WILEY | Hindawi

Research Article

A Data-Driven Method and Hybrid Deep Learning Model for Flood Risk Prediction

Chenmin Ni [ID],[1,2] **Pei Shan Fam** [ID],[2] **and Muhammad Fadhil Marsani** [ID][2]

[1]*School of International Business, Zhejiang Yuexiu University, Shaoxing 312000, China*
[2]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia*

Correspondence should be addressed to Pei Shan Fam; fpeishan@usm.my

Flood disasters occur worldwide, and flood risk prediction is conducive to protecting human life and property safety. Influenced by topographic changes and rainfall, the water level fluctuates randomly and violently during the flood, introducing many noises and directly increasing the difficulty of flood prediction. A data-driven flood forecasting method is proposed based on data preprocessing and a two-layer BiLSTM-Attention network to improve forecast accuracy. First, the Variational Mode Decomposition (VMD) is used to decompose the data for reducing noise and produce suitable Intrinsic Mode Functions (IMFs); Then, an optimized two-layer attention-based Bidirectional Long Sshort-Term memory (BiLSTM-Attention) network is constructed to predict each IMF. Finally, two optimization algorithms are used to obtain the optimized parameters of VMD and BiLSTM intelligently, increasing the self-adaptability. The inertia factor of particle swarm optimization is improved and then used to optimize the five hyperparameters of BiLSTM. The proposed model reduces storage errors for smaller training sets and can achieve good performance. Three water level data sets from the Yangtze River in China are used for comparative experiments. Numerical results show that the peak height absolute error is within 2 cm, and the relative error of peak time arrival is within 30%. Compared with LSTM, BiLSTM, CNN-BiLSTM-attention, etc., the proposed model reduces the root mean square error by at least 50% and has advantages for high-risk forecasting when the water level exceeds the defense line and fluctuates prominently.

1. Introduction

Floods are remarkably destructive disasters and are highly harmful to human beings. Floods frequently occur almost worldwide, causing landslides and mudslides, which usually lead to many deaths and enormous property losses [1, 2]. The percentage of personnel losses caused by natural disasters in Asia is approximately 90%, usually attributed to floods [3]. Flood forecasting can help predict the next flood and inform people to prevent and reduce life and economic losses during floods [4].

Flood prediction systems are usually constructed based on studying the causes of floods and regional hydrological characteristics. Climate complexity, rainfall variability, and topographic features like topography and land cover significantly impact water levels during floods [5–7]. This makes it challenging to predict flood risks accurately. As

many river floods are related to the rapid rise of water levels, researchers attempt to predict water level changes by constructing statistical models of hydrological processes. Many studies are based on meteorological and hydrological characteristics through satellite observation data [4] with the help of GIS systems [8] combined with hydrology, hydrodynamics [9], and other knowledge [10, 11] to establish a flood forecasting system, simulate the process from rainfall to flood formation, and predict water level. This kind of method is remarkably representative, such as the Xin'anjiang model [12], but strongly depends on geographical coordinates and features and requires a large number of geological and water level monitoring data to build the model. Due to the difficulty of data collection and privacy protection, it is challenging for the model to be shared for research. Other prediction research combines statistical methods with other algorithms and predicts with geographic

and climatic characteristic data, such as in [13]. In the literature [14], authors considered the prediction method of combining time series with data topological spatial features and testing with water level data from the Kelantan River in Malaysia, further proving that the prediction effect is improved after the consideration of topological characteristics. In [15], taking Jialu River in Zhengzhou, China, as the research area, a model combining wavelet analysis and Markov is constructed, and it shows that the multiscale flood prediction model can get higher prediction accuracy. In [16], the authors considered all historical alarm flood sequences, established the relationship between the current and upcoming alarms, and constructed an optimization problem to realize water level prediction. Due to the multitude of data factors associated with floods, traditional models may experience a significant decrease in processing speed and capacity when faced with larger datasets. Therefore, achieving accurate predictions with minimal data is crucial in evaluating prediction models.

In recent years, the widespread application of machine learning and artificial intelligence in flood prediction has greatly enhanced its capabilities. In [17], machine learning and hydrodynamic models were integrated for flood hazard mapping, and the authors calculated the required emergency aid considering satellite-derived flood data. In [18], the authors proposed an IoT-based flood state prediction model and used IoT architecture and three machine-learning algorithms to promote the flood data acquisition process. The experimental results show that introducing a machine-learning algorithm improves prediction accuracy. As an extension of machine learning, deep learning can effectively simulate the memory function of the human brain. It is a multilayer neural network. Classic neural network algorithms include Back Propagation (BP) [19], Artificial Neural Network (ANN) [20], Convolutional Neural Network (CNN) [21], Long Short-TermMemory neural network (LSTM) [22], and Recurrent Neural Network (RNN) [23]. BP, ANN, and CNN assume that input is an independent unit without context; however, for some serialized inputs with obvious context characteristics, the new output will depend on the previous input, and a network architecture with a certain "memory ability" is required. In order to endow the network with such memory, the RNN was proposed [23], but it is prone to gradient descent and gradient explosion problems during the data training process. LSTM is a special kind of RNN with a specific memory ability. This network improves the gradient disappearance problem that occurs over time in RNNs during backpropagation, which introduces an internal mechanism called "gating" and enhances the learning capability for long-time sequences. Deep learning algorithms have shown powerful learning capabilities in data prediction, but they are usually black-box models. Due to their opaque internal structure and parameter adjustment process, the network structure and parameter settings are challenging.

The water level data of the river are a dynamic serialized input, and the new output data will depend on the water level data of the previous period. Many researchers have used LSTM for flood prediction. In [24], Hu et al. developed an integrated LSTM and reduced order model framework combined with two data decomposition methods and improved flood prediction speed. Simple time series prediction easily ignores the spatial information of the data. In [25], the authors considered the rainfall and flow at different stations, combined with the latitude and longitude information of the target area. They used a spatiotemporal attention model integrated with LSTM to improve flood prediction performance further. In [26], the authors proposed a fusion model of Convolutional neural networks and LSTM (ConvLSTM) and showed that ConvLSTM is superior to traditional algorithms in flood arrival time and peak flow. In [27], the attention model was applied to the LSTM, and experiments at the Yangtze River site in China demonstrated its superior prediction effect. The water level prediction methods based on LSTM improve the prediction accuracy, but the disadvantage of LSTM lies in its transmission of information only from front to back in one direction and its inability to encode information transmitted from back to front. It is essential to find a prediction model that can capture the long-term dependencies and contextual information in time series data, effectively learn the time correlation of historical meteorological and hydrological data, learn the characteristics of two-way or even multidirectional information flow, and also consider both past and future information. The Bilstm network has bidirectional learning ability and performs better than LSTM in understanding forward and backward dependencies in data [28, 29]. So we use the Bilstm-attention network to build a hybrid model for flood risk prediction.

As the water level fluctuates immediately and violently during the flood and the data changes are random, the randomness and instability of water flow fluctuations increase prediction difficulty. To solve this problem, we first use Variational Mode Decomposition (VMD) [30] to decompose and smooth the data and improve the prediction accuracy. VMD was proposed in 2014 [30]. It can perform a Fourier transform on time series under a fixed variational framework and decompose the original signal into several single-component modal signals with different frequencies called Intrinsic Mode Functions (IMFs). VMD can obtain the set of each modal component and its middle frequency, forming a relatively smooth subsequence. Moreover, VMD essentially extends the classical Wiener filter to multiple adaptive frequency bands. It is also robust in processing nonstationary or noise signals because of the denoising effect of the Wiener filter. In [31], Li et al. applied VMD to composite bearing fault diagnosis and got a better extraction of fault features. The results indicated that VMD decomposition could remove the interference noise in the data and improve the robustness of the model. In order to reduce noise, Xu and Yan et al. [32, 33] applied VMD to flood data preprocessing. Experimental results showed that VMD decomposition could remove the interference noise in the water level data and improve the signal-to-noise ratio, further combining with the deep learning neural network algorithm to improve the flood prediction accuracy [33]. In [34], authors used VMD-LSTM to forecast the stock price index, finding that the VMD-based LSTM's performance

improved. However, the limitation of VMD lies in the advanced setting of two important parameters: the modal number K and the penalty factor α. If the right selection is not made, the decomposition result may be over-decomposed or under-decomposed, leading to a failed data reconstruction. The common methods are to use the center frequency observation method, that is, to determine the value by observing the center frequency under different values [35, 36]. However, this method is accidental and subjective and can only determine the number of modes and not the penalty parameter [37]. The artificial setting methods often cannot obtain the optimal parameter combination, which will lead to some data processing errors.

Based on the discussion above, we address the following research questions: (i) How do you develop a predictive model to address the increased difficulty of predicting water level fluctuations during floods? Achieve higher accuracy with minimal data utilization? (ii) Considering the continuity of water flow, how do you construct a deep learning model with multi-directional learning capabilities and identify an effective approach for optimizing parameter settings? (iii) To what extent does the noise in the water flow signal impact the accuracy of predictions? How can we identify methods to reduce noise and enhance the accuracy of predictions?

To address these questions, a mixed prediction framework, namely, SVIBA (Sparrow Search Algorithm (SSA)-VMD-Improved Particle Swarm Optimization (IPSO)-BiLSTM-attention) is proposed to improve the water level prediction performance. The main innovations and contributions are as follows:

(1) Add VMD to decompose and preprocess the data to remove noise and use the SSA algorithm to intelligently find the two parameters $[K, \alpha]$ of VMD to yield suitable modes (IMFs), which improves the prediction ability of the subsequent BiLSTM neural network.

(2) A two-layer BiLSTM neural network is constructed and the Relu activation function is used instead of the traditional sigmoid activation function, which reduces the gradient decline and gradient disappearance in the prediction process. The inertia factor of traditional particle swarm optimization is improved according to research data characteristics. Compared to other optimization methods, the Improved Particle Swarm Optimization algorithm (IPSO) demonstrates faster convergence speed, effectively avoids local minima, and achieves a better balance between convergence speed and local search accuracy. In response to the complexities involved in BiLSTM network parameter adjustment, the IPSO algorithm is adopted to intelligently optimize the selection of five parameters.

(3) Focus on July, the most flood-prone month of the Yangtze River. Traditional flood forecasting methods usually focus on studying and training multiyear data, mainly using periodicity and seasonality to obtain the data rules. However, multiyear data storage and error records may increase the forecasting difficulty. The proposed SVIBA highlights the prediction of the flood level change in the most likely flood month with less training data, which makes the prediction more targeted. Moreover, SVIBA supports multidimensional data input, increases the selection of characteristic variables, and better grasps the spatiotemporal data characteristics. Comparison tests on the Yangtze River show that the proposed SVIBA model has good stability and prediction advantages in flood prediction, especially for predicting flow peaks with significant fluctuation before and after the occurrence of a flood.

This innovative approach addresses the challenges of noise reduction and accurate forecasting, enabling more precise prediction of flood peak height and arrival time. The accurate flood forecasts, especially for short-term forecasts with less data, provide valuable insights for authorities and policymakers to develop effective flood protection measures, evacuation plans, and resource allocation strategies, ultimately helping to protect human life and property.

2. Study Area and Data

The research area is around Hankou Station of the Yangtze River in Wuhan, Hubei Province, China (as shown in Figure 1). The geographical location is $29°58'–31°22'$ N latitude and $113°41'–115°05'$ E longitude. Wuhan has a population of over 11.2 million people, making it the fifth-largest city in China and the most populous city in Hubei Province. The average elevation of Wuhan is approximately 23 meters, with the lowest point being 12.8 meters in the western Jiangxia District. The terrain in this region is relatively flat, with minimal changes in altitude, making it highly susceptible to flood disasters. The flat topography allows floods to spread easily, resulting in submergence of the area. In addition, as it is located in the middle reaches of the Yangtze River, Wuhan is at significant risk of flooding due to the convergence of the Yangtze and Hanjiang rivers within the city.

