

Ensemble Prediction of China Consumer Price Index Based on MLP and BiLSTM

Zhuoheng Song, Cong Gu*, Canhui Zhang, Shuo Zhao

School of Mathematics and Information Science, Zhongyuan University of Technology, Zhengzhou, Henan, China

**Corresponding Author*

Abstract: Accurate prediction of the Consumer Price Index (CPI) is crucial for policymakers to grasp a country's economic operation rules and has a significant impact on policymaking and resource allocation. However, the nonlinear and nonstationary characteristics of financial data pose significant challenges to achieving accurate and robust predictions. This paper proposes a novel multi-step forecasting ensemble model that integrates Hiking Optimization Algorithm-optimized Variational Mode Decomposition (VMD), Multilayer Perceptron (MLP), and Bidirectional Long Short-Term Memory (BiLSTM) networks called HOA-VMDMLP-BiLSTM (dynamic weighting). Initially, cubic spline interpolation transforms monthly data into weekly frequency to address sample size limitations. Subsequently, the Hiking Optimization Algorithm (HOA) adaptively determines the optimal decomposition modes in VMD, obtaining the independent and stationary components while reducing noise interference. The BiLSTM and MLP networks then respectively process the decomposed modes and interpolated weekly series, with their predictions dynamically weighted through inverse error variance weighting to get the final value. Experimental results show that the Model's determination coefficients (R^2) values for 1 step, 5 step, and 9 step predictions are 0.9964, 0.9776, and 0.9411 respectively. The Diebold-Mariano test rejects the null hypothesis at the 1% significance level, indicating the proposed model's statistical superiority over benchmark methods. Notably, the proposed model demonstrates not only excellent in 1 step prediction but also robust and reliable in multi-step forecasting.

Keywords: CPI Forecast; HOA-VMD; MLP;

BiLSTM; Ensemble Prediction

1. Introduction

Consumer Price Index (CPI), which reflects price fluctuations of goods and services related to residents' daily lives, serves as a crucial indicator for predicting inflation and provides valuable references for assessing national economic development trends. Accurate CPI forecasting not only helps gauge consuming willingness but also offers significant guidance for industrial restructuring, resource allocation optimization, and consumption pattern upgrading. "The Belt and Road" represents an emerging market with substantial potential, where China's pivotal role as the initiator profoundly influences regional development. Current research primarily focuses on policy aspects such as supply chain transfer impacts, engineering project cooperation, cultural export determinants, and business climates. These studies aim to develop a deep learning model as reference framework for forecasting economic trends in "The Belt and Road" countries [1].

Early research on CPI prediction predominantly employed traditional econometric approaches, including basic trend models, autoregressive models, ARIMA, SARIMA, and exponential smoothing. These linear-based models exhibit limitations in capturing complex nonlinear patterns while demonstrating high sensitivity to parameter selection, where different parameters yield substantially varied outcomes. Furthermore, requirements for stationarity in models like ARIMA and SARIMA may lead to information loss through excessive differencing. Given these constraints, machine learning methods have gained prominence among researchers due to their superior prediction accuracy and enhanced generalization capabilities [2].

With the remarkable success of machine learning, particularly neural network models in computer vision applications, scholars

worldwide have progressively introduced these techniques into CPI forecasting tasks. Sun utilized a backpropagation (BP) neural network for monthly CPI forecasting, achieving excellent fitting performance [3]. Lu, Zhao, and Bi integrated fuzzy information granulation with support vector machines (SVM) for CPI prediction, with experimental results indicating superior model accuracy [4]. Zeng innovatively processed prediction errors generated by traditional statistical models through random forest (RF) algorithms, subsequently combining outputs through additive integration [5]. This approach demonstrated that integrating machine learning models to handle nonlinear components significantly outperformed standalone traditional statistical models. Sarangi, Sahoo, and Sinha proposed an artificial neural network (ANN) optimized by particle swarm optimization, with empirical evidence confirming enhanced accuracy in forecasting India's CPI [6].

Furthermore, financial time series exhibit non-stationarity characteristics. Direct modeling without appropriate feature preprocessing substantially hinders prediction accuracy improvements. The decomposition-ensemble framework has emerged as an effective solution, initially decomposing original sequences through signal processing techniques, subsequently feeding these subsequences into machine learning models for individual prediction, and ultimately aggregating partial results for final forecasting. Empirical analyses across various domains have validated the efficacy of this methodology. Wen and Xu employed Empirical Mode Decomposition (EMD) to decompose wind turbine power time series into intrinsic mode functions (IMFs), constructing prediction models based on component characteristics [7]. Their experimental results demonstrated superior performance compared to traditional approaches. However, EMD decomposition suffers from inherent limitations including spurious components and mode mixing phenomena. To address these issues, researchers have introduced Ensemble Empirical Mode Decomposition (EEMD) as an enhanced decomposition methodology. Tai and Liu developed an EEMD-PSO-SVM hybrid model for monthly CPI forecasting, implementing EEMD for noise-filtered decomposition followed by Particle Swarm Optimization (PSO) for parameter hunting in Support Vector Machines (SVM) [8].

Empirical evaluations confirmed this model's superiority over ANN, ARIMA, and standalone SVM. Ying and Wang proposed an EEMD-BP forecasting framework, with comparative studies revealing that the combined model outperformed single-algorithm approaches in CPI forecasting applications [9].

Particularly, Long Short-Term Memory (LSTM) networks demonstrate distinct advantages over Recurrent Neural Networks (RNN). This has prompted increasing adoption of LSTM-based hybrid models and architectural improvements in forecasting applications. Fang et al. proposed a CEEMDAN-Pearson-LSTM model that extracts critical water quality indicators through CEEMDAN decomposition and Pearson correlation analysis, with validation across five monitoring stations confirming its effectiveness in dissolved oxygen prediction [10]. Y. Zhou, Peng, and Bai designed a CNN-LSTM architecture for daily blood collection forecasting in major cities of China, demonstrating optimal performance across evaluation metrics due to its superior dynamic tracking and feature recognition capabilities [11]. Dong and Tang innovatively integrated multi-representational attention and soft-attention mechanisms into LSTM networks, constructing an ATT-LSTM-ATT model for CPI prediction that outperformed benchmark models in both precision and robustness [12]. Furthermore, the advancement of signal processing techniques has popularized Variational Mode Decomposition (VMD) [13]. An adaptive decomposition method with physical meaning, as evidenced by extensive applications [14]. Researchers have addressed VMD's hyperparameter optimization challenges through various metaheuristic approaches including Particle Swarm Optimization (PSO) and Bayesian Optimization (BO) [15, 16]. These methodological developments provide novel perspectives for CPI forecasting research.

