Stock Closing Price Interval Prediction Based on CEEMDAN-WTD-Bilstm-Transformer Model

Xinhong Wu[1, #], Chuchu Sun[2, #, *], Xingyu Hao[3, #]

[1]School of Management, Guangdong University of Technology, Guangzhou, China

[2]College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China

[3]Sydney Smart Technology College, Northeastern University, Qinhuangdao, China

*Corresponding author: sunchuchu0331@163.com, wuxinhong2203@163.com, haoxingyu2231@163.com

[#]These authors contributed equally.

Abstract. Stock price indices usually contain much noise and exhibit significant non-linearity, non-smoothness, and other complex characteristics. To address the complexity and long-term dependence of financial time series forecasting, a novel stock interval forecasting model based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Analysis (CEEMDAN)-wavelet threshold de-noising (WTD) joint de-noising method and Transformer model improved by a Bi-directional Long Short-Term Memory model (BiLSTM) is proposed in this paper. First, based on CEEMDAN, the WTD method is incorporated to extract features from the basic ticker data of financial time series and use them as the input of the Transformer model. Meanwhile, the BiLSTM provides the time-series relationship among the input stock data for the Transformer to forecast the financial time series. Moreover, a hybrid sliding Kernel Density Interval Estimation (KDE) method to demonstrate the forecasting accuracy of financial time series. Using stock price index data, the prediction results are compared with those of traditional neural networks, and the results show that the stock interval prediction of the CEEMDAN-WTD-BiLSTM-Transformer model has a relatively high prediction accuracy, which achieves an improvement of 0.88%~3.62%.

Keywords: BiLSTM; interval prediction; CEEMDAN; wavelet threshold de-noising; stock indices.

1. Introduction

With economic development, financial markets occupy an important position in the economic systems of various countries. Accurate and stable financial forecasting models are crucial for investors to avoid risks and develop profitable investment strategies. Therefore, forecasting financial time series has important research significance. In recent years, many fields, such as academia and investment, have paid increasing attention to financial market research and tried to use current advanced computer technology to solve the information demand problem in financial markets. The research on stock-related forecasting theories and methods has gradually become a hot spot for domestic and international research. Given the volatility of stock price indices, it usually contains a large amount of noise and presents significant nonlinear, non-stationary, and other complex characteristics. It has become a challenging research topic for researchers to understand the stock market trends and asset price movements through the complex phenomena and obtain sufficient information that supports stock forecasting and optimizes investment decisions [1]. In addition, comparative analysis with traditional point-value AR models, VAR models, and model Naïve are found that interval data models have higher forecasting accuracy and smaller forecasting errors for both interval high and interval low. The statistical significance, so a new approach to financial data forecasting, interval-based time series models, as mentioned [2].

Traditional stock investment models generally assume that the prediction sequence is linear or approximately linear, which makes it difficult to predict the price movement of stock indices efficiently and accurately. Compared with econometric methods and machine learning, deep learning uses layer-by-layer feature extraction with unsupervised learning, which has a more robust feature representation and can learn more complex functional representations. In addition, deep learning is easier to alleviate the overfitting problem and has a stronger generalization ability while improving the prediction accuracy of in-sample data [3]. Hochreiter proposed the Long Short-Term Memory model (LSTM) [4] for the problems of traditional Recurrent Neural Networks (RNN), which is a special RNN structure that has been proven to be superior to RNN models in many areas. Researchers have found that LSTM models also have better results for predicting temporal data [5-7]. Ashish et al. constructed a Transformer model based on a multi-headed self-attentive mechanism [8], which brings high accuracy and efficiency compared to traditional sequence models such as recurrent neural networks due to the design of the attention mechanism that captures remote contextual information in the entire sequence, and its deep structure significantly improves the feature extraction capability. To make up for the shortcomings of single models, scholars found that optimization algorithms can be added to machine learning models to make the prediction better. Yi-Yue He et al. [9] combined Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Analysis (CEEMDAN), and LSTM models for forecasting five representative stock indices of the CSI 300 results showed that the models have low error and lag. Yan Hongju [10] used Convolution Neural Network (CNN)and Gate Recurrent Unit (GRU) models for the combined prediction of the SSE index, and the results showed that the model prediction results were higher than deep learning and machine learning algorithms. Li Yafeng [11] et al. proposed an improved LSTM model by adding an attention mechanism to the LSTM layer to improve the prediction effect of the neural network, and the model parameters were tuned by a Genetic Algorithm (GA) to improve the generalization ability of the model. Yan et al. [3] proposed a combination of LSTM neural network and wavelet analysis to build a financial time series forecasting model to eliminate noise and improve the model's prediction ability for out-of-sample financial time series.