The primary causes of Yangtze River flooding are heavy rainfall in the basin and snowmelt from the upper reaches, combined with the narrow river channel and accumulation of sediment, which contribute to the increased water level leading to flooding. National departments and researchers have made a lot of preparations for flood prediction and control [38, 39]. In [40], the authors used four global climate models to simulate future daily flows in the upper reaches of the Yangtze River after the past millennium, and flood variability was explored to find better predictive models. In [41], a particle swarm-support vector machine (PSO-SVM) water level prediction model was constructed by combining particle swarm optimization and support vector machine, and it showed that the predictive ability of SVM after parameter optimization becomes stronger. In [42], the authors divided the 19 stations along the Yangtze River into 6 clusters by dynamic time warping and hierarchical clustering algorithm, used the LSTM and seasonal autoregressive

FIGURE 1: The study area.

integrated moving average model, and then constructed a multistation daily water level prediction model. Deep learning-related algorithms have played an important role in predicting floods in the Yangtze River Basin, and models with short-term prediction capabilities are more favored by modelers because they are less affected by data errors. As the flood season of the Yangtze River in Wuhan occurs from May to September each year, with the last dozen years of major floods occurring in July (as shown in Figure 2). Our objective is to create a scalable flood prediction model that can handle data input with varying dimensions, specifically targeting the flood-prone month of July. This model aims to achieve more accurate predictions with less data while enhancing its predictive capabilities through data preprocessing and parameter optimization.

From [43], three-dimensional data, namely, water level, flow volume, and water level difference from yesterday at the same time, were collected for nearly ten years from April 1, 2012, to September 30, 2020. The data from these months were taken every year as the observation data, sampling by the hour. Figure 2 shows the visualization of 39,528 data points of water level and flow volume, and Table 1 presents the statistical description of water level data in the past 9 years for Hankou station.

According to the government announcement, the defense and warning water levels at the Hankou station are 25 and 27.3 m, respectively (see Figure 3). The water level and water volume changes are random and complex when the water level exceeds the set line, thus increasing the risk and complicating the prediction.

In the last 9 years, the highest water level has been 28.77 m, which is 3.77 m above the defense water level; the standard deviation is approximately 2.9 m; and the fluctuation range of the water level is relatively large. The observation in Figure 2 shows that the rise and fall of water level has a certain periodicity in the past 9 years. The maximum water level value exceeding the warning line appears in July every year. Figure 1 shows that the peak value of water level and water volume in 2016, 2017, and 2020 is high, which markedly exceeds the warning water level. In 2016, 2017, and 2020, major floods in the Wuhan area caused considerable casualties and property losses, especially in 2016 and 2020. Take July, the most likely month for megafloods, as the target for prediction. 2184 data from April to June and 744 relevant data from July are used as the training and test sets, respectively. Luoshan is the upstream station of

Hankou of the Yangtze River, and its water level change can play an early warning role for the downstream. To verify the short-term prediction ability of the algorithm during the flood period, we take 744 data from Luoshan station in July 2020 when the flood occurred, and divide them into training sets and test sets at 7 : 3 for experimental comparison. Figure 3 shows the visualization of the three datasets.

3. Methods

The implementation of the SVIBA model is mainly divided into two parts. First, SSA [44] is used intelligently to obtain the optimized parameters of VMD and enhance its noise removal capability. The decomposed IMFs are obtained using SSA-VMD; then for each IMF, BiLSTM is taken as the primary prediction model that can learn the information before and after the data [45]. We further improve the particle swarm optimization algorithm [46] and use this improved method to optimize the hyperparameters of BiLSTM. Finally, the weight attention mechanism in BiLSTM is added to improve prediction performance.

3.1. Data Preprocessing Based on SSA-VMD. This part introduces data denoising, preprocessing, and parameter optimization theory. VMD is a signal decomposition method that can effectively extract different frequency and time domain features by decomposing the signal into multiple modal components. It offers high decomposition accuracy and stability, enabling robust handling of signals with noise and interference and delivering precise decomposition results. As the water level fluctuation is significant and random during the flood period, producing a series of noises. VMD is used to smooth data to denoise and increase the accuracy of water level prediction. However, manually selecting VMD parameters is time-consuming, and choosing the appropriate parameters is difficult. Therefore, SSA-VMD is used to intelligently obtain the optimization parameters K, α, and optimal decomposition components of VMD.

3.1.1. VMD Theory. The core of VMD is to construct and solve the following variational constraint problem [30, 47–49]:
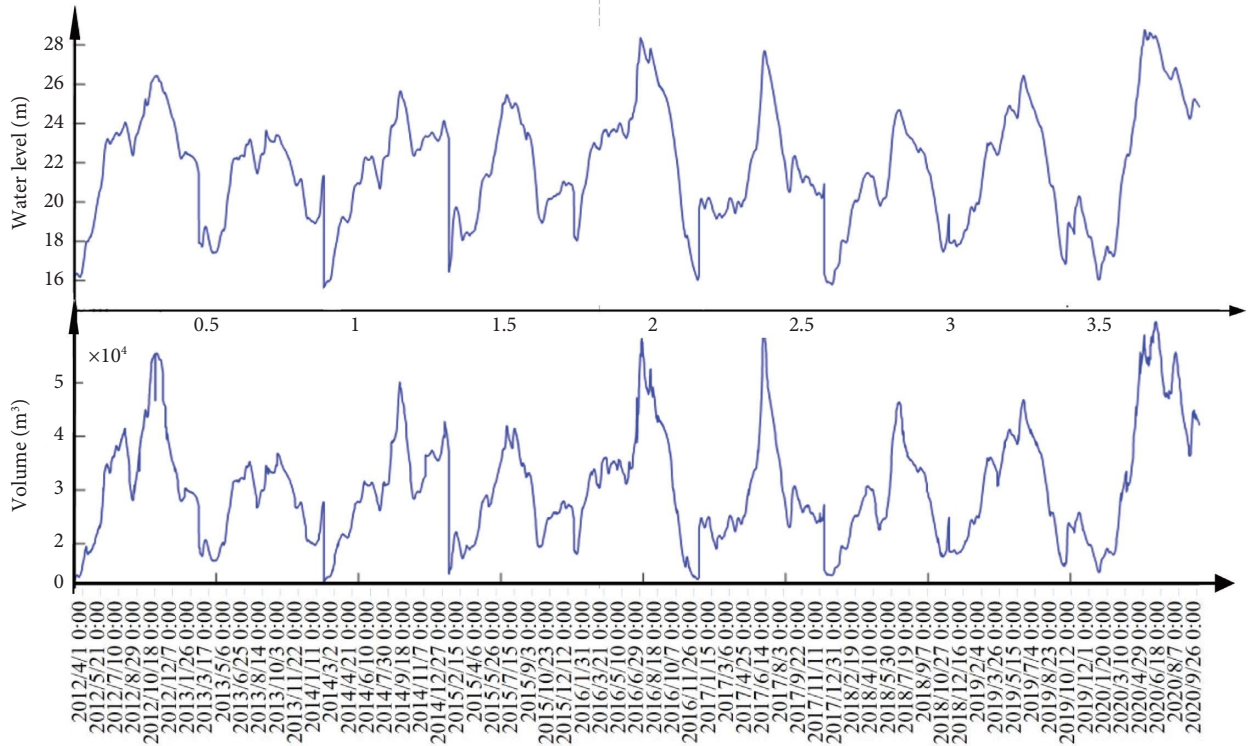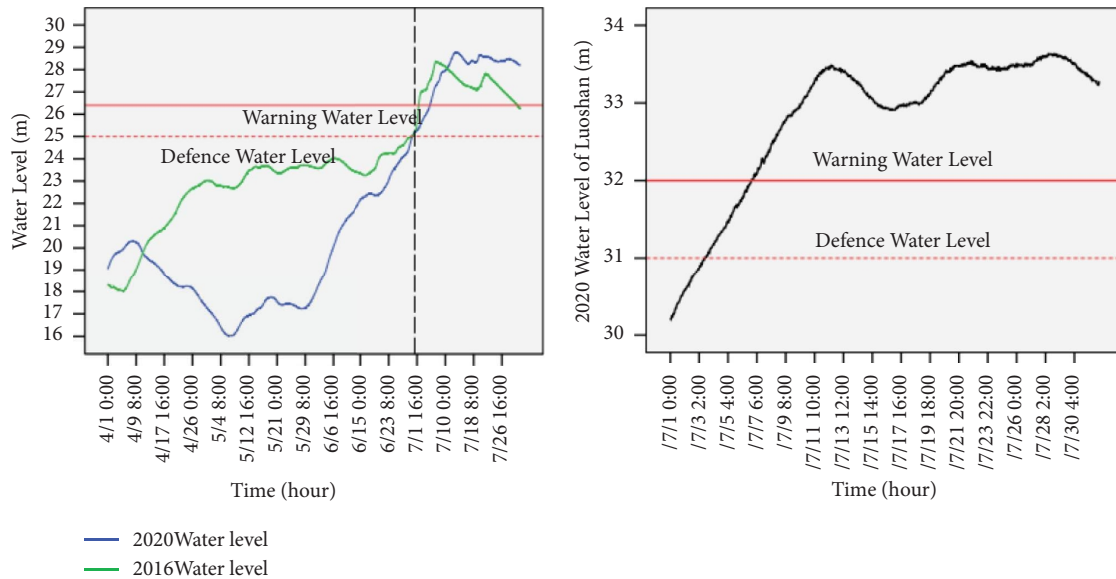
Figure 2: Water level and flow volume visualization in Hankou from 2012 to 2020.

Table 1: Water level data description from 2012 to 2020 (/m).

Average value	95% confidence interval of the mean value	Median value	Variance	Standard deviation	Minimum value	Maximum value	Full range	Interquartile distance
21.7592	(21.7307, 21.7879)	22.01	8.379	2.8946	15.38	28.77	13.39	4.13

(a)

(b)

Figure 3: Water level visualization of Hankou (a) and Luoshan (b).

$$\begin{cases} x(t) = \left[\left(\delta(t) + \dfrac{j}{\pi t} \right) u_k(t) \right] e^{-jw_k t}, \\\\ \min\limits_{\{u_k\},\{w_k\}} \sum\limits_k \left\| \partial_t x(t) \right\|_2^2, \\\\ \sum\limits_k u_k = f(t), \end{cases} \tag{1}$$

where k denotes the number of modes; u_k and w_k are the kth mode after decomposition and its corresponding center frequency, respectively; $\delta(t)$ and ∂_t are the Dirac delta function and the partial derivative, respectively; j is an imaginary number; $f(t)$ denotes the original data; and $e^{-jw_k t}$ is converted into simple harmonic through the Euler transform, which is used to adjust the spectrum of each mode component to its corresponding baseband [44].

The quadratic penalty factor α of VMD is to ensure the accuracy of the signal reconstruction and help reduce noise interference. The Lagrange multiplication operator $\lambda(t)$ is used to transform the constrained variational problem (1) into an unconstrained variational problem [48], as shown in the following equations (2) and (4) [50, 51]:

$$L(\{u_k\}, \{w_k\}, \lambda(t)) = \alpha \sum_k \left\| \partial_t x(t) \right\|_2^2 + \left\| f(t) - \sum_k u_k \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k \rangle. \tag{2}$$

The alternating-direction multiplication iterative algorithm is then used to find the saddle point of the extended Lagrangian function by combining the Fourier transform to optimize the modal components and the center frequency and identifying the optimal solution by alternating optimization search iterations. The final initial signal is decomposed into K-modal components, and the detailed VMD algorithm is as follows [30]:

(1) Initialize $\{\hat{u}_k^1\}$, and n and set them to 0. Select the appropriate number of modes K and penalty parameters α;

(2) Let $n = n + 1$. Update $\hat{u}_k^1$ and $\hat{w}_k^1$ according to (3) and stop the internal circulation until the number of decomposition reaches K;

$$\begin{cases} \hat{\lambda}^{n+1}(\omega) \longleftarrow \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum\limits_k \hat{u}_k^{n+1}(\omega) \right), \\\\ \hat{u}_k^{n+1} = \dfrac{\hat{f}(\omega) - \sum\limits_{i \neq k} \hat{u}_i(\omega) + (\hat{\lambda}(\omega)/2)}{1 + 2\alpha (\omega - \omega_k)^2}, \\\\ \omega_k^{n+1} = \dfrac{\int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega}. \end{cases} \tag{3}$$

(3) Update λ^1 according to (3);

(4) Determine whether the termination conditions are met in accordance with inequality.

$$\sum_k \frac{\left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2}{\left\| \hat{u}_k^n \right\|_2^2} < \varepsilon, \tag{4}$$

where $\hat{u}_k^n(w)$ is the signal corresponding to the original signal $u_k^n(w)$ after Fourier transform decomposition, $\hat{w}_k^n$ is the center of the amplitude modulation and frequency modulation signal ω_k^n after Fourier transform, and ε is the convergence judgment accuracy. If it does not meet the inequality (4), then return to step (iii). Otherwise, output K variable modal components.