Accordingly, this paper proposes a novel hybrid model named HOA-VMD-MLPBiLSTM (Dynamic Weight) to address the nonlinear and non-stationary characteristics in CPI forecasting, which integrates Hiker Optimization Algorithm (HOA), Variational Mode Decomposition (VMD), Multilayer Perceptron (MLP), and Bidirectional Long Short-Term Memory (BiLSTM). The implementation framework consists of three main phases: First, the HOA is introduced to determine the optimal number of

decomposition modes in VMD. Subsequently, the decomposed modal components are fed into the BiLSTM model for prediction training, while the original sequence is simultaneously processed through the MLP model. Finally, dynamic weights are assigned to both models through the inverse error variance method to generate final predictions. This architecture enables the model to simultaneously capture implicit latent patterns through BiLSTM and explicit global features through MLP, thereby enhancing overall predictive capability. Experimental results demonstrate superior performance of the proposed model in both one-step and multi-step forecasting scenarios. Furthermore, Diebold-Mariano (DM) tests confirm statistically significant differences in predictive accuracy between this model and comparative benchmarks.

The contributions of this study are primarily reflected in two aspects:

First, due to the challenges in collecting and calculating CPI data, making the sample sizes are typically limited. While previous research predominantly focused on developing small-sample prediction models. Our experimental results demonstrate that predictions based on cubic spline-interpolated data achieve satisfactory performance. Furthermore, we introduce the HOA method to determine the optimal number of VMD modes by minimizing energy entropy, thereby obtaining independent and stationary decomposition results to enhance data quality for subsequent model training.

Second, we design an ensemble forecasting model combining MLP and BiLSTM. These two models respectively capture extrinsic trend features and intrinsic detailed features of CPI data, the inverse error sum of squares method is employed to determine model weights, then ultimately generating final predictions. The core concept of this framework lies in utilizing trend features to govern detailed features, ensuring precise and robust forecasting performance.

The paper is structured as follows: Section 2 reviews relevant methodologies, including the HOA-VMD mechanism, theoretical foundations of MLP and BiLSTM, and model integration strategies. Section 3 describes the details of data preprocessing, hyperparameter selection, and multi-perspective experimental analyses. Section 4 gives the conclusions and future research directions.

2. Methodology

2.1 HOA-VMD Decomposition Framework

2.1.1 Hiking optimization algorithm

The Hiker Optimization Algorithm, proposed by Oladejo et al. [17], is a metaheuristic optimization algorithm inspired by hikers' strategies to ascend mountainous terrains while avoiding steep slopes to maintain hiking velocity. Its mathematical foundation stems from Tobler's Hiking Function (THF) [18], an exponential model quantifying terrain steepness to determine optimal traversal speed. The THF is defined as:

$$W_{i,t} = 6e^{-3.5|S_{i,t}+0.05|} \quad (1)$$

$$S_{i,t} = \frac{dh}{dx} = \tan \theta_{i,t} \quad (2)$$

where $W_{i,t}$ denotes the velocity of hiker i at time t , $S_{i,t}$ denotes the steepness of the terrain at the hiker i at time t , dh and dx denotes the change in elevation and horizontal distance traversed by the hiker, respectively, and $\theta_{i,t} \in [0, 50^\circ]$ corresponds to the slope angle of the terrain that hiker i at time t .

HOA updates its parameters based on the hikers' social mindset and their individual cognitive capabilities. The movement velocity of each hiker is governed by Tobler's hiking function, which incorporates three key factors: the leader's position in the hiker group, the current hiker's position, and scan factor (SF). The mathematical formulation is expressed as:

$$W_{i,t} = W_{i,t-1} + \gamma_{i,t}(\beta_{best} - \alpha_{i,t}\beta_{i,t}) \quad (3)$$

where $\gamma_{i,t} \in [0, 1]$ is a random variable, denote the control coefficient. β_{best} denote the position of the leader in the hiking group, $\beta_{i,t}$ denote the position of hiker i at time t . $\alpha_{i,t}$ denote the SF of hiker i . Specifically, $\alpha_{i,t}$ defined as a random integer within the interval $[1, 2]$. The SF regulates the distance between individual hikers and group leaders, where the specified range ensures that group members neither cluster excessively near the leader nor diverge too far from it. This configuration achieves dual objectives: maintaining search effectiveness while enabling individual hikers to visually perceive the leader's position and receive its guidance signals. Considering the velocity, the position update formula is:

$$\beta_{i,t+1} = \beta_{i,t} + W_{i,t} \quad (4)$$

2.1.2 Variational mode decomposition

Variational Mode Decomposition exhibits strong robustness and adaptive signal decomposition capabilities, with well-established applications in non-stationary signal processing, vibration signal analysis, and image processing [19]. Compared to conventional Empirical Mode Decomposition (EMD), VMD effectively mitigates mode aliasing and noise interference in signals. Given the nonlinear and non-stationary nature of fixed-based CPI index data, the VMD algorithm can efficiently extract trend and cyclical features while decomposing complex nonlinear sequences into multiple stationary components. The core principle of VMD involves decomposing raw signals into several intrinsic mode functions (IMFs) and a residual term through three steps: variational problem construction, solution derivation, and parameter optimization. The decomposition efficacy is evaluated using the energy entropy (He) of IMFs, where minimizing this entropy enhances decomposition precision and modal independence.

(1) Variational problem construction

Assuming the original CPI data consists of a finite number of intrinsic mode functions (IMFs), $u_k(t)$ denotes the decomposed components, formulate and solve the total bandwidth summation of $u_k(t)$, then converting it into a constrained variational problem. The procedure formula is:

$$\min_{\{u_k\}, \{w_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \quad (5)$$

$$s.t. \sum_{k=1}^K u_k(t) = f(t)$$

where $f(t)$ denote the original CPI data, K denote the number of decomposed modes, $\{u_k\}, \{w_k\}$ correspond to the set of decomposed intrinsic mode functions and their central frequency sequences, respectively. ∂_t denote the gradient operation, $\delta(t)$ denote the Dirac distribution, $\otimes$ denote the convolution operator, and j denote the imaginary unit.

