation sensor web, spatiotemporal big data intelligence, geographic simulation decisions, smart city, and smart watershed.

Dr. Chen has presided over national key research and development plans, national natural fund projects, etc. He was the recipient of many awards, such as the American Science and Technology R&D Innovation Award and National Science and Technology Progress Innovation Team Award.

Hongchu Yu received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2020.

She is currently an Associate Researcher with the School of Navigation, Wuhan University of Technology, Wuhan, China. She was a Lecturer with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, China. She has authored or coauthored many papers in related journals include *Applied Energy*, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, *Transportation Research Part D -Transport and Environment*, etc. Her research interests include maritime data modeling and forecasting, traffic geography, spatial optimization, shipping network dynamic modeling and forecasting, and ship behavior modeling.