3.1.2. SSA-VMD Optimization Algorithm. SSA is a new swarm intelligence optimization algorithm proposed by Xue and Shen [44] in 2020 based on the predation behavior of sparrows. This algorithm is not limited by the continuous differentiability of the objective function and has good stability and convergence. SSA iteratively finds the most optimal value of an objective function by comparing fitness values and uses the formulas described in [44] to renew those positions for the producers, the scroungers, and those who are aware of the danger.

The key to the design of the optimization algorithm using SSA is to construct a suitable fitness function. The envelope entropy [51] can be a good evaluation of signal sparsity, because it reflects the probability distribution characteristics of the signal decomposition. The envelope entropy value is considerable in the presence of noise and limited feature information in the IMF; otherwise, the envelope entropy value is small. The minimum envelope entropy is chosen in this paper as the target fitness function of SSA, and SSA-VMD is used to obtain the optimized parameters of VMD and the decomposition components.

The algorithm flow chart of SSA-VMD is shown in Figure 4.

3.2. Water Level Prediction Model Based on IPSO-BiLSTM-Attention. The main prediction model is constructed in this part. The PSO algorithm [46] is improved, and the IPSO is then used to optimize five hyperparameters of BiLSTM and provide the IPSO-BiLSTM-Attention algorithm process. x_t denotes the input data in the following discussion.
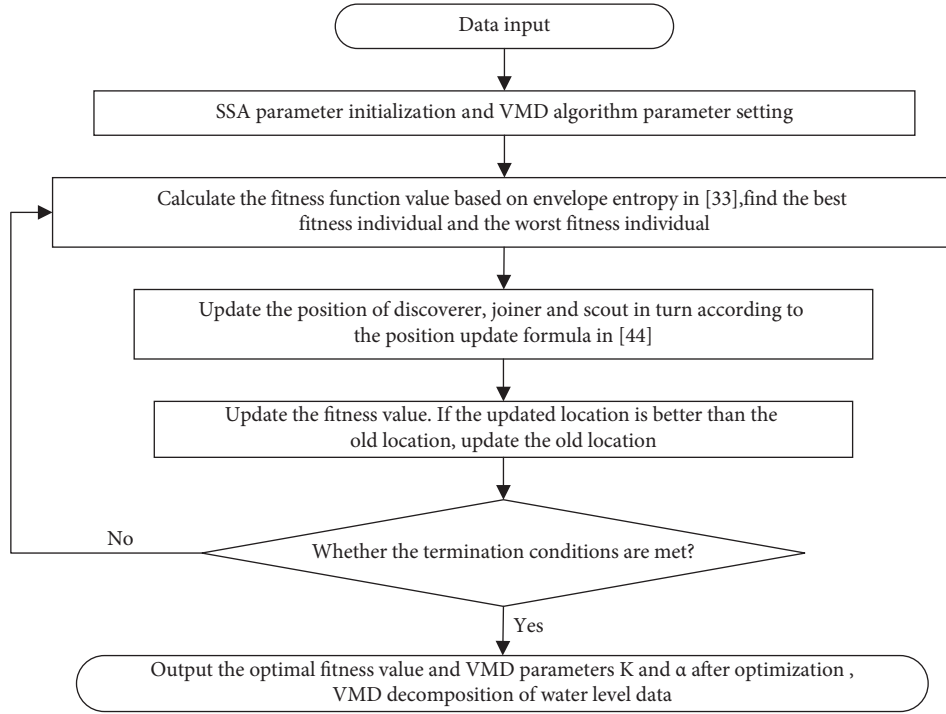
FIGURE 4: SSA-VMD algorithm flow chart.

3.2.1. BiLSTM Theory. BiLSTM refers to the bidirectional LSTM network. The LSTM network comprises several LSTM cells. Figure 5 illustrates the working state of one cell [22].

σ is the activation function, and the common form of σ is Sigmoid$(x) = 1/1 + e^{-x}$, which converts any value to the interval $[0, 1]$. The output double tangent activation function $\tan h$ is used to help adjust the value flowing through the network such that it remains between -1 and 1. t and $t - 1$ indicate the current and previous moments, respectively. Three data inputs are available: x_t, the output value h_{t-1}, and the united state C_{t-1}. The LSTM has two data outputs: the short-time output value h_t and the long-time unit status C_t. The LSTM cell is controlled by the following three control switches: forgetting gate ft, input gate i_t, and output gate o_t. The update process of a cell is as follows [26, 47]:

$$\begin{cases} \mathbf{f}_t = \sigma_1\left(W_f x_t + U_f h_{t-1} + b_f\right), \\ \mathbf{i}_t = \sigma_2\left(W_i x_t + U_i h_{t-1} + b_i\right), \\ \tilde{C}_t = \tan h\left(W_c x_t + U_c h_{t-1} + b_c\right), \\ \mathbf{o}_t = \sigma_3\left(W_o x_t + U_o h_{t-1} + b_o\right), \\ h_t = o_t \otimes \tan h\left(C_t\right), \\ C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t, \end{cases} \quad (5)$$

where W_k and U_k are the weight matrices and $k = f, i, o, c. b_f, b_i, b_o$ denote the bias vectors of the forget, input, and output gates, respectively. b_c is the bias vector of the candidate cell state $\tilde{C}_t$, and C_t represents the new memory cell state. $\otimes$ represents the matrix product. The forget gate determines how much of the cell state at the previous time C_{t-1} remains at the current time C_t [52], and
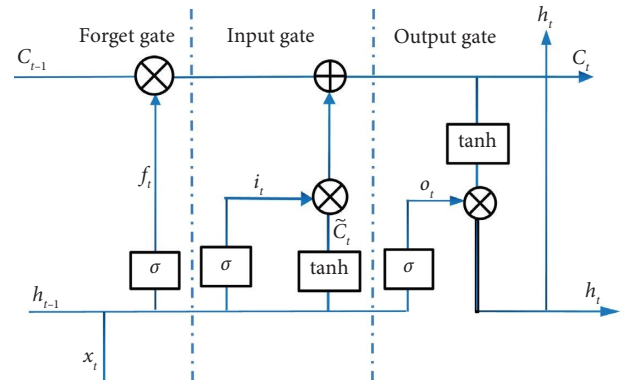

FIGURE 5: Structure of one LSTM cell.

the input gate determines how many network inputs x_t are saved to the cell state C_t; o_t controls how much of the status C_t outputs to the current output value h_t. Figure 5 provides the routes of C_t and h_t, which represent the transmission lines of long- and short-term memories, respectively.

One disadvantage of LSTM lies in its transmission of information only from front to back in one direction and its incapability to encode information from back to front. Considering the continuity of water flow and the influence of terrain characteristics, the current flood water level has a particular relationship with the water level before and after this time. BiLSTM is used as the main prediction function. BiLSTM comprises forward and backward LSTM and can effectively learn the data information and improve prediction accuracy. The single-layer BiLSTM comprises two LSTMs in two directions: the forward input to process the

sequence information and the reverse input to process the sequence information. After the processing is completed, the output of the two LSTMs is contacted as the output of BiLSTM. The operation formula of BiLSTM at the time t noted as S_t is as follows [53]:

$$
\begin{cases}
h_t = \overrightarrow{\text{lstm}}(x_t, s_{t-1}), \\
h_i = \overleftarrow{\text{lstm}}(x_t, s_{i-1}), \\
S_t = u_t s_t + v_t s_i + w_t,
\end{cases}
\tag{6}
$$

where $\overrightarrow{\text{lstm}}$ and $\overleftarrow{\text{lstm}}$ represent the forward and the backward transfers of LSTM network, respectively; s_t represents the forward hidden layer state, while s_i represents the backward hidden layer state; u_t denotes the output weight of the hidden layer in the forward propagation unit; and v_t represents the output weight of the hidden layer in the backward propagation unit; and w_t is the optimization parameter of hidden layer offset at the current time. The information transmission structure of BiLSTM is given in Figure 6.

In Figure 6, S_t and S_i represent the hidden layer states of the LSTM forward and backward, respectively, and y_t is the output information after the splicing of two LSTM units, $t \in N$.

3.2.2. Attention Mechanism.

The information-processing mechanism of human vision inspires the attention mechanism, which was first applied to image processing [54]. According to the attention mechanism, the visual system finds the focus area in the overall image to provide additional attention while suppressing the acquisition of useless information to improve computing efficiency and enhance prediction performance. In deep learning prediction, the weighted attention mechanism assigns different weights to data according to requirements, highlights relevant influencing factors, and helps the model make accurate judgments [23, 47]. Figure 7 provides the weighted attention mechanism.

The transformation of the attention mechanism comprises the following formulas:

$$
\begin{cases}
e_{ti} = V \tan h(W h_t + U h_i + b), \\
a_{ti} = \dfrac{\exp(e_{ti})}{\sum_{k=i}^{t} \exp(e_{tk})}, \\
C_t = \displaystyle\sum_{i=1}^{t} a_{ti} h_i, \\
s_t = f(C_t, h_i, x_t),
\end{cases}
\quad i = 1, 2, \ldots, t-1, \tag{7}
$$

where $(h_1, h_2, \ldots, h_t)$ denotes the hidden layer state corresponding to the input sequence $(x_1, x_2, \ldots, x_t)$, and the envelope entropy value is considerable in the presence of noise and limited feature information in the IMF; otherwise, the envelope entropy value is small. The minimum envelope includes the level of attention given to the entire input

sequence, focusing on the part of the input sequence surrounding the t-th word. a_{ti} is the weight value of the hidden layer output value corresponding to the current time t, s_t is the hidden layer state, and U, V, W, b are the training parameters and will be continuously adjusted with the training process of the model.

In the BiLSTM-Attention model, the similarity between input and hidden states is calculated to measure the attention distribution. The weighted average value of the input information is then calculated following the attention distribution. The training parameters e_{ti} are calculated, and the standardized weight is the output. The output of the BiLSTM hidden layer and a_{ti} is then obtained by using the activation function. At last, form the new output vector C_t.

3.2.3. Using IPSO to Optimize the Parameters of the BiLSTM-Attention Model.

A double-layer BiLSTM-Attention neural network is established to extract and predict the data features of flood water levels. The batch size parameter improves training accuracy in selecting the gradient descent direction. The two-layer neural network involves many super parameters, and improper settings will affect the prediction effect. The PSO algorithm [46] is improved, and this IPSO is then used to optimize five hyperparameters of BiLSTM: learning rates, iteration times, batch sizes, and the number of neurons in the two hidden layers.

PSO originates from the research modeling and simulation of the predatory behavior of birds [46]. The realization process starts with a random solution. The solution to each optimization problem is searched as particles in a D-dimensional space; the velocity and position of the particles are updated by the fitness function value of each particle; and the global optimum solution is searched iteratively by tracking the current optimum solution.

The change in particle speed of traditional PSO always maintains the same level; therefore, the algorithm converges slowly, and the search needs to be more robust. Thus, the algorithm is susceptible to falling into a local optimum solution. In [55], Shi and Eberhart treated the velocity term of the PSO algorithm by introducing inertia weights. According to [55, 56], in combination with the periodic characteristics of the Yangtze River, the inertial factor is modified as follows to ensure the convergence of the PSO algorithm:

$$
w_t = w_{\max} - (w_{\max} - w_{\min}) \cdot \dfrac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{8}
$$

where $w_{\max}$ and $w_{\min}$ are the maximum and minimum values of inertia factor, respectively, which are generally taken in the interval $(0, 1)$, $x = (\pi \cdot t)/(4 \cdot T_{\text{maxiter}})$, t is the number of iterations, T_{maxiter} is the maximum iterated algebra.