In this paper, a novel stock interval forecasting model based on the CEEMDAN-wavelet threshold de-noising (WTD) joint de-noising method and Transformer model improved by a Bi-directional LSTM (BiLSTM) is proposed for the dynamic instability and long-term dependence of financial time series, which improves the decomposition noise reduction efficiency and reduces the loss of useful high-frequency signals through the CEEMDAN-WTD double-layer decomposition model, and then uses the improved Transformer model for interval prediction of financial time series to improve the prediction accuracy. The specific advantages and innovations are as follows:

(1) The CEEMDAN and WTD method are combined to decompose the time series data for noise reduction.

(2) A Bi-LSTM neural network replaces the position encoding in the original Transformer model, which extracts the time-series features of the input data based on the entire time series, providing the Transformer with the time series relationships between the input stock data. Through that, the time series information in the stock series can be better extracted by Transformer.

(3) The hybrid sliding Kernel Density Interval Estimation (KDE) method is used for stock price interval prediction.

(4) Compared with other neural network models, results show that the prediction accuracy of the Transformer-based prediction model is high.

This paper begins with a brief introduction to the components of the model combination: the CEEMDAN decomposition technique, the wavelet threshold de-noising method, the Transformer model improved by BiLSTM, and the hybrid sliding KDE method. Based on the above discussion, this paper proposes the forecasting framework and applies it to the interval prediction of the closing price of one of the major stock indices in China, the Shanghai Stock Exchange Index. The model results were then compared with other traditional models, leading to the conclusion.

2. Methodology

2.1 Ceemdan-WTD Double-Layer Denoising Model

CEEMDAN Decomposition Method

EMD has achieved outstanding results in non-stationary and nonlinear signal processing [12], but there is mode mixing in signal smoothing [13]. EEMD adds noise based on EMD to offset the mode mixing phenomenon [14], but this method has a long calculation time and large reconstruction error. CEEMDAN and its improved method add the IMF component of white noise at each decomposition. The added noise is reduced step by step, and there is less residual noise in the intrinsic mode component, which effectively reduces the reconstruction error. At each stage of decomposition, there is a global stopping criterion, so the decomposition is most efficient [15].

The implementation steps of CEEMDAN are as follows:

Step1: Firstly, white noise $\varepsilon_0 w_i(n)$ is added to the original signal $x(n)$, and the obtained signal $X_i(n)$ can be expressed as:

$$X_i(n) = x(n) + \varepsilon_0 w_i(n). \tag{1}$$

Equation (1), i is the number of times Gaussian white noise added, $i = 1,2,\cdots,N$. Firstly, the first-order IMF component IMF_1 is calculated, and the EMD decomposition is performed on $X_i(n)$ with white noise at the i order to obtain the IMF_{i1}. The results obtained by N decomposition are integrated and averaged to obtain the first-order IMF_1:

$$IMF_1 = \frac{1}{N} \sum_{N}^{i=1} IMF_{i1}. \tag{2}$$

At the same time, the first-order residual is:

$$r_1(n) = x(n) - IMF_1. \tag{3}$$

Setp2: A set of new time series $r_1(n) + \varepsilon_1 E_1(w_i(n))$is constructed by adding some auxiliary white noise in the residual $r_1(n)$. The time series is processed by EMD to obtain IMF_{i2}. The process is repeated N times, and the IMF with order 2 and the residual term is calculated by ensemble average. $E_j(\cdot)$ is denoted as the j_{th} IMF component obtained by EMD. The formulas are as follows:

$$IMF_2 = \frac{1}{N} \sum_{N}^{i=1} E_1\{r_1(n) + \varepsilon_1 E_1[w_i(n)]\}, \tag{4}$$

$$r_2(n) = r_1(n) - IMF_2. \tag{5}$$

Step3: A set of new time series $r_{k-1}(n) + \varepsilon_{k-1} E_{k-1}[w_i(n)]$ is constructed by adding auxiliary white noise on the basis of residual $r_{k-1}(n)$. EMD is used to process the time series to obtain the IMF_{ik}. Repeat this process N times, and the results are integrated to obtain the IMF with order k and the residual term. The formulas are as follows:

$$IMF_{ik} = \frac{1}{N} \sum_{N}^{i=1} E_1\{r_{k-1}(n) + \varepsilon_{k-1} E_{k-1}[w_i(n)]\}, \tag{6}$$

$$r_k(n) = r_{k-1}(n) - IMF_k. \tag{7}$$

Step4: Repeat step 3 until the residual cannot be decomposed, the original signal $x(n)$ is decomposed into:

$$x(n)=\sum_{L}^{k=1} IMF_k+R(n). \tag{8}$$

Where L is the maximum decomposition order of $x(n)$, and $R(n)$ is the final residual.