(2) Solution derivation

To solve the variational extremum problem in Step (1), we introduce a quadratic penalty factor α to mitigate the effects of Gaussian smoothing and employ a Lagrangian multiplier λ to facilitate the transformation into an unconstrained variational problem. The converted formula is expressed as:

$$L(\{u_k\}, \{w_k\}, \lambda) \quad (6)$$

$$= \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k t} \right\|_2^2$$

$$+ \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle$$

where $\langle \cdot \rangle$ denote the inner product operator. The Alternating Direction Method of Multipliers (ADMM) is employed to resolve the variational problem, then iteratively updating the saddle point of the Lagrangian function, which corresponds to the optimal modal solutions. The iterative update formula is:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2}$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k(w)|^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \quad (7)$$

$$\hat{\lambda}^{n+1}(w) = \hat{\lambda}^n(w) + \tau(\hat{f}(w) - \sum_{k=1}^K \hat{u}_k^{n+1}(w))$$

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^{n+1}\|_2^2} < \varepsilon$$

where $\hat{f}(w)$, $\hat{u}_k(w)$ and $\hat{\lambda}(w)$ denote the Fourier transforms of $f(t)$, $u_k(t)$ and $\lambda(t)$ respectively. τ denote the noise tolerance. ε denote the convergence accuracy ($\varepsilon > 0$), the iteration process terminates when the relative error becomes smaller than the convergence accuracy, ultimately yielding K Intrinsic Mode Function (IMF) components.

(3) Parameter optimization

The modal number K significantly influences the decomposition performance of Variational Mode Decomposition [20]. Excessive modes may introduce spurious components, while insufficient modes might fail to adequately separate different frequency constituents within the signal. Through the Hybrid Optimization Algorithm, the optimal K is determined when the Energy Entropy (He) reaches its minimum. The He formula is:

$$He = - \sum_{j=1}^K p_j \lg p_j \quad p_j = \frac{E_j}{\sum_{j=1}^K E_j}, \quad E_j = \sum_{j=1}^N |X_j|^2 \quad (8)$$

where p_j denote the energy proportion of each Intrinsic Mode Function (IMF) relative to the total energy. E_j denote the energy of IMF _{j} . X_j denote the amplitude of IMF _{j} . To summarize, based on the iterative optimization framework of the HOA, the schematic diagram of the HOA-VMD is illustrated as Figure 1.

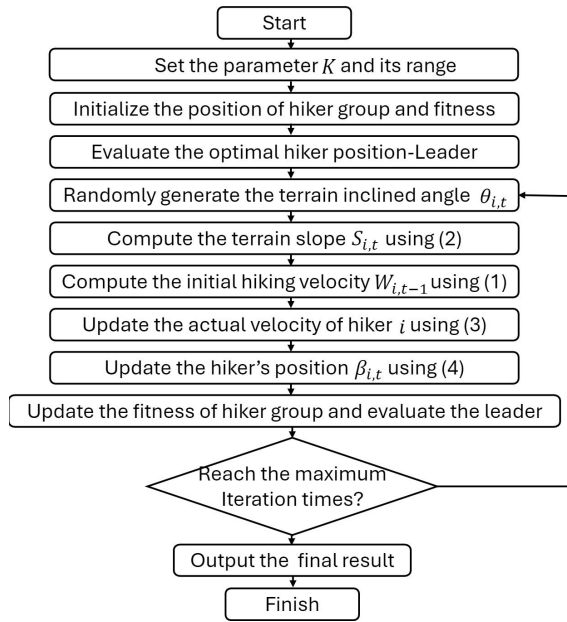


Figure 1. HOA-VMD Flowchart

2.2 Multilayer Perception

The Multilayer Perceptron (MLP), as a classical machine learning model, typically comprises an input layer, one or more hidden layers, and an output layer. Firstly, the input layer receives a set of input data, where each feature corresponds to a distinct weight. The input signals undergo a linear combination of these weights, followed by the addition of a bias term b , the formula is $\sum_{i=1}^n w_i x_i + b$. Second, the weighted summation result is transmitted to neurons in the hidden layer. Each neuron in this layer incorporates a nonlinear activation function, which applies a nonlinear transformation to its input. Then the transformed signal is subsequently propagated to neurons in the output layer, the formula is $f(\sum_{i=1}^n w_i x_i + b)$. Third, the activation function of the output layer converts the received data into the final output. The architectural diagram of the MLP framework is illustrated in Figure 2.

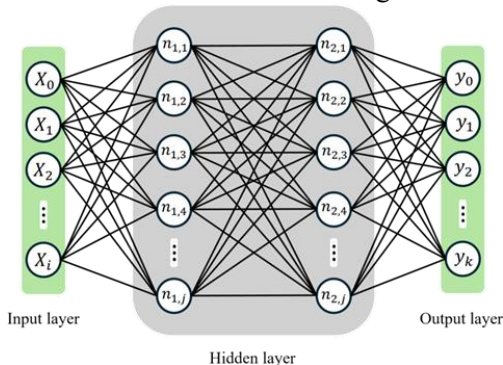


Figure 2. Structure of MLP

2.3 Bidirectional Long Short-Term Memory Network

2.3.1 Long short-term memory

Long Short-Term Memory networks (LSTM), an optimized variant of Recurrent Neural Networks (RNN), address the gradient vanishing and explosion issues inherent in traditional RNN models when processing lengthy input sequences. The LSTM architecture comprises recurrent memory units containing one or more cell structures (forget gate, input gate, output gate), as illustrated in Figure 3. Owing to its inherent memory mechanism, LSTM demonstrates superior capability in modeling temporal dependency characteristics of Consumer Price Index (CPI) data.