The inertia factor is close to the maximum when the iteration number is relatively small based on (8), and the particles can then quickly scan the entire search area and find the approximate range of the optimal solution. With the increasing of iterations, $(\pi \cdot t)/(4 \cdot T_{\text{maxiter}})$ gets small and
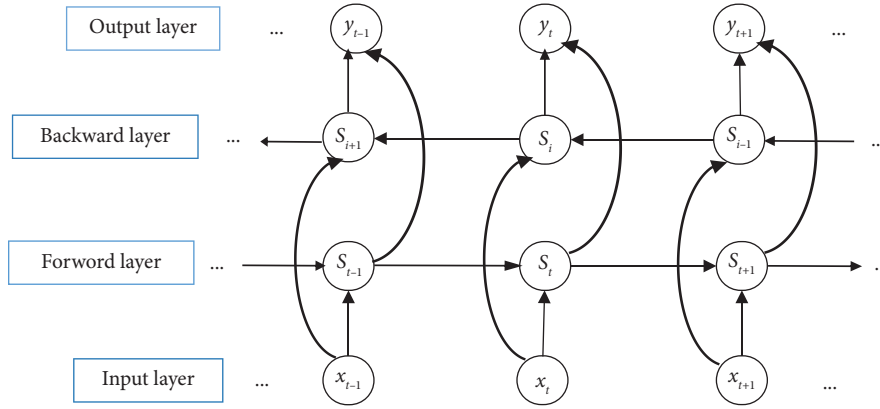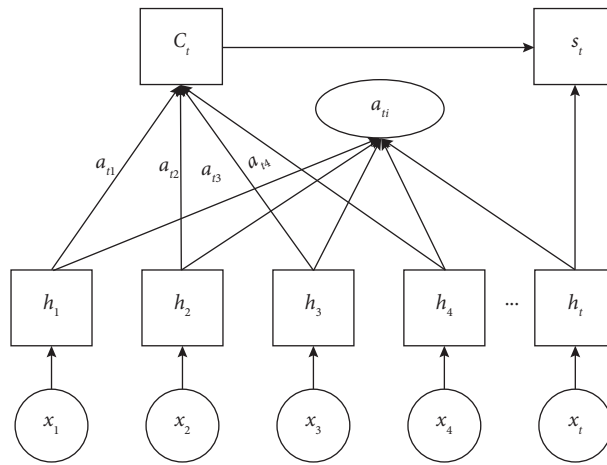
FIGURE 6: Structure of the BiLSTM network.

FIGURE 7: Structure of weighted attention.

the step length gets short, which ensures the accuracy of the local search. Correcting the nonlinear inertia factor can help the search reach a dynamic balance in convergence speed and local search accuracy. The specific steps of the IPSO algorithm for optimizing the hyperparameters of the BiLSTM network are shown below.

3.3. Research Framework. Figure 8 illustrates the specific research framework of the proposed model, SVIBA.

First, three-dimensional data variables are inputted, with the water level data as the primary variable and the flow and difference between the previous day's rise and fall water levels as the characteristic variables. Before data input, the missing data values are replaced with the proximal point average method, the outliers are eliminated, and the training and test sets are separated. After these sets are divided, the input data are normalized by (9) to eliminate the impact of dimensions between variables.

$$x_i^* = \frac{x_i - \bar{x}}{\sigma}, \tag{9}$$

where x_i^* is the input data after standardization, $\bar{x}$ is the mean value, and σ is the standard deviation.

SSA is then used to find the suitable VMD parameters $[K, \alpha]$. After VMD decomposition, the VMD decomposed components IMFs are obtained according to $[K, \alpha]$. The ReLU activation function is used on the BiLSTM hidden layers instead of the sigmoid function to avoid the problem of gradient disappearance or sliding. The inertia factor of the traditional PSO algorithm is formulated to develop the proposed model; the proposed IPSO is then used to optimize the double-layer BiLSTM hyperparameters on each IMF, which improves the operation speed and prediction capability of the entire model. The weighted attention mechanism is added after the BiLSTM neural network to strengthen the targeted extraction of feature information. Finally, all the VMD prediction components are integrated to obtain the prediction results.

4. Experiments and Results

MSE is taken as the loss function. The Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), the Mean Absolute Error (MAE), and goodness of fit R^2 shown in (10) are selected as objective evaluation indexes, and the comparison curve is used as the subjective evaluation index.

Input: Water level data
Begin
(1) Initialize particle swarm population size, number of iterations, weight factor, and other related parameters, and determine the
    initial optimization range of BiLSTM;
(2) Calculate and compare the fitness value, inertia factor, and boundary conditions of each particle;
(3) Perform an iterative update of particle velocity and position to find the optimal global value of particles according to [46];
(4) Determine if the maximum number of iterations has been reached;
(5) If Yes, export the optimized parameters; otherwise, return to step (2).
End
Output: Output BiLSTM optimization hyperparameters.
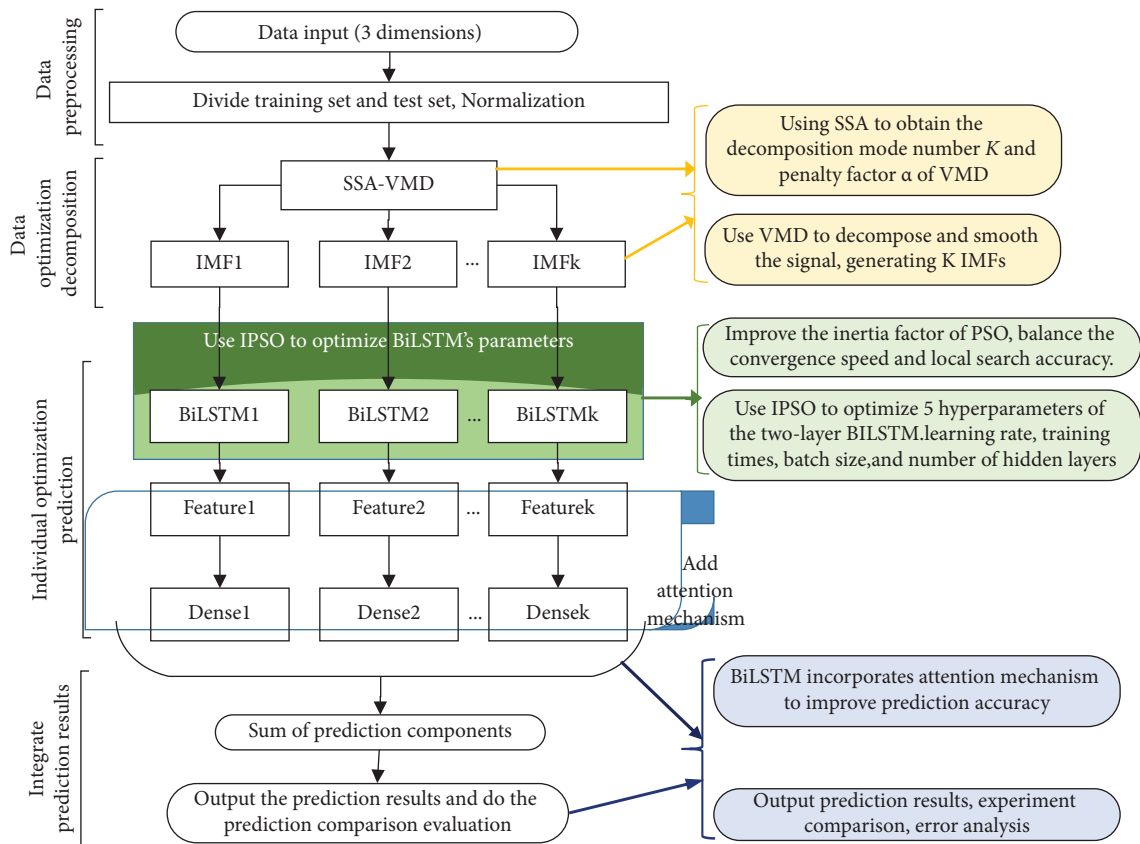
ALGORITHM 1: IPSO-BiLSTM-attention.

FIGURE 8: Prediction flow chart of SVIBA.

Input: Water level and characteristic variable data
Begin
(1) Divide the data into the training set and test set according to Part 2, Normalize data by (9);
(2) Obtain decomposition mode IMFS according to SSA-VMD shown in 3.1 and 4.2;
(3) For each IMF, use IPSO to optimize its BiLSTM parameters shown in Algorithm 1 and 4.3.1;
(4) Add the weighted attention mechanism to BiLSTM and use the optimized parameters to achieve the BiLSTM-Attention prediction
    results for each IMF as shown in 4.3.1;
(5) Integrate and add all prediction components to obtain the prediction results.
End
Output: Output prediction results.

ALGORITHM 2: SVIBA.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right)^2},$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - \hat{x}_i}{x_i}\right|,$$

(10)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \hat{x}_i\right|,$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2},$$

where x_i denotes the i-th input actual value, $\hat{x}_i$ denotes the predicted value, n denotes the number of training data, and $\bar{x}$ denotes the average value of $x_1, x_2, \dots x_n$. Small RMSE, MAPE, MAE and R^2 close to 1 indicate the superiority of the prediction model.

4.1. Water Level Data Decomposition Based on SSA-VMD.
In 2020, a major flood occurred around the Yangtze River basin, wherein more than 17,000 houses collapsed due to the flood, and the rainfall led to the continuous rise of the water level in the Yangtze River, causing tremendous losses to life safety and property security for people. The 2020 experimental data are inputted into the SSA-VMD model for decomposition and noise reduction based on the data preprocessing in Section 3.1 to effectively predict the water level change in July 2020. The initial population of the SSA algorithm is set as 15, and the number of iterations as 30; the upper limit of $[K, \alpha]$ is set to [50, 3]; and the lower limit is set to [2800, 10]. Based on the minimization of the envelope entropy as the objective function, after 30 iterations, the minimum local entropy of the entropy is 10.9376, and the optimal values of K and α are 4 and 713.72, respectively. The optimal adaptive evolution curve of envelope entropy and the optimization process of K and α are shown in Figures 9(a)–9(c). After decomposition based on optimized parameters, each IMF's component curves and Fourier spectra are shown in Figures 9(d) and 9(e).

Figure 9 shows that VMD decomposes the water level data into four modal IMFs, corresponding to the Fourier spectra from the lowest to the high-frequency components IMF1–IMF4. Each component curve is relatively straightforward, with almost no modal aliasing. The correlation coefficient between IMF1–IMF4 and the original data is calculated, and the four correlation coefficients are 0.9981, 0.2611, 0.0606, and 0.0260 in order. The low-frequency component IMF1 has the strongest linear correlation with the original data. IMF1 contains the primary information regarding the data, whereas components IMF2–IMF4 have a small linear relationship with the original signal. However, experiments show that these components also play a crucial role in reconstructing the original signal. The places where their signal strongly fluctuates are mainly concentrated in the extreme points, wherein the water level direction

changes. These points, especially the maximum points, which are the main cause of the flood, should be investigated; therefore, a nonlinear relationship between these components and the original water level data may exist. Using the same optimization algorithm, the VMD optimization parameters for the 2016 data set in Hankou are $[K, \alpha]_{2016} = [3, 551.56]$, and for July 2020 data set in Luoshan are $[K, \alpha] = [9, 730.39]$.

4.2. SVIBA Model Experiment and Multimodel Comparison

4.2.1. Parameter Setting and Optimization. This experiment is based on Python 3.6 software, and the hardware is the 12th-generation I7 processor, 16 G memory, and Windows 11 operating system. The five parameters of the BiLSTM network are sequentially recorded as a five-dimensional array (learning rate, training times, batch size, number of nodes in the first hidden layer, and number of nodes in the second hidden layer) to facilitate the expression. The lower limit of optimization is set as LB = [0.001, 20, 20, 16, 16], the upper limit of optimization UB = [0.01, 100, 60, 100, 100], and the number of iterations is 60. Let $w_{\max} = 0.8, w_{\min} = 0.6$ be the inertia factors of Equation (8). In the part of network internal parameter optimization, the adaptive moment estimation (Adam) optimizer is used to update the network parameters of each layer. In addition, the ReLU function is used as the activation function for the BiLSTM hidden layers to reduce the gradient disappearance problem of the Sigmoid activation function in the backpropagation of long-time series prediction.