2.1.1 Wavelet Threshold De-noising

The wavelet threshold de-noising method is a multi-scale signal analysis method, which is one of the common methods for non-stationary signal processing [16]. The basic idea is: After the wavelet transform of the signal, the noise is suppressed by setting a suitable threshold to retain wavelet coefficients larger than the threshold and remove those smaller than the threshold [17]. The complicated thresholding noise reduction method tends to cause signal reconstruction oscillation and produce the pseudo-Gibbs phenomenon. The soft threshold method has a constant deviation from the original wavelet coefficient after de-noising, reducing the similarity between the denoised and original signals [18]. Therefore, this paper chooses the soft and hard threshold compromise method to process the wavelet coefficients, which makes up for the shortcomings of the two methods.

The basic operation steps are as follows:

Step1: Select the appropriate wavelet basis function and decomposition layers to decompose the signal into a series of wavelet coefficients.

Step2: Appropriate threshold function is selected to process the wavelet coefficients obtained by decomposition, and the wavelet coefficients with smaller thresholds are deleted as noise.

Step3: The denoised signal is obtained by inverse wavelet transform of the processed wavelet coefficients.

After using this algorithm to process the signal collected from the experiment, it is found that the sym8 wavelet is ideal for five-layer decomposition.

The threshold function is:

$$\tilde{w}_{j,k}=\begin{cases} w_{j,k}-a\lambda, w_{j,k}\geq\lambda \\ 0, |w_{j,k}|<\lambda \\ w_{j,k}+a\lambda, w_{j,k}\leq-\lambda \end{cases}. \tag{9}$$

In formula (9) $W_{j,k}$ is a wavelet coefficient, λ is a threshold, $a=0.9$.

The threshold formula selected is:

$$T=\frac{\mathrm{median}(cD1)}{0.6745}\sqrt{2\ln N}. \tag{10}$$

2.1.2 CEEMDAN-WTD Combined Noise Reduction

The CEEMDAN method decomposes the original signal into a series of IMF components from high frequency to low frequency. The traditional method first selects the demarcated IMF components by the correlation coefficient criterion, then decomposes the signal into high-frequency, and low-frequency IMF components, and finally rejects the high-frequency IMF components directly as noise signals for noise reduction. This method results in the loss of valuable signals at high frequencies and residual noise in the IMF component used to reconstruct the signal [19]. Therefore, CEEMDAN and wavelet threshold noise reduction are combined.

Considering the complexity of determining the delimitation IMF by the correlation coefficient criterion and the uncertainty of using a single index [20], this paper adopts the evaluation index T, which combines the root mean square error with the smoothness. It considers the detailed information and approximation information of the signal to determine the K value of the boundary IMF component [21]. The residual term is regarded as the last IMF component to facilitate the expression. The formulas are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{N-1}^{t=0}\left(IMF_k(t) - x(t)\right)^2}, \tag{11}$$

$$r = \frac{\sum_{n-2}^{t=0}\left(IMF_k(t+1) - IMF_k(t)\right)^2}{\sum_{n-2}^{t=0}\left(x(t+1) - x(t)\right)^2}. \tag{12}$$

In the formula, $IMF_k(t)$ is the k_{th} IMF component and $x(t)$ is the original data sequence.

Normalize these two indicators:

$$P_{RMSE} = \frac{RMNE}{\Sigma(RMNE)}, \tag{13}$$

$$P_r = \frac{r}{\Sigma(r)}. \tag{14}$$

These two indicators need to be weighted, and the weighting formulas [22] are:

$$CV_{P_{\mathrm{RAHSE}}} = \frac{\sigma_{P_{\mathrm{ROSSE}}}}{\mu_{P_{\mathrm{RHSSE}}}}, \tag{15}$$

$$CV_{P_r} = \frac{\sigma_{P_r}}{\mu_{P_r}}, \tag{16}$$

$$W_{P_{\mathrm{RHSE}}} = \frac{CV_{P_{\mathrm{RASE}}}}{CV_{P_{\mathrm{RSSS}}} + CV_{P_r}}, \tag{17}$$

$$W_{P_r} = \frac{CV_{P_r}}{CV_{P_{\mathrm{riss}}} + CV_{P_r}}. \tag{18}$$

In the formula, σ and μ are standard deviation and mean, CV is coefficient of variation, W is weight.