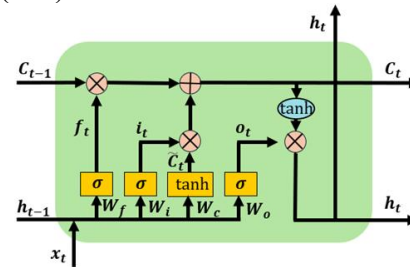


Figure 3. Structure of LSTM

2.3.2 Bidirectional long short-term memory network

Given that Long Short-Term Memory (LSTM) networks can only learn information along the forward temporal axis, while in time series prediction tasks such as Consumer Price Index (CPI) forecasting, both historical and future data may influence predictions at the current timestep, we employ a Bidirectional Long Short-Term Memory Network (BiLSTM) to comprehensively capture temporal features [21]. This architecture utilizes dual LSTM hidden layers for forward and reverse computations, thereby overcoming the limitations of unidirectional LSTM in learning comprehensive temporal information. The structure of the BiLSTM is illustrated in Figure 4.

For the forward LSTM, each timestep's input vector is processed to generate a forward hidden state sequence (Equation 9), which encodes all information from the sequence's starting position to the current timestep. Conversely, the backward LSTM produces a backward hidden state sequence (Equation 10) that encapsulates information from the sequence's endpoint to the current timestep. The final BiLSTM output is derived through joint computation of these bidirectional hidden state sequences.

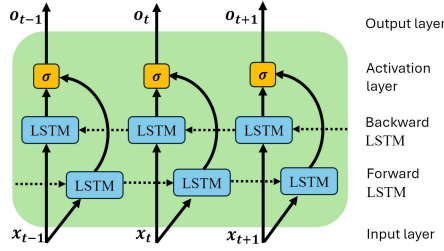


Figure 4. Structure of BiLSTM

$$\bar{h}_t = \overrightarrow{LSTM}(\bar{h}_{t-1}, x_t, \bar{B}_{t-1}) \quad (9)$$

$$\tilde{h}_t = \overleftarrow{LSTM}(\tilde{h}_{t+1}, x_t, \tilde{B}_{t+1}) \quad (10)$$

$$o = \sigma(M_k[\bar{h}_t, \tilde{h}_t] + \alpha_k) \quad (11)$$

where $\bar{h}_t$ and $\tilde{h}_t$ denote the forward and backward hidden sequence, respectively. o denotes the output of the BiLSTM. x_t denote the input value at timestep t . $\bar{h}_{t-1}$ denote the hidden state of the forward LSTM at the $t-1$ th input point, $\tilde{h}_{t+1}$ corresponds to the hidden state of the backward LSTM at the $(t+1)$ th input point. $\bar{B}_{t-1}$ and $\tilde{B}_{t+1}$ denote the cell states of the forward and backward LSTM at the $(t-1)$ th and $(t+1)$ th input point, respectively. σ denote the activation function, M_k denote the weight matrix, and α_k denote the bias vector.

2.4 Multi-step Forecasting

Multi-step ahead prediction involves utilizing historical data to forecast values over future times. Taking CPI prediction as an example, single-step prediction employs a $d \times K$ dimensional sequences (where K denote the number of decomposition modes, $K = 1$ if no decomposition is applied) prior to time N as input to predict the CPI value at $N + 1$. Multi-step prediction maintains identical input dimensions but extends the forecasting horizon to H subsequent steps. This study specifically adopts the direct strategy for multi-step value prediction.

2.5 Inverse Error Variance Method

The Inverse Error Variance Method indicates that a larger error sum of squares in a prediction model signifies poor model performance and corresponds to a smaller weight coefficient. Conversely, a smaller error sum of squares suggests better performance in forecasting tasks, which should be assigned a larger weight coefficient. After obtaining two different sets of predictions from the MLP and HOA-VMD-BiLSTM models, it is necessary to calculate the

weighted coefficients for these two prediction methods to achieve the optimal prediction outcome. The formula is:

$$\alpha_i = \frac{D_i^{-1}}{\sum_{i=1}^m D_i^{-1}} \quad i = 1, 2, \dots, m \quad (12)$$

$$D_i = \sum_{t=1}^N (y(t) - y_{ip}(t))^2 \quad t = 1, 2, \dots, N$$

where α_i denote the weight coefficient of the i th model, D_i denote the sum of the squared errors of the i th model, $y(t)$ denote the actual CPI value and $y_{ip}(t)$ denote the predicted value from the i th model. The predictive result of the final combined model is computed using the following formula:

$$y_p = \alpha_1 y_{1p} + \alpha_2 y_{2p} \quad (13)$$

2.6 Proposed Method

This study proposes a robust and novel hybrid model, termed HOA-VMD-BiLSTMMLP (Dynamic weight), abbreviated as HV-D(M-B), for forecasting the CPI. First, to address the limited sample size of CPI data, cubic spline interpolation was applied to the original time series, with the interpolated sequence subsequently serving as the new original sequence. Subsequently, considering the critical sensitivity of VMD to decomposition mode quantity, the HOA was employed to determine the optimal decomposition parameter K , thereby generating more stationary and independent subcomponents. Finally, to ensure robust and accurate predictions, a hybrid model learning architecture was implemented: the VMD-derived subcomponents were fed into a BiLSTM network to capture intrinsic patterns, while the original sequence was processed through a MLP to learn global characteristics; dynamic weight allocation between the two models was achieved through the inverse sum of squared errors method and then obtain the final predict value. Furthermore, given the importance of both short- and long-term CPI forecasting for policymakers to conduct comprehensive assessments, this research incorporates one-, five-, and nine-step-ahead predictions. The comprehensive framework is illustrated in Figure 5.

3. Experimental Design and Analysis

3.1 Data Processing

The dataset was sourced from the Belt and Road

Initiative module of the CSMAR (China Stock Market & Accounting Research) Database, comprising fixed-base Consumer Price Index (CPI) monthly data for China from January 2000 to July 2024. The experimental environment was configured with Python 3.12, PyTorch 2.6.0, CUDA 12.6, and an RTX 4060 GPU. The dataset was partitioned into training and test sets using an 8:2 split ratio, with chronological ordering preserved to maintain temporal integrity. The cubic spline interpolated CPI time series is shown in Figure 6, where the horizontal axis indicates weekly sequence indices and the vertical axis represents CPI values.