After SSA-VMD decomposition in section 3.1, the distribution types of IMF1–IMF4 and their relationships with the original water level data are different. However, since the extreme points of the curve fluctuation are closely related to the change in the original water level, the hyperparameters of these components must be carefully set. The IPSO algorithm of 3.2.3 is then used to optimize the five BiLSTM parameters for each IMF. Take IMF4, which has the most significant difference from the original water level data distribution, as an example; its optimization process is shown in Figure 10.

Figure 10(a) shows that the optimal fitness function value is stable at 0.01569 after 60 iterations, and the combined parameters reach an optimized state after 12 optimizations. After further iterations, the model's learning rate, training times, batch size, and the number of nodes in the first and second layers reach stable values, as shown in Figures 10(b)–10(f). The five optimization parameters of IMF4 are found to be [0.002, 23, 44, 82, 77]. Similarly, the parameter optimizations of other IMFs are completed. Figure 11 shows the prediction result and loss curve of IMF4 by IPSO-BiLSTM-Attention using optimization parameters.

Figure 11 reveals that the optimized parameter prediction curves fit the original data well, and the change in the loss curve is also relatively close. Despite the significant frequency variations in IMF4, the model attained a superior prediction result due to the optimization of model parameters, stable parameter values, a close change in the loss curve, and an excellent fit to the data. Similarly, the
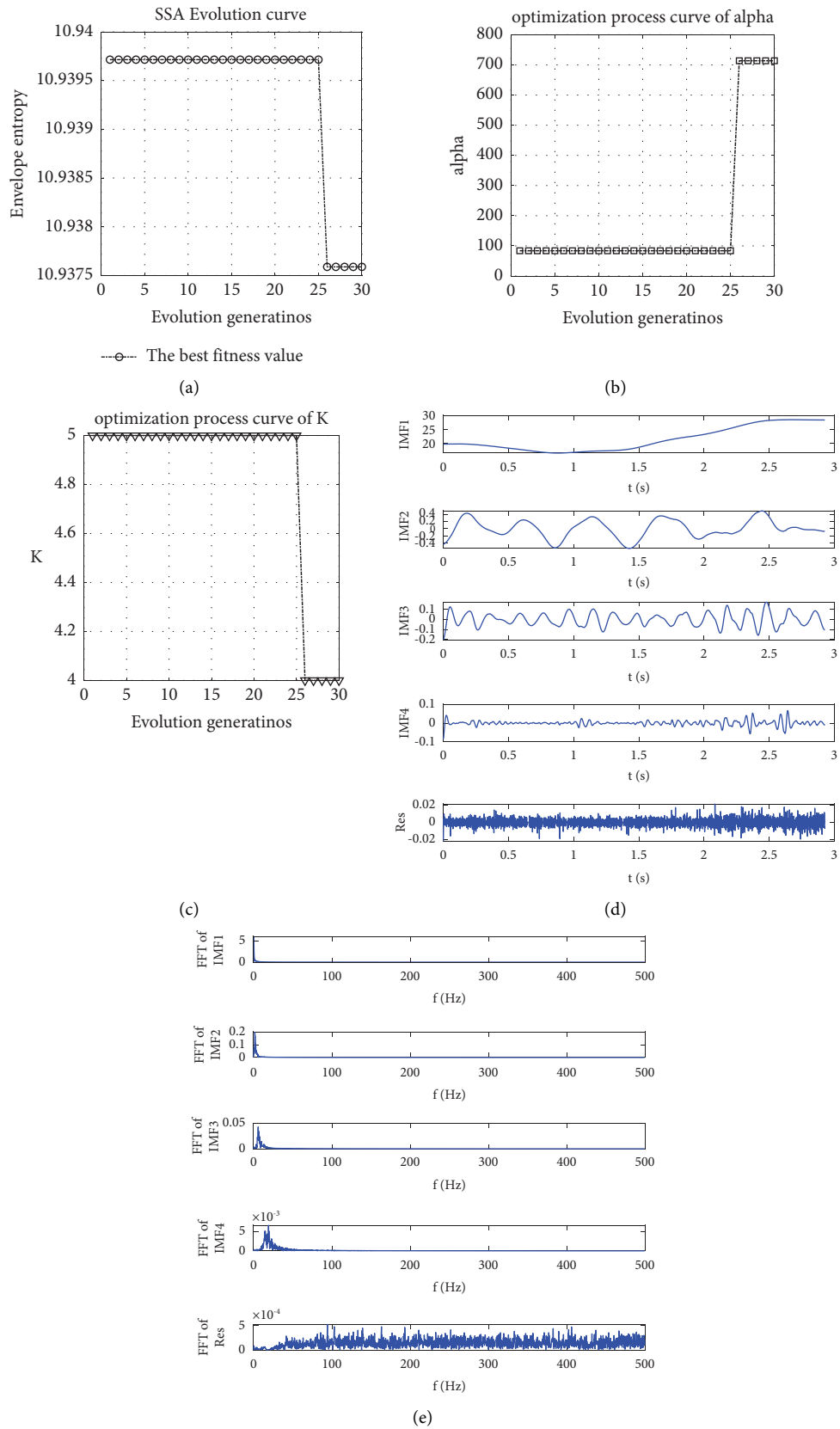
Figure 9: SSA-VMD optimization process and decomposition results. (a) Best fitness curve. (b) Optimization process of α. (c) Optimization process of K. (d) IMF curves after SSA-VMD. (e) Fourier spectra corresponding to each IMF.
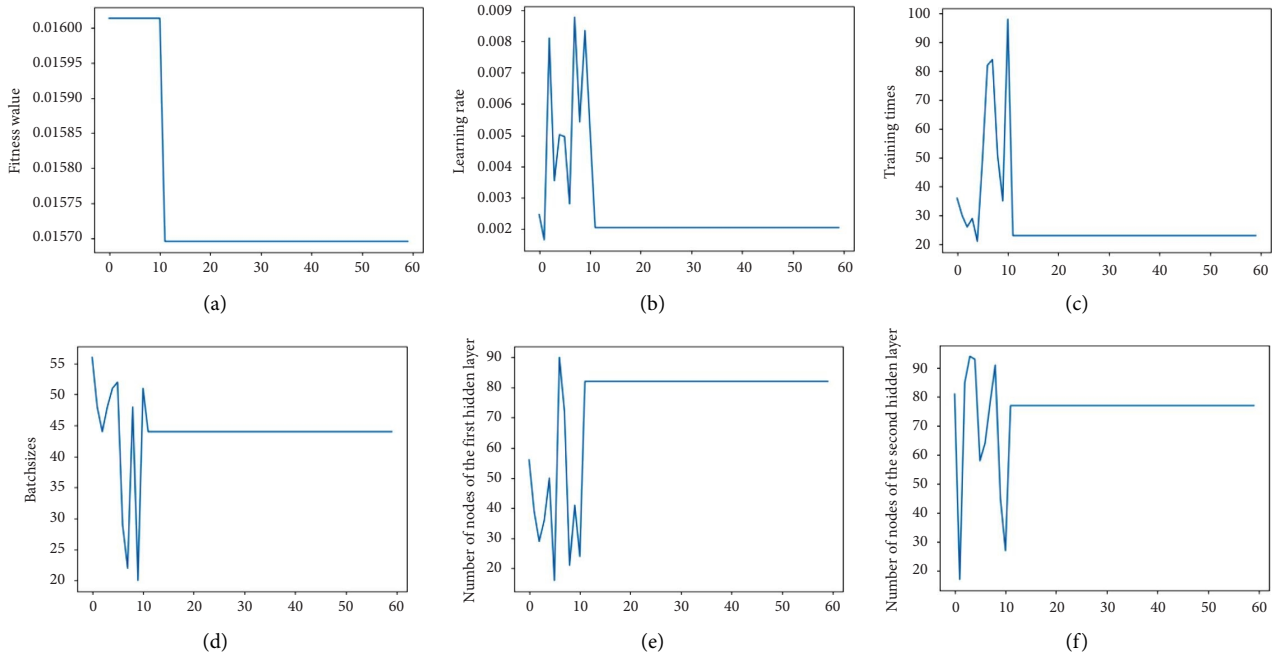
FIGURE 10: IPSO optimization process of IMF4 on the 2020 data set. (a) The fitness value. (b) The learning rate. (c) The curve of training times. (d) The optimization of batchsize. (e) Number of nodes of the first hidden layer. (f) Number of nodes of the second hidden layer.
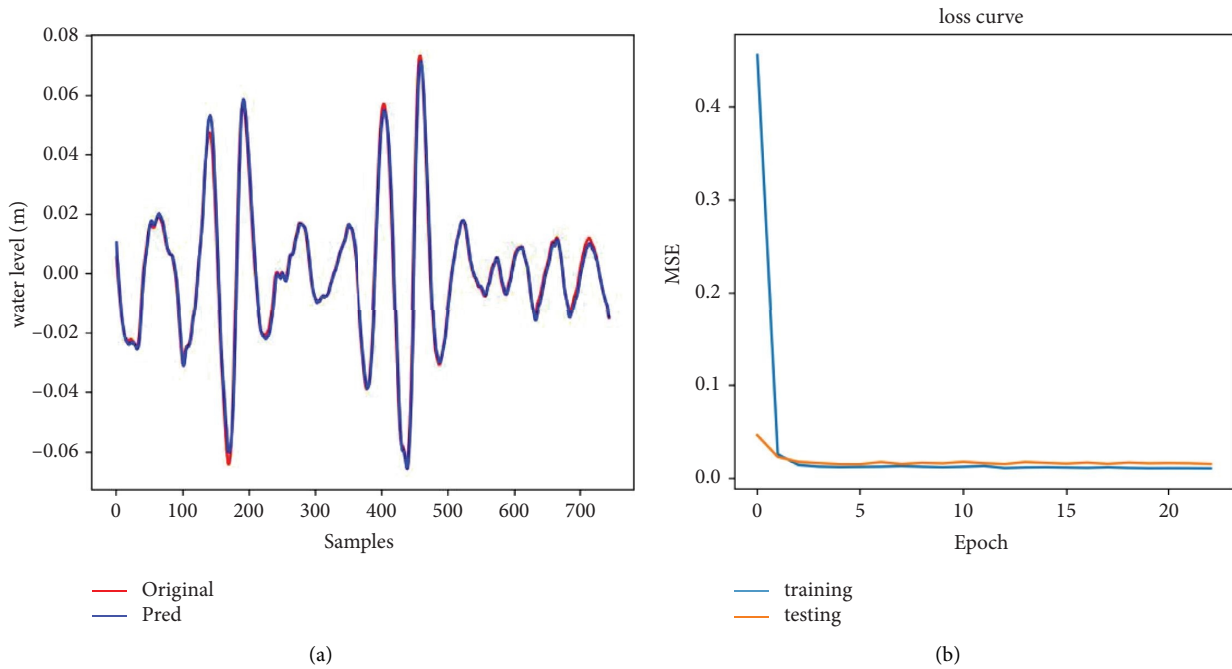

FIGURE 11: Forecast results for IMF4 (a) and loss curve (b) by IPSO-BiLSTM-attention.

optimization parameters and prediction results of other IMFs are obtained. Finally, the prediction components are integrated and added to output the prediction results.

4.2.2. Multimodel Comparison.
The control variable method was used to facilitate comparison. The comparison models LSTM, BiLSTM, and BiLSTM-attention all use double-layer networks. The same part of the main model employs the same

parameters. The five similar initial super parameters are set as [0.001, 100, 128, 10, 10], and the unoptimized comparison parameters for VMD are set as $[K, \alpha]$ = [4,1500]. The number of iterations of BP is 1000, hidden_layer_sizes is 100, and the prediction step is 1 hour.