The expression of composite index T is:

$$T = W_{P_{\mathrm{RASE}}} \times P_{RMSE} + W_{P_r} \times P_r. \tag{19}$$

The expression of signal and noise threshold is:

$$R_{k-1} = \left|\frac{T_{k-1}}{T_k}\right|. \tag{20}$$

Where $k \geq 2$, when the threshold value is $1 \leq R_{k-1} \leq 3$ for the first time, it is determined as the boundary point. The first k IMF components are regarded as noise, and the IMF components after k are regarded as signals [21].

The specific process of the improved algorithm is shown in Fig. 1.

Step1: The CEEMDAN algorithm decomposes the original data sequence and obtains the IMF components with frequency from high to low.

Step2: The k value of the boundary IMF component is obtained by calculating the index T, and the IMF component is decomposed into signal and noise.

Step3: The filtered noise IMF component is processed by wavelet threshold de-noising.

Step4: The noise signal after threshold de-noising is reconstructed with the signal IMF component to obtain the denoised signal.

2.2 Models for Forecasting

2.2.1 Transformer

Ashish et al. constructed a Transformer model based on a multi-headed self-attentive mechanism [23]. Compared with traditional sequence models such as recurrent neural networks, the design of the attention mechanism allows the Transformer to capture remote contextual information throughout the sequence, bringing high accuracy and efficiency, while its deep structure significantly improves the feature extraction capability.

The Transformer consists primarily of positional encodings, encoders, and decoders. To make use of the order of the sequence, the Transformer adds "positional encodings" to the input embedding at the bottoms of the encoder and decoder stacks. The encoder and decoder consist primarily of a multi-headed attention mechanism layer and a fully-connected layer, between which the data is processed using residual connectivity and normalization. The multi-headed attention mechanism uses multiple attention mechanisms to perform multiple different linear transformations of the K, Q, V matrices and then stitches together the results obtained by each attention mechanism to enhance the performance of the model by focusing on different aspects of the information and integrating them [23]. The output of both the multi-headed attention mechanism and fully connected layers requires residual concatenation and normalization. Residual concatenation can solve the problem of gradient disappearance and degradation of the weight matrix [25]. Normalization allows the model to be unaffected by the input feature distribution and to converge faster [26].

$$Q = \text{Linear}\,(X) = XW^{0} \tag{21}$$

$$K = \text{Linear}\,(X) = XW^{K} \tag{22}$$

$$V = \text{Linear}\,(X) = XW^{V} \tag{23}$$

$$\text{SelfAttention}\,(Q,K,V) = \text{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{24}$$

Where X is the input matrix and W^Q, W^K, W^V are linear transformation weight matrices, the $softmax$ normalization function is used to calculate the attention weights.

$$\text{head}_i = \text{SelfAttention}\,(QW_i^{0}, KW_i^{K}, VW_i^{V}) \tag{25}$$

$$\text{MultiHead}(Q,K,V) = \text{Concat}\,(\text{head}_1, \cdots, \text{head}_h)W^{0} \tag{26}$$

Where W_i^0, W_i^K, W_i^V are the linear transformation weight matrices of the i-th head of Q, K, V, respectively, W^0 is the linear transformation weight matrix after stitching the multi-headed attention matrix.

2.2.2 Transformer-Based Prediction Model Improved by BiLSTM

In time series, the data at a point in time is associated with the data at the past time and with the data at the future time. The state of the bidirectional LSTM recurrent neural network at time t is related to the past moment $(t-1, t-2, \ldots, 0)$ to the future moment $(t+1, t+2, \ldots, T)$ [27]. Therefore, in order for the model to better extract the time series information in the stock series, a Bi-LSTM recurrent neural network is used to replace the position encoding in the original Transformer model. It extracts the time-series features of the input data based on the entire time series, providing the Transformer with the time series relationships between the input stock data.

The encoder maps an input sequence into a latent representation, and the decoder uses the representation and other inputs to generate a target sequence [24]. The multi-headed attention mechanism layer in the encoder enables the model to quickly obtain global time-series information and achieve parallel computation, effectively solving the long-range dependency problem while increasing computational efficiency and enabling the model to have stronger feature extraction capability. Since the objective of this paper is to predict the closing price of the SSE index, the model needs to learn the features that represent the stock trend rather than decoding them into another sequence, so the model in this paper is mainly built based on the encoder, and a fully connected layer replaces the decoder. The model structure is shown in Fig. 2 and Fig. 3.