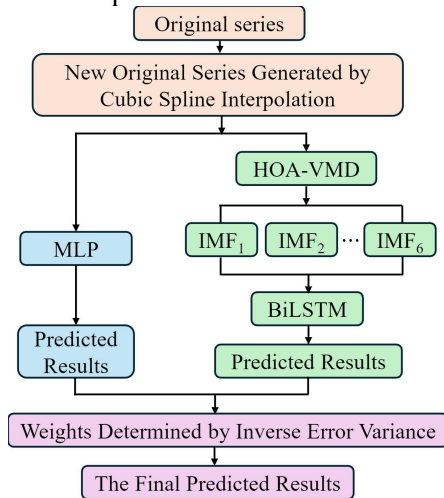


Figure 5. Decomposition-Prediction Process

Prior to model training, the data are normalized to the $[0,1]$ range through minmax scaling. This preprocessing step mitigates noise interference during neural network training, enhances the precision of parameter updates, and consequently stabilizes the model training process. The normalization procedure is mathematically formulated as follows:

$$\tilde{z}_i = (z_i - \min z_i) / (\max z_i - \min z_i) \quad (14)$$

where z_i denote the original input data and $\tilde{z}_i$ denote the normalized data.

The descriptive statistical analysis results of China's Consumer Price Index (CPI) are shown in table 1. The following conclusions can be drawn from the data presented: The Jarque-Bera (J-B) test statistic indicates that China's CPI rejects the null hypothesis of normal distribution at the 1% significance level. The Ljung-Box test results demonstrate that China's CPI rejects the null hypothesis of no autocorrelation at the 1% significance level, revealing that the series exhibits both 10th-order lag autocorrelation and long memory characteristics. Furthermore, the BDS test results show that China's CPI rejects

the independent and identically distributed null hypothesis at the 1% significance level, suggesting the presence of nonlinear dependencies in the data series.

Additionally, we conducted an Augmented Dickey-Fuller (ADF) test to examine the stationarity characteristics of China's Consumer Price Index. As shown in Table 2, the critical values at the 1%, 5%, and 10% significance levels are -3.4360, -2.8640, and -2.5681 respectively. The test statistic (-0.1207) exceeds all critical values and p-value greater than 0.05, leading us to fail to reject the null hypothesis that the series is non-stationary.

Based on the comprehensive analysis of Table 1 and Table 2, we conclude that China's Consumer Price Index exhibits non-normality, autocorrelation, nonlinear dependencies, and non-stationarity characteristics. These findings collectively reveal fundamental properties of the series and provide a comprehensive understanding of its statistical behavior.

Table 1. Results of Statistical Analysis of CPI of China

Index	CPI
Sequence Length	1177
Mean	106.0294
Standard deviation	17.8504
Maximum	133.5752
Minimum	79.2490
Jarque-Bera	0.96e ^{2***}
Q(10)	1.16e ^{4***}
BDS	2.15e ^{2***}

Note: The null hypothesis rejected at the significance levels of 10%, 5%, 1% is indicated by *, **, and ***, respectively. J-B is the Jarque-Bera statistic. Q(10) refers to the Ljung-Box statistic with sequence correlation up to 10 orders. BDS is the Brock-Dechert-Scheinkman statistic.

Table 2. ADF Test Results for CPI of China

	T statistic	Prob	The 1% T statistic	The 5% T statistic	The 10% T statistic
CPI	-0.1207	0.9473	-3.4360	-2.8640	-2.5681

3.2 Series Decomposition

To effectively address noise interference in forecasting, we adopted a hybrid method integrating Variational Mode Decomposition (VMD) with the Hiking Optimization Algorithm (HOA) to enhance the accuracy and robustness of decomposition results. The HOA parameters were configured with a population size of 30 and maximum iterations of 25. The convergence process of HOA illustrated in Figure 7, which

shows that fitness stabilized after the 2nd generation, indicating algorithm convergence to optimal solutions. And decomposition results are presented in Figure 8. Furthermore, to ensure model stability and mitigate noise impacts on predictive performance, residual components from the decomposition were discarded. The resulting intrinsic mode functions (IMFs 1-6) were subsequently fed into the BiLSTM network for model training.

3.3 Hyperparameter Configuration

During model training, window size, epochs, batch size, and neuron count significantly impact model prediction performance. Specifically, the window size refers to the length of input data when predicting the next value, determining the model's ability to capture short-term or long-term dependencies in time series. Epochs indicate how many times the training data is fully traversed, where an appropriate epoch allows gradual learning of data patterns to enhance predictive capability. Batch size represents the number of samples used per training step, with proper batch sizes improving computational efficiency, reducing training oscillations, and increasing gradient descent direction accuracy. Neuron count denotes the number of nodes in each neural network layer, where suitable neuron quantities mitigate overfitting risks while enhancing model expressiveness. Through the inverse sum of squared errors method, the weights for MLP and BiLSTM in the proposed ensemble model were determined as 0.1 and 0.9, respectively. The hyperparameters of the proposed model are detailed in Table 3.

3.4 Comparative Analysis of Experimental Results

In this paper, we developed a novel ensemble model named HOA-VMD-MLP-BiLSTM (Dynamic weight), abbreviated as HV-D(M-B), to predict China's Consumer Price Index. The weights of MLP and BiLSTM were calculated as 0.1 and 0.9 respectively using the inverse sum of squared errors method. This model was compared with six baseline models: HOA-VMD-MLP-BiLSTM (Fixed weight), abbreviated as HV-F(MB); HOA-VMD-MLP-LSTM (Dynamic weight), abbreviated as HV-D(M-L); MLP; BiLSTM; HOA-VMD-BiLSTM, abbreviated as HV-B; HOA-VMD-LSTM, abbreviated as HV-L. Additionally, we

conducted 1-, 5-, and 9-step forecasting across all models to provide comprehensive long-term predictions. This approach delivers precise guidance for policymakers to assess future trends and better respond to market fluctuations.