(1) Experiment on the 2020 Data Set of Hankou. To verify the model's generalization capability, different data periods were chosen at the same site for short-term prediction

comparison experiments. First, the experiment was conducted on the 2020 data set, which led to substantial flood losses. Table 2 illustrates comparative experimental results for predicting the July water level change at Hankou station from April to June 2020. M7 is the CNN-BiLSTM-Attention algorithm [28], and M8 is the algorithm proposed by SVIBA.

Compared to BP, LSTM, BiLSTM, and BiLSTM-Attention, the introduction of VMD effectively improves the prediction accuracy of the BiLSTM-Attention algorithm. The RMSE value of VMD-BiLSTM-Attention (VBA) is 21% lower than that of the BiLSTM-Attention algorithm. The RMSE of the SSA-VBA algorithm after optimizing the parameters by SSA is 7% lower than that of no optimization, whereas that of SVIBA after optimizing the hyperparameters by IPSO is 86% lower than that of no optimization prediction. The improvement effect of other prediction indicators is also observed. Figure 12 shows the prediction curve comparison of M1–M8 on the 2020 data set.

Figure 12 shows that most of the water levels in July 2020 exceeded the warning value, and the water level seriously fluctuated, resulting in high prediction difficulty and a high risk index. The pretreatment after VMD decomposition and noise reduction markedly improved the accuracy of water level prediction. The algorithm shows improved prediction performance after parameter optimization. The SVIBA algorithm in this paper has achieved the best prediction accuracy compared with other algorithms, especially when the water level fluctuates strongly during the flood. Hence, this algorithm shows a superior water level tracking prediction effect.

(2) Experiment on the 2016 Data Set of Hankou. Comparison experiments were conducted on the 2016 data set to further verify the predictive capability of the model. Table 3 shows the predicted index values and the visualization of some indicators, and Figure 13 shows the prediction curves and details.

From Table 3, the change in water level data in 2016 is relatively stable, and each model shows improved prediction accuracy. From the RMSE prediction index, the BiLSTM-Attention with attention is approximately 32% lower than BiLSTM. After data preprocessing with VMD, compared with the RMSE without VMD parameter optimization, the RMSE of VBA prediction is reduced by 6.7%, and the SVBA optimized by SSA parameters is reduced by 31.8%. The RMSE of the algorithm optimized by IPSO is reduced by approximately 50.8% compared with the hyperparameters without optimization. Compared with the flood data with large fluctuations in 2020, the role of VMD is reduced, which may be related to the selection of basic parameters of VMD or the stability of water level fluctuation. The model's prediction capability is markedly improved after data preprocessing and parameter optimization. Combined with Figure 13, the prediction curve of the proposed SVIBA algorithm in this paper is remarkably close to the actual water level value, and the prediction fitting curve of flood peaks is closer to the actual value than other prediction curves.

(3) Experiment on the July 2020 Data Set of Luoshan. Luoshan also suffered a great flood disaster in July 2020. The training and prediction analysis of the flood period data in July is conducted to test the learning and prediction ability of SVIBA. The comparison of objective evaluation indicators and supervisor prediction curves of different algorithms is shown in Table 4 and Figure 14.

From Table 4 and Figure 14, when the data set becomes smaller, M7 is relatively stable, and its learning ability is better in small data sets. The learning ability of BP, LSTM, and BiLSTM is weakened, and their four evaluation indicators and the prediction curves are relatively poor. After adding VMD and Attention mechanism, M5 gets better, but the VMD without parameter optimization cannot improve the prediction performance very well. After SSA optimization, the RMSE of M6 is 37.44% lower than that of M3, and $R2$ is 19.77% higher. The prediction ability of the SVIBA framework has been dramatically improved for the IPSO optimization; compared with M6, RMSE decreased by 21.9%, R2 increased by 4%, and MAE is about 39% lower than M2, showing good prediction stability. Although it is worse than that on the Hankou station dataset, the comprehensive comparison shows that SVIBA had a more minor error during the flood with stronger randomness. Its relative short-term prediction performance is excellent.

5. Discussions

In this part, we take Hankou data set as the research object to discuss the sensitivity and prediction error and then give the possibility of expanding our model in practical applications.

(1) Different Prediction Steps Can Detect The Model's Generalization Capability. The comparison experiments were constructed with six different prediction steps, as listed in Table 5, to test the generalization capability of SVIBA. The objective evaluation index values under different prediction steps are obtained, and Figure 15 shows the subjective prediction curve.

Table 5 shows that the SVIBA model provides good generalization capability under different prediction steps, and most of the prediction index values are better than those of traditional algorithms. Different data sets reflect different forecast change rules. When the prediction step is 5 h, the prediction error of the data set in 2020 is larger than that when the step size is 10 and 15 h, but the prediction result in 2016 is just the opposite. This difference may be related to the larger fluctuation of the test set in 2020 compared with 2016. The prediction error shows a rising trend as the prediction step increases. When the step length is greater than 10, the prediction error increases with the rise of the prediction step and R^2 decreases in both data sets.

Figures 15(a) and 15(b) illustrate a clear trend: as the prediction step size increases, so does the prediction error. This pattern is especially pronounced in the initial 50% of the test set, as shown in Figures I and II. A rise in prediction points leads to greater variability in subsequent long errors, notably at the flood peak. After VMD preprocessing, the SVIBA algorithm prediction curve becomes smoother,

TABLE 2: Comparison table of water level prediction errors for July 2020 in Hankou.

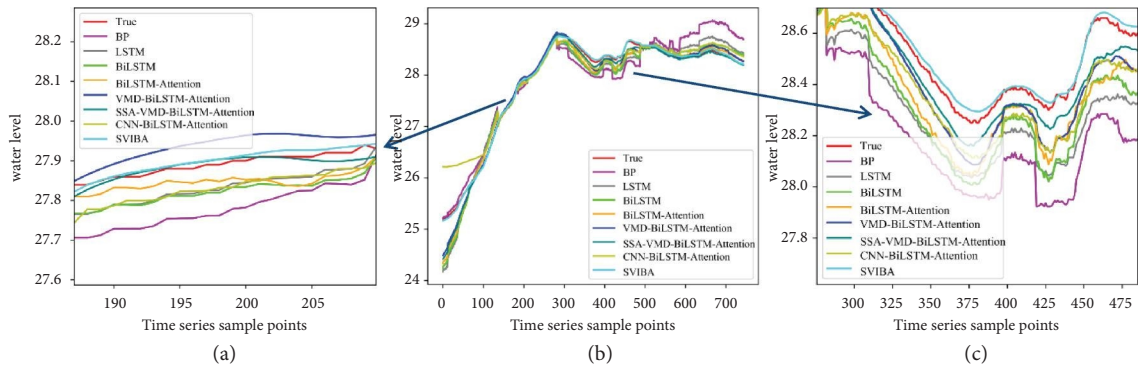| Number	Model	RMSE (m)	MAPE (%)	MAE (m)	R^2
M1	BP	0.3112	0.8958	0.2521	0.9046
M2	LSTM	0.2576	0.6964	0.1907	0.9346
M3	BiLSTM	0.2227	0.5647	0.1538	0.9511
M4	BiLSTM-attention	0.1842	0.4131	0.1116	0.9666
M5	VMD-BiLSTM-attention	0.1466	0.3488	0.0948	0.9788
M6	SSA-VMD-BiLSTM-attention	0.1363	0.2579	0.0718	0.9789
M7	CNN-BiLSTM-attention	0.2782	0.6387	0.1712	0.9255
M8	SVIBA	0.0192	0.0545	0.0152	0.9996

FIGURE 12: Comparison curves of different models in July 2020, Hankou. (a) Prediction details for points185–210. (b) The main prediction curve. (c) Prediction details for points 275-485.

TABLE 3: Comparison table of water level prediction errors for July 2016 in Hankou.

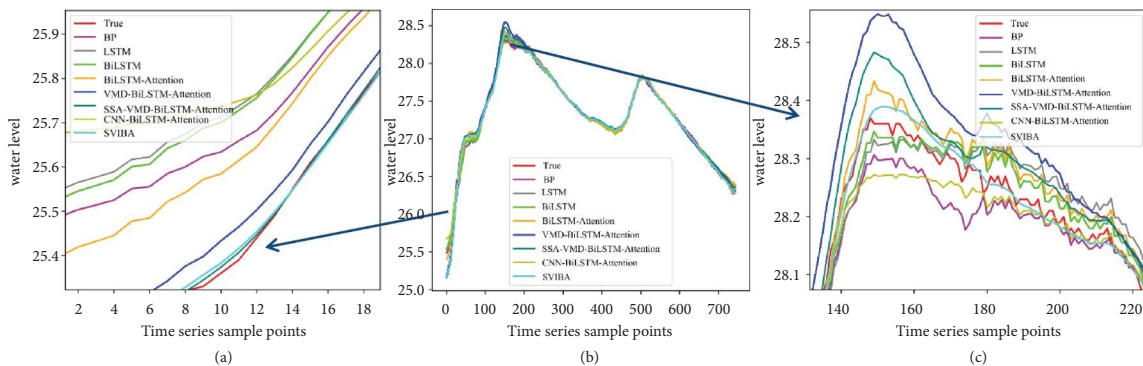| Number	Model	RMSE (m)	MAPE (%)	MAE (m)	R^2
M1	BP	0.0683	0.1412	0.0379	0.9872
M2	LSTM	0.0753	0.1515	0.0406	0.9845
M3	BiLSTM	0.0728	0.1406	0.0376	0.9855
M4	BiLSTM-attention	0.0495	0.1152	0.0310	0.9932
M5	VMD-BiLSTM-attention	0.0462	0.1021	0.0281	0.9943
M6	SSA-VMD-BiLSTM-attention	0.0315	0.0818	0.0224	0.9973
M7	CNN-BiLSTM-attention	0.0459	0.1157	0.0195	0.9851
M8	SVIBA	0.0155	0.0412	0.0113	0.9993

FIGURE 13: Comparison curves of different models in July 2016, Hankou. (a) Prediction details for points 2–19. (b) The main prediction curve. (c) Prediction details for points 135–222.

resulting in a lower prediction fitting degree to the peak point, as shown in Figure (b) III. However, at a 1-hour prediction step, the algorithm demonstrates remarkable stability; the predicted curve closely aligns with the actual curve, indicating high accuracy. This accuracy may stem from rapid water flow during floods and factors like

TABLE 4: Comparison table of water level prediction errors for July 2020 at Luoshan station.

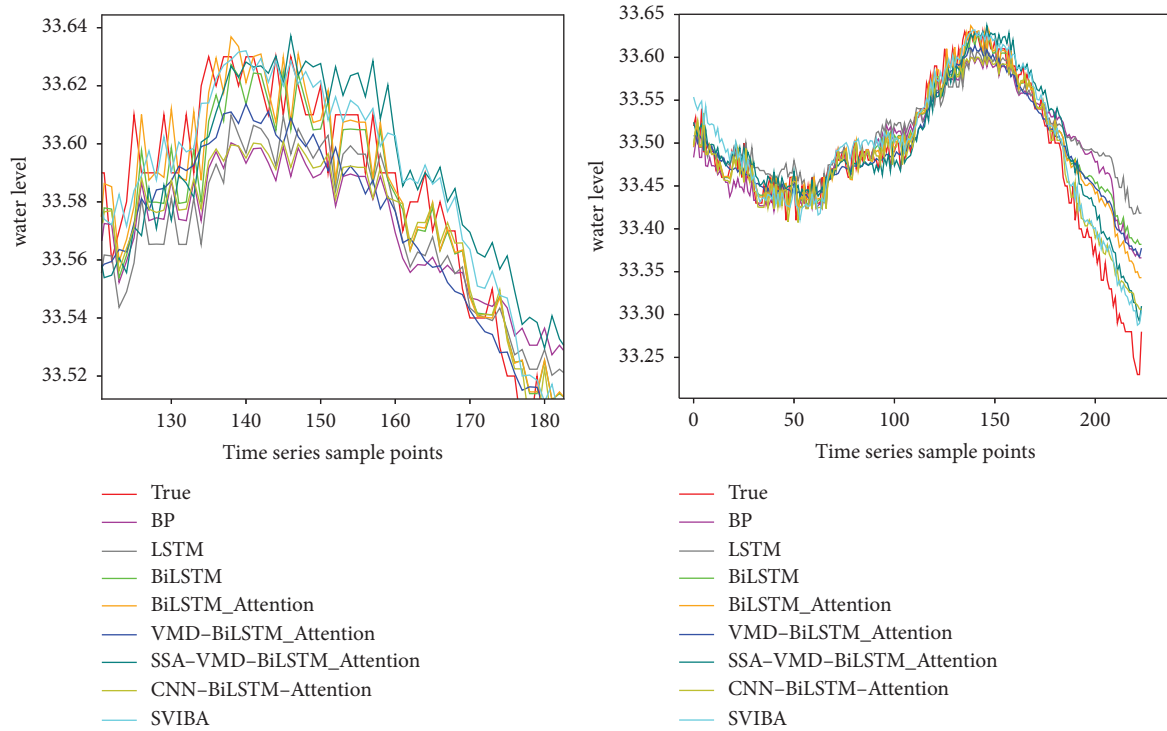| Number	Model	RMSE (m)	MAPE (%)	MAE (m)	R^2
M1	BP	0.0489	0.0999	0.0334	0.6955
M2	LSTM	0.0595	0.1152	0.0385	0.5500
M3	BiLSTM	0.0438	0.0830	0.0277	0.7556
M4	BiLSTM-attention	0.0367	0.0739	0.0247	0.8289
M5	VMD-BiLSTM-attention	0.0398	0.0761	0.0254	0.7984
M6	SSA-VMD-BiLSTM-attention	0.0274	0.0633	0.0212	0.9050
M7	CNN-BiLSTM-attention	0.0242	0.0563	0.0188	0.9251
M8	SVIBA	0.0214	0.0502	0.0168	0.9417