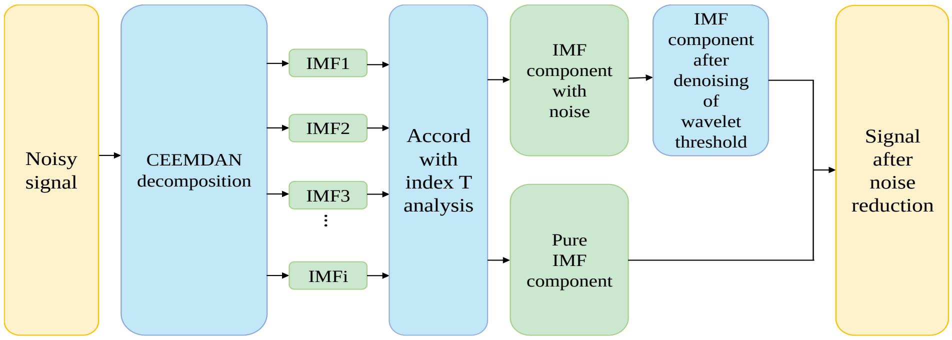

Figure. 1 Process of the CEEMDAN-WTD

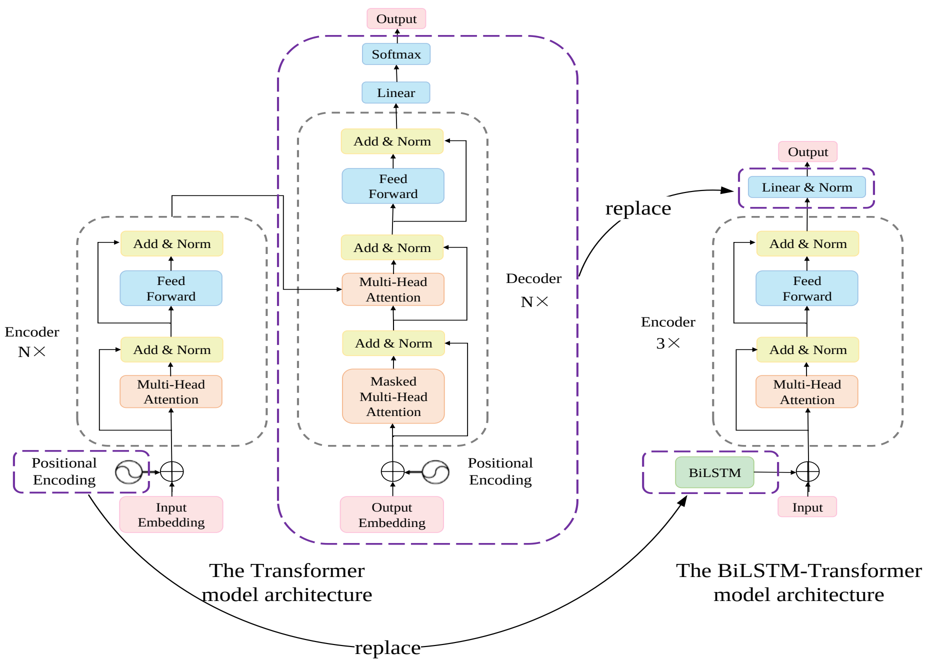

Figure. 2 Comparison of Transformer and BiLSTM-Transformer

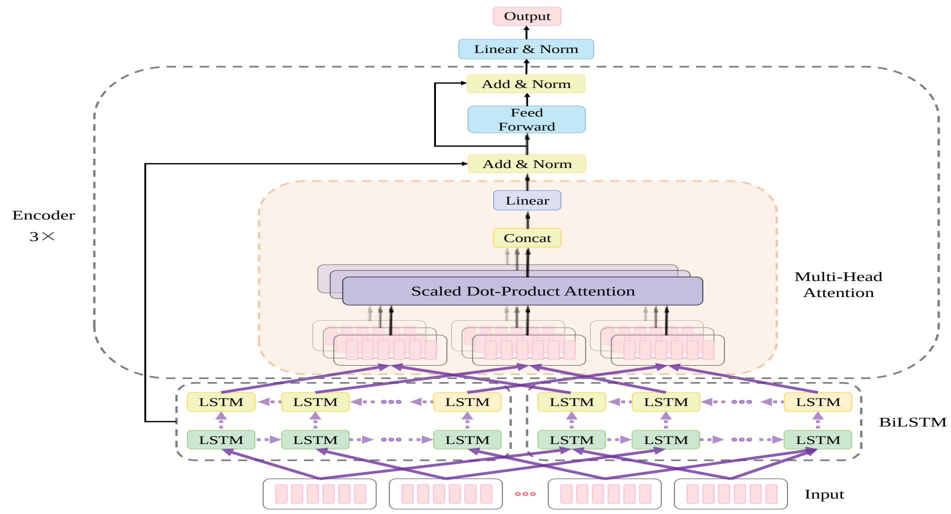

Figure. 3 Details of the BiLSTM-Transformer ModelHybrid Sliding Kernel Density Interval Estimation