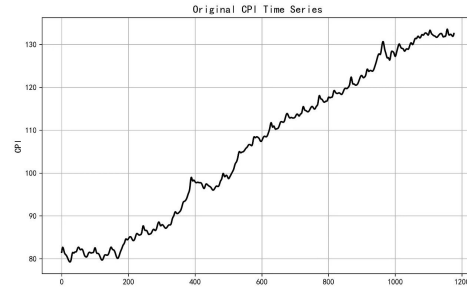


Figure 6. Cubic Spline Interpolated CPI Time Series

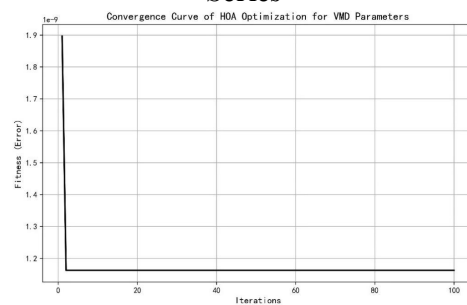


Figure 7. Convergence Diagram of the HOA for VMD

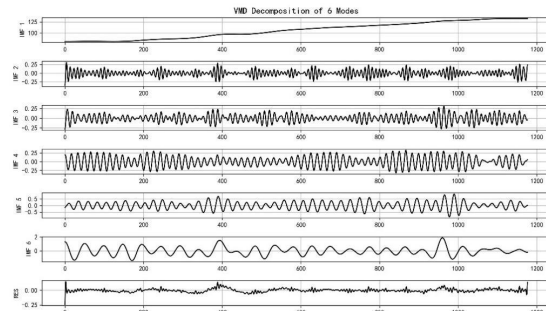


Figure 8. HOA-VMD Decomposes the CPI Data

3.4.1 Ablation study

For model evaluation, the predictive performance was assessed using loss functions (RMSE, MAE, MAPE) calculated from forecasting results. These metrics provide an intuitive understanding of model prediction capabilities, thereby demonstrating the effectiveness of the proposed approach. Error values closer to zero indicate smaller prediction errors and superior model performance. The mathematical definitions of these three loss functions are shown below:

$$RMSE = \sqrt{\sum_{t=1}^N (y_t - \hat{y}_t)^2 / N} \quad (15)$$

$$MAE = \sum_{t=1}^N |y_t - \hat{y}_t| / N \quad (16)$$

$$MAPE = 100\% * \sum_{t=1}^N |(y_t - \hat{y}_t) / y_t| / N \quad (17)$$

where y_t denote the true value at time t , $\hat{y}_t$ denote the predicted value, and N denote the number of samples in the test set.

Additionally, this study utilizes the coefficient of determination R^2 to assess the goodness-of-fit between predicted values and true values. Values of R^2 closer to 1 indicate a stronger model fit. The mathematical expression for R^2 is shown below:

$$R^2 = 1 - \sum_{t=1}^N (y_t - \hat{y}_t)^2 / \sum_{t=1}^N (y_t - \bar{y})^2 \quad (18)$$

where $\bar{y}$ denote the mean value of original test set.

The loss function values of each model for 1-, 5-, and 9-step forecasting are shown in Table 4. By analyzing the tabular data, the following conclusions can be drawn.

The comparative analysis of 1-step and 9-step forecasting demonstrates superior precision in decomposition-based approaches over non-decomposition forecasting. This evidence

confirms that preprocessing raw data through HOA-VMD decomposition enhances model learning capacity, thereby improving predictive performance. Specifically, HOA-VMD-BiLSTM outperforms HOA-VMD-LSTM, indicating BiLSTM's superior predictive capability in this research scenario. Furthermore, the implementation of dynamic weighting mechanisms enhanced predictive accuracy compared to fixed-weight configurations.

For 5-step forecasting, while HOA-VMD-BiLSTM emerges as the optimal model, our proposed methodology achieves second-best performance with closely comparable results.

The differences in four loss functions among all models for 1-step forecasting are shown in Figure 9. The proposed model achieves the smallest RMSE, MAE, and MAPE values (closest to zero), while R^2 attains the largest value (closest to 1) compared to other models. This demonstrates the proposed model's accuracy and effectiveness in forecasting China's Consumer Price Index.

Table 3. Parameters of the Proposed Model

	Window size	Epochs	Batch size	Neuron number			
				Hidden layer 1	Hidden layer 2	Hidden layer 3	Dense layer
BiLSTM	48	100	32	128	64	None	64
MLP	48	100	32	128	64	32	None

Table 4. Four Loss Function Results for Each Model

Model	1-step				5-step				9-step			
	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
MLP	0.8797	0.5683	0.4899	0.37%	0.7105	0.8563	0.7237	0.55%	0.6785	0.8535	0.7392	0.56%
BiLSTM	0.8455	0.7533	0.7078	0.54%	0.7203	1.0065	0.8052	0.62%	0.6382	1.1492	0.9098	0.70%
HV-B	0.9806	0.2284	0.2137	0.16%	0.9813	0.2175	0.1719	0.13%	0.9054	0.4629	0.3870	0.29%
HV-L	0.9449	0.3846	0.3467	0.26%	0.9227	0.4426	0.3701	0.28%	0.8303	0.6201	0.5220	0.40%
HV-D(M-L)	0.9887	0.1742	0.1440	0.11%	0.9638	0.3028	0.2539	0.19%	<u>0.9384</u>	<u>0.3735</u>	<u>0.3162</u>	<u>0.24%</u>
HV-F(M-B)	0.9673	0.2964	0.2503	0.19%	0.9260	0.4329	0.3654	0.28%	0.8698	0.5431	0.4505	0.345
HV-D(M-B)	0.9964	0.0979	0.0838	0.06%	0.9776	<u>0.2379</u>	<u>0.1922</u>	<u>0.15%</u>	0.9411	0.3653	0.2878	0.22%

Note: boldface indicates optimal values and underlined indicates suboptimal values.

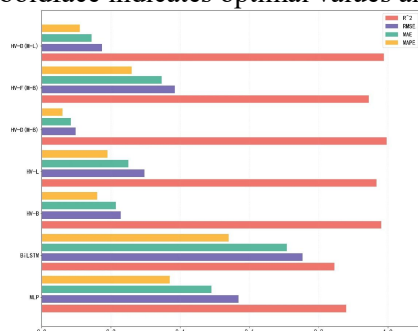


Figure 9. Comparison of Four Loss Function for Each Model in 1-step Prediction

3.4.2 Diebold-Mariano test

This study employs the Diebold-Mariano test to demonstrate predictive capability differences

between HOA-VMD-MLP-BiLSTM (Dynamic weight) and comparative models. The test's core principle examines whether the mean value of loss differential series between prediction errors significantly deviates from zero, thereby determining the statistical significance of predictive performance variations across models. Assuming model A and model B generate forecasts form time series $\{y_t\}_{t=1}^T$, denote as $\{\hat{y}_{A,t}\}_{t=1}^T$ and $\{\hat{y}_{B,t}\}_{t=1}^T$, with corresponding forecast errors $e_{A,t} = y_t - \hat{y}_{A,t}$ and $e_{B,t} = y_t - \hat{y}_{B,t}$. The loss differential sequence is defined as $d_t = L(e_{A,t}) - L(e_{B,t})$, where $L(\cdot)$ denote the loss