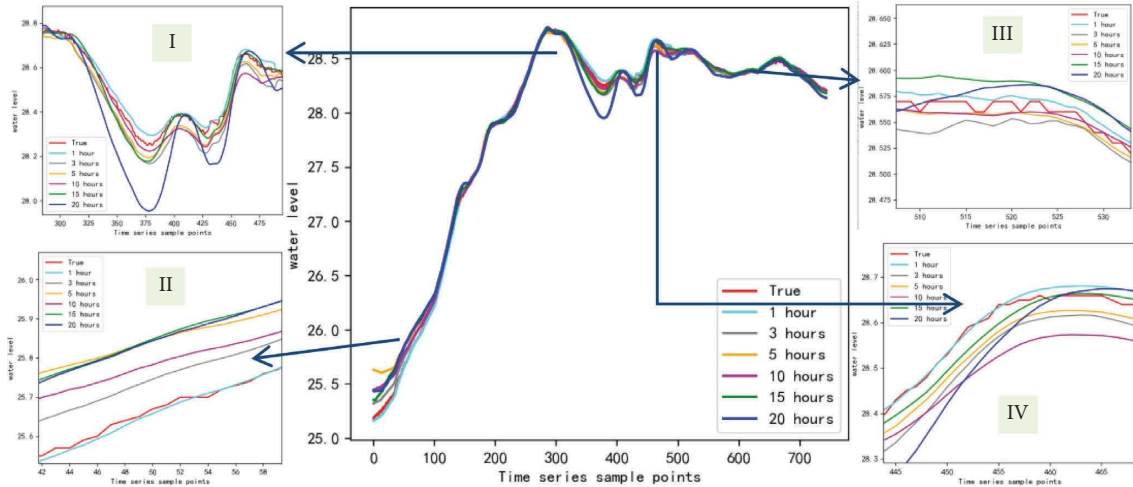FIGURE 14: Comparison curves of different models in July 2020, Luoshan.

TABLE 5: Comparison of prediction errors of the SVIBA model under different prediction steps.

| Data set	Prediction step	RMSE (m)	MAPE (%)	MAE (m)	R^2
2020	1 hour	0.0192	0.0547	0.0152	0.9996
	3 hours	0.0497	0.1370	0.0376	0.9976
	5 hours	0.0898	0.1794	0.0476	0.9921
	10 hours	0.0617	0.1351	0.0362	0.9962
	15 hours	0.0686	0.1583	0.0422	0.9953
	20 hours	0.1063	0.2638	0.0721	0.9888
2016	1 hour	0.0155	0.0412	0.0113	0.9993
	3 hours	0.0341	0.0931	0.0253	0.9968
	5 hours	0.0291	0.0758	0.0206	0.9977
	10 hours	0.0493	0.1263	0.0343	0.9934
	15 hours	0.0506	0.1434	0.0390	0.9930
	20 hours	0.0631	0.1693	0.0461	0.9891

real-time rainfall, which influence water level changes within 1-hour intervals. Consequently, a 1-hour prediction step is taken for further flood peak prediction error analysis.

(2) Prediction Error of the Model For the Time and Height of Flow Peaks. Table 6 shows the error analysis of the SVIBA model for 2016 and 2020 data sets when the prediction time is 1 hour.

(a)
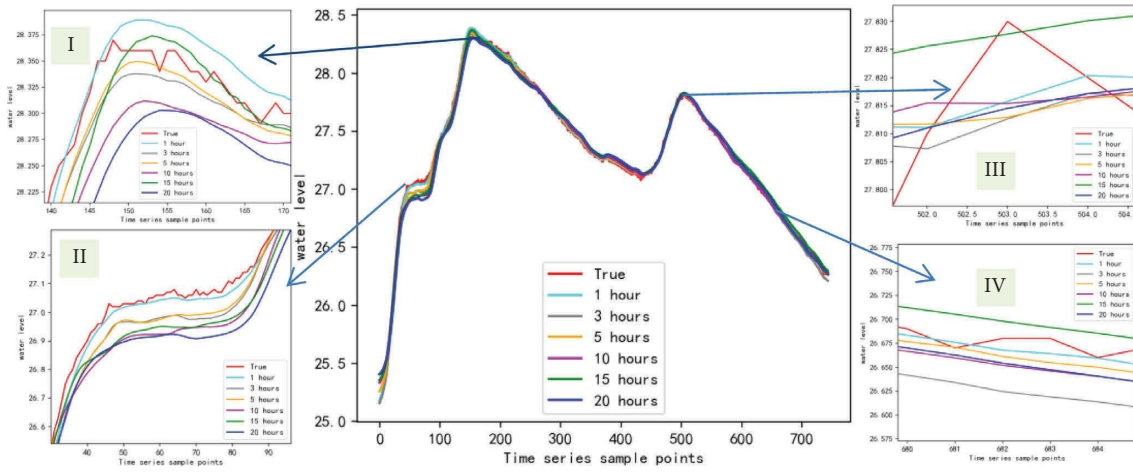
(b)

FIGURE 15: Comparison of SVIBA model prediction curves under different prediction steps. (a) Comparison of prediction curves for 2020 data sets under different prediction steps. (b) Comparison of prediction curves for 2016 data sets under different prediction steps.

TABLE 6: SVIBA model: the peak errors.

Peak time	Predictive peak time	Time Re (%)	Observed peak height (m)	Predictive height (m)	Predictive peak height (m)	Predictive height Re (%)	Peak height Re (%)	Predictive height Ae (cm)	Peak height Ae (cm)
2016/7/7 4:00	2016/7/7 7:00	30	28.37	28.3744	28.3894	0.02	0.07	0.44	1.94
2016/7/21 23:00	2016/7/22 0:00	10	27.83	27.8158	27.8203	0.05	0.03	1.42	0.97
2020/7/12 23:00	2020/7/13 0:00	10	28.77	28.7573	28.7583	0.04	0.04	1.27	1.17
2020/7/20 3:00	2020/7/20 5:00	20	28.66	28.673	28.6794	0.05	0.07	1.3	1.94
2020/7/28 18:00	2020/7/28 19:00	10	28.5	28.5101	28.5131	0.04	0.05	1.01	1.31

Table 6 shows that the maximum error of the flood peak prediction time of SVIBA is three h, and the minimum error is one h. The time prediction error on July 7, 2020, is three h, and the difference between the predicted and actual values is only 0.44 cm, which achieves a remarkably high prediction

accuracy. The absolute error of prediction from the maximum point of the flood peak is less than 2 cm, and the relative error of prediction height is less than 0.1%, which reaches a good prediction standard. However, a certain lag is generally observed in prediction time, which may be the

prediction error caused by using the previous data to predict future data.

(3) Model Advantages and Application Expansion. Compared to traditional prediction algorithms, the SVIBA model utilizes data-driven techniques such as VMD, SSA, and BiLSTM to capture and analyze the complex patterns and trends in flood data, which enhances the model's ability to make accurate predictions with less data. The proposed IPSO algorithm optimizes the parameters of the BiLSTM model, enhancing its predictive capability and accuracy. Using a weighted attention mechanism in the BiLSTM neural network improves the model's ability to focus on essential features and enhance the prediction accuracy. The experiment found that VMD has a better preprocessing effect when the flood fluctuation is larger and more severe.

As the SVIBA model supports input with $n(n \geq 2)$ dimensions, it can be applied to other flood-prone areas or river systems by incorporating local data and characteristics. Furthermore, as it can get accurate predictions with less data, it can integrate into existing flood management systems to save time and money. If we can find more data, the model can also assess the potential impact of climate change on flood events by incorporating climate forecasts and historical data.

6. Conclusions

This study concentrates on July, the month most susceptible to flooding along the Yangtze River. Smaller data sets are used to predict the water level during the flood. Three-dimensional variables are utilized as the data input to enhance the understanding of the spatial and temporal trend characteristics. VMD is used for data preprocessing and denoising. SSA optimizes VMD parameters $[K, \alpha]$ to yield suitable modes (IMFs), and a double-layer BiLSTM model is used on each IMF for prediction. The inertia factor of particle swarm optimization is improved, and this IPSO is used to optimize the parameters of BiLSTM. The weighted attention mechanism is used on the BiLSTM neural network to improve the processing capability of the prediction model. Notably, the model achieves an absolute error within 2 cm for peak height and a relative error within 30% for peak time arrival, while reducing the root mean square error by at least 50%. The comprehensive model SVIBA proposed provides better prediction results than other traditional models concerning statistical indexes, prediction curves, and error analysis, especially in flood periods. Additionally, the following three conclusions are summarized.

(1) VMD preprocessing can denoise and improve the flood prediction capability of the model. The selection of parameters in VMD directly affects the subsequent prediction accuracy. After SSA optimization parameters, the decomposition effect of VMD will be stable, and the prediction results will improve. When using a neural network prediction model to predict water levels, the prediction has a specific time lag at the point of significant fluctuation. The lag is improved after VMD preprocessing.

(2) The occurrence of flood has both periodicity and randomness. In addition to considering the periodicity information, the extraction and accurate use of the randomness characteristics of the data in a short time may be more effective for the subsequent flood prediction

(3) BiLSTM with Relu activation function is better than sigmoid in water level prediction, with an attention mechanism outperforming a no-attention mechanism in water level prediction, and the weighted attention can improve the model's capability to select important information. IPSO can optimize superior hyperparameter combinations and plays a considerable role in improving the prediction performance of BiLSTM-Attention.

Compared with the traditional models, such as BP, LSTM, BiLSTM, BiLSTM-Attention, and CNN-BiLSTM-Attention, the proposed prediction model has achieved good results but also has some limitations. Although the prediction accuracy has been dramatically improved compared with other models, the prediction lag of the flood peak still exists. In addition, when the prediction step size increases, the prediction accuracy will decline, making the reasonable selection of step size more important under different data sets. In the SVIBA model, variables with a more significant correlation with the primary prediction variable have a more considerable impact on the prediction effect, and multidimensional data input can obtain higher prediction accuracy. Rainfall and geographic information characteristics greatly influence flood variation, adding rainfall and some geographical features as the characteristic variables of the model. Further studying the short-term prediction will be considered to build a comprehensive statistical model to achieve better flood warnings. Our model supports multidimensional input and can be used for similar multidimensional data predictions. The input dimensions can be expanded, allowing researchers to optimize predictions based on different datasets.

Data Availability

The data presented in this study are available on http://113.57.190.228:8001/?WxUg5ztDmi=1653974255331#!/web/Report/RiverRepor-t.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] W. Hong, "Research on flood disaster risk management system," doctoral dissertation, Wuhan University, Wuhan, China, 2012.