In contrast, to point estimation, confidence interval estimation can quantify the variability of the forecast results due to uncertainty factors so that the actual observations fall within the forecast interval corresponding to a certain confidence level, providing decision-makers with more information to better understand the possible uncertainty and risk faced by the predicted quantity in future changes [28]. The kernel density estimation (KDE) method is an effective non-parametric method. However, kernel density estimation is mainly based on the estimation of the density of a sample with a moving window, and the accuracy of the kernel density estimation can be affected by choosing too large or too small a window bandwidth. In this paper, we use the mixed sliding kernel density estimation (MSKDE) method [28] to introduce the ideas of "mixing" and "sliding" into kernel density estimation. In MSKDE, "mixing" and "sliding" ideas are introduced into kernel density estimation. The probability density functions of the prediction errors estimated by different bandwidths are combined using appropriate weighting factors, and sliding means that the samples for kernel density estimation are sliding with the time series. The hybrid sliding kernel density estimation purposefully addresses the problem that ordinary kernel density bandwidths and fixed sample windows can impact the estimation accuracy, and therefore a more reasonable prediction interval can be generated using this method. The probability density function of the prediction error at time $t+1$ for the hybrid sliding kernel density estimation expression is as follows.

$$f_{\mathrm{MSKD},t+1}(\varepsilon)=\sum_{k=1}^{M}\beta_k\frac{1}{nh_k}\sum_{i=1}^{n}K\left(\frac{\varepsilon-\varepsilon_i}{h_k}\right) \tag{27}$$

Where N is the total number of samples; h_k is the k-th bandwidth; $\mathcal{E}_i$ is the given sample of prediction errors; $K(\cdot)$ is the kernel function; $f_k(\varepsilon)$ is the probability density function estimated from h_k; M is the total number of different bandwidths; and β_k is the weighting factor of $f_k(\varepsilon)$ and satisfies:

$$\sum_{k=1}^{M}\beta_k=1. \tag{28}$$

Note: N is the total number of samples replaced by: n is the number of samples with the nearest prediction error at time$t+1$.

Integration of $f_{\mathrm{MSKD},t+1}(\varepsilon)$ gives the prediction error probability distribution function $F_{\mathrm{MSKDE},t+1}(\xi)$.Given α (0<α<1), a confidence interval for the true value of the closing price with a confidence level of $1-\alpha$ is:

$$\left[\hat{L}_3,\hat{U}_3\right]=\left[P_{\mathrm{pred}}+\hat{F}_{\mathrm{MSKD}}(\alpha_1),P_{\mathrm{pred}}+\hat{F}_{\mathrm{MSKD}}(\alpha_2)\right]. \tag{29}$$

Where: $\hat{L}_3$ and $\hat{U}_3$ are the upper or lower limits of the confidence interval derived by the MSKDE method, respectively; $\hat{F}_{\mathrm{MSKD}}(\alpha_1)$ is the inverse function of $F_{\mathrm{MSKDE},t+1}(\xi)$; $\alpha_1=\dfrac{\alpha}{2},\alpha_2=1-\dfrac{\alpha}{2}$.

The confidence interval estimate at a certain confidence level can be obtained from the stock closing price point forecast value and the probability density curve of the closing price forecast error.

The specific calculation steps are as follows.

Step1: perform a closing price point forecast to obtain a point forecast value for the closing price.

Step2: find the points corresponding to α_1 and α_2 of the prediction error according to the probability density curve of the closing price prediction error.

Step3: calculate the confidence interval of the closing price prediction value according to equation (29).

2.2.3 Framework of Prediction

In this paper, a stock price interval prediction method based on CEEMD-WTD-BiLSTM-Transformer is proposed for the stock closing price prediction problem, and the specific prediction process is shown in Fig. 4.

(1) The stock closing price is decomposed into a series of high-frequency to low-frequency IMF components using the CEEMDAN method.

(2) Wavelet threshold noise reduction is applied to the noisy IMF components to obtain the denoised multiple IMF components

(3) Using the BiLSTM-Transformer model to predict the IMF components, all the predicted components are summed to obtain the point prediction value of the stock closing price.