function (e.g., Mean Squared Error (MSE) or Mean Absolute Percentage Error (MAPE)). The Diebold-Mariano test establishes the null hypothesis $H_0: E[d_t] = 0$ that no significant difference exists in predictive accuracy between the two models. The test statistic is formulated as:

$$DM = \frac{\bar{d}}{\sqrt{\hat{s}_d / T}} \quad (19)$$

where $\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t$ denote the sample mean, $\hat{s}_d$ denote the sample variance estimator of the differential sequence. Under weak stationarity and appropriate moment conditions, the DM statistic asymptotically follows the standard normal distribution. The null hypothesis is rejected when $|DM| > z_{\alpha/2}$.

The Diebold-Mariano (DM) test results between the proposed model and comparative models for 1-, 5-, and 9-step forecasting are shown in table 5. All models reject the null hypothesis at the 1% significance level across different predicted steps, which statistically validates the accuracy and reliability of the proposed methodology.

3.4.3 Comparative analysis of multi-step predictions in the proposed model

Furthermore, the comparative analysis of the multi-step prediction performance for the HOA-VMD-MLP-BiLSTM model with dynamic weighting across 1-, 5-, and 9-step forecasting are shown in Figure 10. As quantitatively demonstrated in Table 1, the proposed model achieves RMSE values of 0.0979, 0.2379, and

0.3653 for 1-, 5-, and 9-step predictions, respectively; MAE values of 0.0838, 0.1922, and 0.2878, respectively; MAPE values of 0.06%, 0.15%, and 0.22%, respectively; and R2 values of 0.9964, 0.9776, and 0.9411, respectively. Notably, the 1-step prediction exhibits optimal performance, with error indices (RMSE, MAE, MAPE) approaching zero and R^2 values closer to 1. At the same time, through visual comparison of multi-step prediction results, it can also be observed that the proposed model's 1-step predictions are overall closer to the true values.

3.4.4 1-step prediction comparison across all models

Finally, to visually demonstrate the predictive performance of each model, we plot the 1-step prediction results of all models as shown in figure 11. The horizontal axis represents the sample size of the test set, and the blue line with circular markers denotes the predictions of the HOA-VMD-MLP-BiLSTM (Dynamic weight) model. The figure reveals that the results of our proposed model more closely match the original data. However, the predictions from other models exhibit larger deviations from the original data. Therefore, we conclude that the proposed model demonstrates superior predictive performance.

This study validates the superiority of the HOA-VMD-MLP-BiLSTM (Dynamic weight) model through four loss functions, Diebold-Mariano tests, and multi-step prediction results, with its advantage being obvious in one-step predictions.

Table 5. DM Test Results of HOA-VMD-MLP-BiLSTM (Dynamic Weight) and Other Models

	1-step			5-step			9-step		
	DM ₁	DM ₂	DM ₃	DM ₁	DM ₂	DM ₃	DM ₁	DM ₂	DM ₃
MLP	6.52e-20***	1.43e-35***	1.57e-35***	7.93e-22***	1.97e-39***	1.42e-38***	1.44e-14***	6.88e-29***	1.82e-28***
BiLSTM	5.03e-20***	1.67e-30***	1.88e-30***	2.83e-19***	1.07e-30***	4.52e-30***	2.87e-18***	6.33e-28***	8.13e-28***
HV-B	7.33e-17***	2.06e-22***	2.14e-22***	1.23e-16***	2.10e-25***	1.80e-25***	1.84e-13***	8.01e-18***	1.02e-17***
HV-F(M-B)	1.55e-08***	2.68e-09***	2.65e-09***	5.03e-07***	1.89e-07***	1.37e-07***	5.27e-07***	1.44e-06***	1.19e-06***
HV-L	6.57e-20***	1.07e-30***	1.08e-30***	5.98e-17***	4.53e-33***	6.66e-33***	2.42e-18***	2.10e-32***	1.00e-31***
HV-D(M-L)	2.56e-14***	3.98e-19***	4.12e-19***	2.02e-14***	1.93e-20***	2.19e-19***	1.58e-13***	8.81e-16***	3.40e-15***

Note: the DM test statistics of MSE, MAPE and MAE loss functions are expressed by DM₁, DM₂ and DM₃ respectively. Rejection of null hypothesis at the significance levels of 10%, 5%, 1% is indicated by *, **, and ***, respectively.

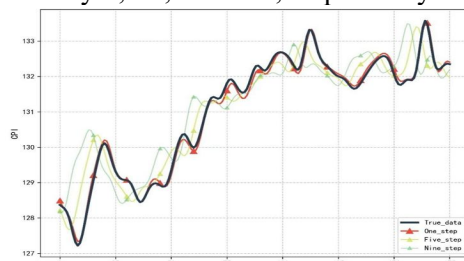


Figure 10. Multi-step Prediction Results of the Proposed Model

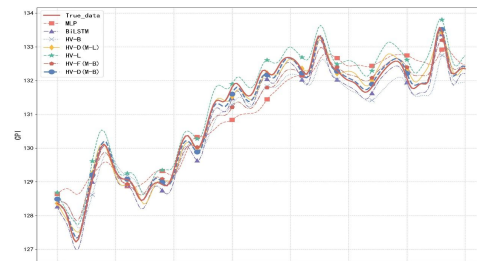


Figure 11. 1-step Prediction Results for Each Model

4. Conclusion

Accurate prediction of the Consumer Price Index is a critical tool for governments to anticipate inflation, for policymakers to implement rational economic regulations, and for investors to evaluate market stability. To provide decision-makers with accurate and robust support, therefore, we propose a robust hybrid model, the HOAVMD-MLP-BiLSTM (Dynamic weight) model, for forecasting China's Consumer Price Index. This model utilizes a novel and efficient optimization algorithm to determine the mode number of Variational Mode Decomposition (VMD), feeds the decomposed mode functions and original series into BiLSTM and MLP models respectively to obtain predictions, and dynamically integrates the strengths of both models through adaptive weighting. Let the global trends learned by the MLP guide the detailed features captured by the BiLSTM, making the model achieve optimal predictive performance. Based on the research findings, the HOA-VMD-MLP-BiLSTM (Dynamic weight) model consistently outperforms other models in the following aspects: its RMSE, MAE, and MAPE values are closest to 0, and R^2 values are closest to 1 in both 1-step and 9-step predictions. For 5-step predictions, these metrics rank second-best, with minimal gaps from the optimal values. Moreover, it maintains high accuracy even in 9-step predictions, achieving the objective of developing models with high precision and robustness. Furthermore, Diebold-Mariano tests confirm the model's statistically significant superiority across all models, demonstrating its reliability as an economic forecasting tool.