[2] S. Mobini, B. Pirzamanbein, R. Berndtsson, and R. Larsson, "Urban flood damage claim analyses for improved flood damage assessment," *International Journal of Disaster Risk Reduction*, vol. 77, Article ID 103099, 2022.

[3] T. Skevas, R. Massey, and S. L. Hunt, "Farm impacts of the 2019 Missouri River floods and economic valuation of flood risk reduction," *Journal of Environmental Management*, vol. 344, Article ID 118483, 2023.

[4] S. K. Jain, P. Mani, S. K. Jain et al., "A Brief review of flood forecasting techniques and their applications," *International Journal of River Basin Management*, vol. 16, no. 3, pp. 329–344, 2018.

[5] N. Golding, C. Hewitt, P. Zhang, M. Liu, J. Zhang, and P. Bett, "Co-development of a seasonal rainfall forecast service: supporting flood risk management for the Yangtze River basin," *Climate Risk Management*, vol. 23, pp. 43–49, 2019.

[6] J. Zheng and G. Huang, "Towards flood risk reduction: commonalities and differences between urban flood resilience and risk based on a case study in the Pearl River Delta," *International Journal of Disaster Risk Reduction*, vol. 86, Article ID 103568, 2023.

[7] Z. Zhou, W. Tang, M. Li, W. Cao, and Z. Yuan, "A novel hybrid intelligent SOPDEL model with comprehensive data preprocessing for long-time-series climate prediction," *Remote Sensing*, vol. 15, no. 7, p. 1951, 2023.

[8] W. Liu, "Research on GIS based urban storm water simulation and prediction method and application," Master's thesis, Central South University, Changsha, China, 2010.

[9] K. Mondal, S. Bandyopadhyay, and S. Karmakar, "Framework for global sensitivity analysis in a complex 1D-2D coupled hydrodynamic model: highlighting its importance on flood management over large data-scarce regions," *Journal of Environmental Management*, vol. 332, Article ID 117312, 2023.

[10] D. Koujiawei, Z. ZhouJie, and Zhangxiangyu, "Flood forecasting and early warning system in Hongze Lake area based on WebGIS," *Journal of water resources and water engineering*, vol. 28, no. 6, pp. 145–150+157, 2017.

[11] G. A. Coelho, C. M. Ferreira, and J. L. Kinter III, "Multiscale and multi event evaluation of short-range real-time flood forecasting in large metropolitan areas," *Journal of Hydrology*, vol. 612, Article ID 128212, 2022.

[12] Y. Zhang, Y. Zhang, K. Shi, Y. Zha, Y. Zhou, and M. Liu, "A landsat 8 OLI-based, semianalytical model for estimating the total suspended matter concentration in the slightly turbid Xin'anjiang reservoir (China)," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 1, pp. 398–413, 2016.

[13] M. Rahman, N. Chen, A. Elbeltagi et al., "Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh," *Journal of Environmental Management*, vol. 295, Article ID 113086, 2021.

[14] S. M. S. Syed Musa, M. S. Md Noorani, F. Abdul Razak, M. Ismail, M. A. Alias, and S. I. Hussain, "Using persistent homology as preprocessing of early warning signals for critical transition in flood," *Scientific Reports*, vol. 11, no. 1, p. 7234, 2021.

[15] J. Zhang, Y. Wang, Y. Zhao, and H. Fang, "Multi-scale flood prediction based on GM (1,2)-fuzzy weighted Markov and wavelet analysis," *Journal of Water and Climate Change*, vol. 12, no. 6, pp. 2217–2231, 2021.

[16] Y. Xu and J. Wang, "A maximum-entropy-based method for alarm flood prediction," *Journal of Process Control*, vol. 107, pp. 58–69, 2021.

[17] M. Rahman, N. Chen, M. M. Islam et al., "Development of flood hazard map and emergency relief operation system using hydrodynamic modeling and machine learning algorithm," *Journal of Cleaner Production*, vol. 311, Article ID 127594, 2021.

[18] F. M. Aswad, A. N. Kareem, A. M. Khudhur, B. A. Khalaf, and S. A. Mostafa, "Tree-based machine learning algorithms in the Internet of Things environment for multivariate flood status prediction," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 1–14, 2021.

[19] Y. Zhao, "Research and application on BP neural network algorithm," in *Proceedings of the 2015 International Industrial Informatics and Computer Engineering Conference*, pp. 1444–1447, Shaanxi, China, January 2015.

[20] G. Kumar, U. P. Singh, and S. Jain, "Hybrid evolutionary intelligent system and hybrid time series econometric model for stock price forecasting," *International Journal of Intelligent Systems*, vol. 36, no. 9, pp. 4902–4935, 2021.

[21] T. Chou, T. Hoang, Y. Fang et al., "Swarm-based optimizer for convolutional neural network: an application for flood susceptibility mapping," *Transactions in GIS*, vol. 25, 2020.

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

[23] C. Dyer, A. Kuncoro, M. Ballesteros, and N. A. Smith, "Recurrent neural network grammars," 2016, https://arxiv.org/abs/1602.07776.

[24] R. Hu, F. Fang, C. C. Pain, and I. M. Navon, "Rapid spatiotemporal flood prediction and uncertainty quantification using a deep learning method," *Journal of Hydrology*, vol. 575, pp. 911–920, 2019.

[25] F. Noor, S. Haq, M. Rakib et al., "Water level forecasting using spatiotemporal attention-based long short-term memory network," *Water (Switzerland)*, vol. 14, no. 4, p. 612, 2022.

[26] C. Chen, J. Jiang, Z. Liao, Y. Zhou, H. Wang, and Q. Pei, "A short-term flood prediction based on spatial deep learning network: a case study for Xi County, China," *Journal of Hydrology*, vol. 607, Article ID 127535, 2022.

[27] Wangyingfei, huangyingping, Xiaomin, B. Xiong, S. Zhou, and Z. Jin, "Study on LSTM method for predicting water level of Yangtze River in flood season based on attention mechanism," *Journal of Three Gorges University (NATURAL SCIENCE EDITION)*, vol. 44, no. 03, pp. 13–19, 2022.

[28] Q. Nie, D. Wan, and R. Wang, "CNN-BiLSTM water level prediction method with attention mechanism," *Journal of Physics: Conference Series*, vol. 2078, Article ID 012032, 2021.

[29] D. Sahadevan, P. Ponnusamy, V. P. Gopi, and M. K. Nelli, "Ground-based 4d trajectory prediction using bi-directional LSTM networks," *Applied Intelligence*, vol. 52, no. 14, pp. 16417–16434, 2022.

[30] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[31] H. Li, X. Wu, T. Liu et al., "Composite fault diagnosis for rolling bearing based on parameter-optimized VMD," *Measurement*, vol. 201, Article ID 111637, 2022.

[32] Z. Xu, J. Zhou, L. Mo et al., "A novel runoff forecasting model based on the decomposition-integration-prediction framework," *Water (Switzerland)*, vol. 13, no. 23, p. 3390, 2021.

[33] D. Yan, R. Jiang, J. Xie, J. Zhu, J. Liang, and Y. Wang, "A multivariate and multistage streamflow prediction model based on signal decomposition techniques with deep learning," *Journal of Coastal Research*, vol. 37, no. 6, p. 2021, 2021.

[34] H. Niu, K. Xu, and W. Wang, "A hybrid stock price index forecasting model based on variational mode decomposition

and LSTM network," *Applied Intelligence*, vol. 50, no. 12, pp. 4296–4309, 2020.

[35] Z. Zhihua, W. Huanning, H. Yan et al., "Research on non-invasive load identification method based on VMD," *Energy Reports*, vol. 9, pp. 460–469, 2023.

[36] S. Chen, M. Ren, and W. Sun, "Combining two-stage decomposition based machine learning methods for annual runoff forecasting," *Journal of Hydrology*, vol. 603, Article ID 126945, 2021.

[37] J. Li, X. Cheng, Q. Li, and Z. Meng, "Adaptive energy-constrained variational mode decomposition based on spectrum segmentation and its application in fault detection of rolling bearing," *Signal Processing*, vol. 183, Article ID 108025, 2021.

[38] J. Xia and J. Chen, "A new era of flood control strategies from the perspective of managing the 2020 Yangtze River flood," *Science China Earth Sciences*, vol. 64, 2020.

[39] R. Sun, Z. Gong, W. Guo, A. A. Shah, J. Wu, and H. Xu, "Flood disaster risk assessment of and countermeasures toward Yangtze River Delta by considering index interaction," *Natural Hazards*, vol. 112, no. 1, pp. 475–500, 2022.

[40] R. Huo, H. Chen, L. Li et al., "Flood variability in the upper Yangtze River over the last millennium—insights from a comparison of climate-hydrological model simulated and reconstruction," *Science China Earth Sciences*, vol. 66, no. 3, pp. 547–567, 2023.

[41] Y. Qin, Y. Lei, X. Gong, and W. Ju, "A model involving meteorological factors for short- to medium-term, water-level predictions of small- and medium-sized urban rivers," *Natural Hazards*, vol. 111, no. 1, pp. 725–739, 2021.

[42] Z. Yuan, J. Liu, Y. Liu, Q. Zhang, Y. Li, and Z. Li, "A two-stage modelling method for multi-station daily water level prediction," *Environmental Modelling & Software*, vol. 156, Article ID 105468, 2022.

[43] http://113.57.190.228:8001/?WxUg5ztDmi=1653974255331#!/web/Report/RiverRepor-t.

[44] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.

[45] Y. Shen, B. Tang, B. Li, Q. Tan, and Y. Wu, "Remaining useful life prediction of rolling bearing based on multi-head attention embedded Bi-LSTM network," *Measurement*, vol. 202, Article ID 111803, 2022.

[46] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the ICNN'95-International Conference on Neural Networks*, pp. 1583–1589, Perth, WA, Australia, November 2002.

[47] A. Yan, J. Gu, Y. Mu, J. Li, S. Jin, and A. Wang, "Research on photovoltaic ultra short-term power prediction algorithm based on attention and LSTM," *IOP Conference Series: Earth and Environmental Science*, vol. 675, no. 1, Article ID 012078, 2021.

[48] C. Zhang, Y. Zhang, C. Hu, Z. Liu, L. Cheng, and Y. Zhou, "A novel intelligent fault diagnosis method based on variational mode decomposition and ensemble deep belief network," *IEEE Access*, vol. 8, pp. 36293–36312, 2020.

[49] L. Han, R. Zhang, X. Wang, A. Bao, and H. Jing, "Multi-step wind power forecast based on VMD-LSTM," *IET Renewable Power Generation*, vol. 13, no. 10, pp. 1690–1700, 2019.

[50] L. Chang, H. Cao, and C. Shen, "Dual-mass MEMS gyroscope parallel denoising and temperature compensation processing based on WLMP and CS-svr," *Micromachines*, vol. 11, no. 6, p. 586, 2020.

[51] P. Zhang, W. Zhang, X. Zhao, X. Wu, and N. Liu, "Application of WOA-VMD algorithm in bearing fault diagnosis," *Noise and vibration control*, vol. 41, no. 4, pp. 86–93+275, 2021.

[52] R. Zhao, Y. Yin, Y. Shi, and Z. Xue, "Intelligent intrusion detection based on federated learning aided long short-term memory," *Physical Communication*, vol. 42, Article ID 101157, 2020.

[53] W. Lin, B. Zhang, H. Li, and R. Lu, "Multi-step prediction of photovoltaic power based on two-stage decomposition and BILSTM," *Neurocomputing*, vol. 504, pp. 56–67, 2022.

[54] S. Zhang, S. Chen, J. Zhang, Z. Cai, and L. Hu, "Image annotation of ancient Chinese architecture based on visual attention mechanism and GCN," *Multimedia Tools and Applications*, vol. 81, no. 28, pp. 39963–39980, 2022.

[55] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, Washington, DC, USA, July 1999.

[56] H. Li and Y. Gao, "Particle swarm optimization velocity equation improvement and adaptive mutation strategy," *Computer engineering and application*, vol. 46, no. 13, pp. 47–50, 2010.