(4) Use the mean absolute percentage error (MAPE), mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2) as evaluation indicators to compare the forecasting effectiveness of this model with the LSTM model, Transformer model, BiLSTM-Transformer model, and CEEMD-BiLSTM-Transformer model

(5) Using the hybrid sliding kernel density estimation model to calculate the prediction interval of the closing price point forecasts at the specified confidence level values.

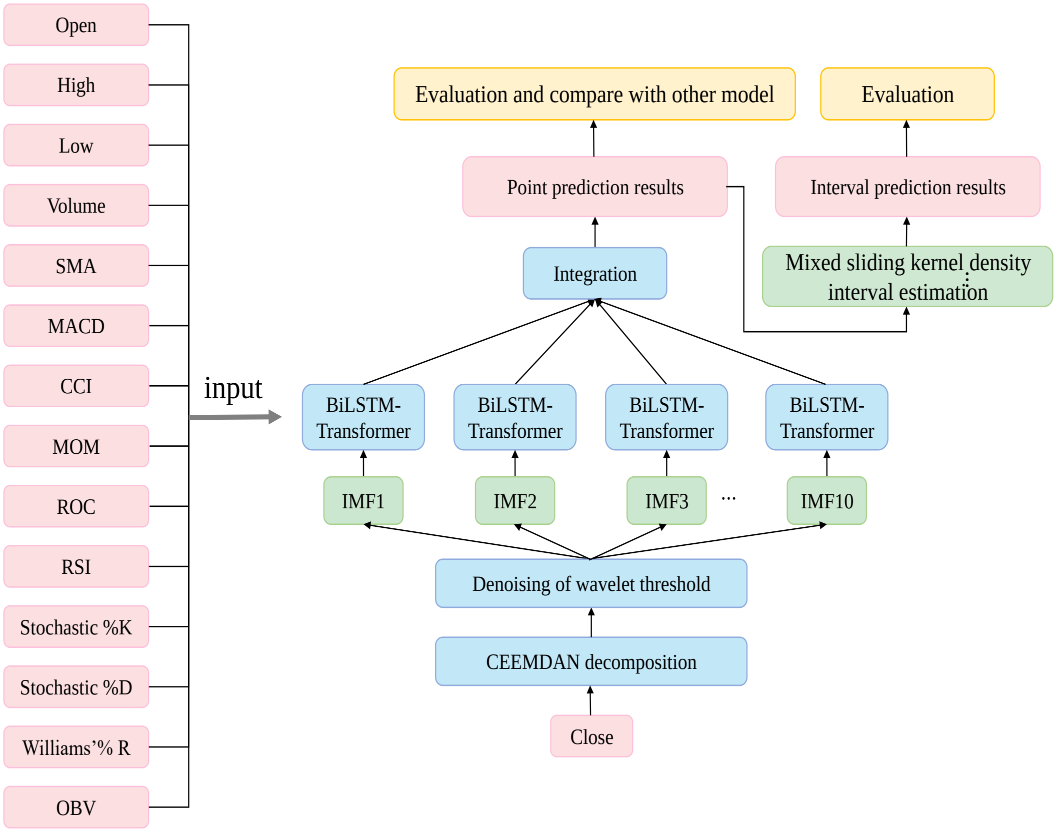

Figure. 4 Predictive FrameworkCase study

2.3 Sample selection and data sources

The experimental data are selected from the SSE index and a more representative stock in China. The data are selected for the interval from 1991/01/11 to 2021/01/13, and the data are obtained from the website of Yingwei Finance (https://cn.investing.com/). The reason for selecting the broad market index as the data is that the broad market is the centralized performance of the overall financial market volatility, which can exclude the chance of violent fluctuations of individual stocks [29]. The daily trading data of the SSE index contains the daily opening (Open), high (High), low (Low), closing (Close), and volume (Volume).

The closing price is often the price commonly accepted by market participants, and it contains much helpful information and has the function of predicting the price direction of the next trading day, so it is important to study the closing price [30]. Therefore, in this paper, the closing price of the SSE index on the next trading day is selected as the prediction object. The closing price of the SSE index from 1991 to 2021 is shown in Fig. 5. From Fig. 5, it can be seen that the SSE index first shows an oscillating trend, then a peak, and finally still shows an oscillating trend.