In conclusion, this study holds vital significance for economic forecasting for governments, enterprises, and individuals. Firstly, the multi-step prediction model proposed herein enables comprehensive analysis of data to reveal long-term trends, providing decision-makers with reliable information to support economic regulation, business expansion and contraction strategies, and reduction of subjective misjudgments. Furthermore, the HOA-VMD-MLP-BiLSTM (Dynamic weight) model is applicable not only to Consumer Price Index (CPI) forecasting but also to other datasets with similar patterns, particularly CPI data from The Belt and Road countries, which share analogous characteristics. Finally, despite current global

economic stability, persistent uncertainties necessitate accurate predictions to optimize resource allocation. Meanwhile, this research has limitations: the model exclusively utilizes raw CPI data without incorporating related economic indicators (e.g., inventory, international crude oil prices, interest rates, Producer Price Index) that could enhance predictive accuracy. To address this, future work will blend multiple time series data sources.

Acknowledgments

This paper is supported by Henan Province Graduate Education Reform and Quality Improvement Project (YJS2025GZZ14).

References

- [1] Zhang, Y. Evaluation of Influencing Factors of National Engineering Project Cooperation along the 'Belt and Road' Based on Entropy Weight TOPSIS. *Science Technology and Industry*, 2025, 25(05): 216-223.
- [2] Alvarez-Diaz, M. and Gupta, R. Forecasting US Consumer Price Index: Does Nonlinearity Matter. *Applied Economics*, 2016, 48(46): 4462-4475.
- [3] Sun, H. Comparative Analysis of China's CPI Dynamics under Two Financial Crises. *Statistics & Decision*, 2010, 26(1): 116-118.
- [4] Lu, S., Zhao, B., Bi, J. CPI Forecasting Integrating Fuzzy Information Granulation and SVM Time Series Regression. *Statistics & Decision*, 2015, 14: 82-84.
- [5] Zeng, L. CPI Prediction Based on ARIMA-RF Combined Model. *Modern Information Technology*, 2023, 7(13):13-17.
- [6] Sarangi, P.K., Sahoo, A.K., Sinha, S. Modeling Consumer Price Index: A Machine Learning Approach. *Macromolecular symposia*, 2022, 40(1): 2100349.
- [7] Wen, X. and Xu, Y. EMD-based Analysis and Predictive Modeling of Wind Turbine Power Characteristics. *Acta Energaie Solaris Sinica*, 2021, 42(11): 293-298.
- [8] Tai, X. and Liu, Y. Monthly CPI Forecasting using a Hybrid EEMD-PSO-SVM Model. *Statistics & Decision*, 2019, 3: 30-33.
- [9] Ying, H. and Wang, S. CPI Forecasting using a Hybrid EEMD-BP Model. *Contemporary Economics*, 2022, 39(10): 97-106.
- [10] Fang, Z., Yang, H., Deng, Q., Fang, H., Zhuang, Y., Lin, S., Yao, W. Dissolved

- Oxygen Prediction Research based on Correlation Optimization with the Coupled LSTM Model. *Safety and Environmental Engineering*, 2025: 1-17.
- [11] Zhou, Y., Peng, J., Bai, Y. Data-driven Time Series Feature Extraction and Prediction of Daily Blood Collection Volumes. *Journal of Systems Science and Mathematical Sciences*, 2025: 1-22.
- [12] Dong, M. and Tang, X. Study on CPI Prediction by LSTM Model based on Double-layer Attention Mechanism. *Chinese Journal of Management Science*, 2025: 1-14.
- [13] Dragomiretskiy, K. and Zosso, D. Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, 2013, 62(3): 531-544.
- [14] Zhang, Y., Zhao, Y., Kong, C., Chen, B. A New Prediction Method based on VMD-PRBF-ARMA-E Model Considering Wind Speed Characteristic. *Energy Conversion and Management*, 2020, 203: 112254.
- [15] Feng, Y., Zeng, H., Deng, H., Tu, P. A Step-type Landslide Displacement Prediction Model based on Creep Trend Influence and Feature Optimization Algorithm. *Chinese Journal of Rock Mechanics and Engineering*, 2025, 44(03): 705-720.
- [16] Zhou, W., Li, J., Cai, Y., Zheng, Z., Wen, L. Step-like Landslide Displacement Prediction using Bovmd-p-boxgboost. *South-to-North Water Transfers and Water Science & Technology*, 2025: 1-19.
- [17] Oladejo, S.O., Ekwe, S.O., Mirjalili, S. The Hiking Optimization Algorithm: A Novel Human-based Metaheuristic Approach. *Knowledge-Based Systems*, 2024, 296: 111880.
- [18] Goodchild, M.F. (2020). Beyond tobler's hiking function. *Geographical Analysis*, 52(4), 558-569.
- [19] Li, H., Xiao, B., Jin, K., Chen, Z., Yu, L. Construction and Comparative Analysis of Water Quality Prediction Model of the Sanmenxia Reservoir of the Yellow River. *Environmental Engineering*, 2024, 42(12): 1-7.
- [20] Zhang, F., Zhang, Q., Qin, B., Zhu, G., Deng, X., Liu, D. Bearing Life Prediction based on VMD and TCN-SeNet-BiLSTM Networks. *Machine Tool & Hydraulics*, 2025, 53(1): 15-23.
- [21] Zhang, J., He, J., Cai, Q., Kang, Z., Yang, Y., Wang, T., Yang, L. Prediction of Cod Concentration in Wastewater Treatment Plant Effluent based on Secondary Decomposition and BiLSTM. *CIESC Journal*, 2025: 1-16.