In addition to price and volume, previous studies [30-33] also widely used technical indicators derived from price and volume as input variables. In addition to the basic market indicators, some technical indicators were initially selected from previous studies, and their effects on model prediction results were observed through repeated experiments. Finally, 10 technical indicators were selected. The names of the selected technical indicators and their calculation formulae are shown in Table Ⅰ. Among them, c_t Indicates closing price at time t , EMA (Exponential Moving Average) denotes the exponential sliding average, its calculation formula is

$$EMA(k)_t = EMA(k)_{t-1} + 2/(k+1) \times (c_t - EMA(k)_{t-1}). \tag{30}$$

So

$$DIFF_t = EMA(12)_t - EMA(26)_t. \tag{31}$$

HH_{t-n} and LL_{t-n} represent the maximum value of the highest price and the minimum value of the lowest price in the past n time intervals, respectively; u_t represents price change upwards at time t; d_t represents price change downward at time t; TP (Typical Price) indicates the typical price, its calculation formula is

$$TP_t = (h_t + l_t + c_t)/3. \tag{32}$$

h_t represents the highest price at time t, l_t denotes the lowest price at time t, ATP is a simple sliding average of TP, MD (Mean Deviation) is the average deviation of TP, v_t represents the volume of transactions at time t.

Table. 1 Technical Indicators and Calculation Formula

Technical indicators	calculation formula
SMA	$(c_{t-9} + \cdots + c_{t-1} + c_t)/10$
MACD	$MACD(n)_{t-1} + \frac{2}{n+1}(DIFF_t - MACD(n)_{t-1})$
CCI	$(TP_t - ATP)/0.015MD$
MOM	$C_t - C_{t-n}$
ROC	$(c_t - c_{t-n})/c_{t-n} \times 100$
RSI	$100 - 100/\left(1 + \frac{\sum_{i=0}^{n-1} u_{t-i}}{n} / \frac{\sum_{i=0}^{n-1} d_{t-i}}{n}\right)$
Stochastic %K	$(c_t - LL_{t-n})/(HH_{t-n} - LL_{t-n}) \times 100$
Stochastic %D	$\left(\sum_{i=0}^{n-1} \%K_{t-i}\right)/n$
Williams'% R	$(HH_{t-n} - c_t)/(HH_{t-n} - LL_{t-n}) \times 100$
OBV	$OBV_t = \begin{cases} OBV_{t-1} + v_t, \text{if } c_t > c_{t-1} \\ OBV_{t-1} - v_t, \text{if } c_t < c_{t-1} \\ OBV_{t-1}, \text{otherwise} \end{cases}$

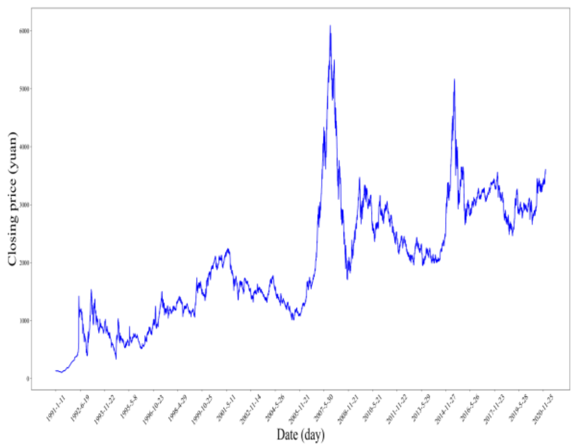

Figure. 5 SSE Closing Price

2.4 Applying the CEEMDAN-WTD Algorithm to Decompose the Data

The CEEMDAN decomposition of the acquired close data signal was performed and the 10 IMF components obtained are shown in Fig. 6, from which it can be observed that there is a very obvious high-frequency noise signal in the IMF component, with fluctuations in the original signal shown at the IMF4 component.

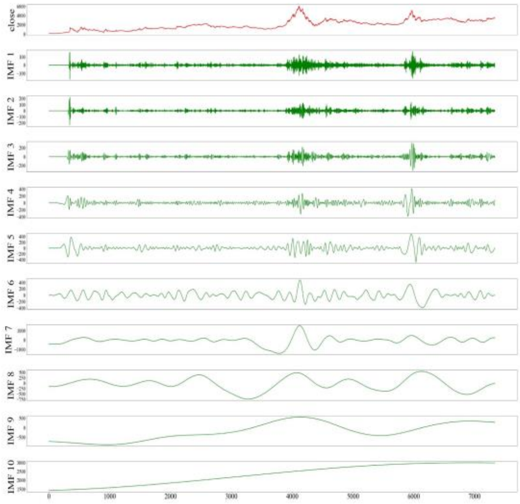

Figure. 6 Modal Components Obtained from CEEMDAN Decomposition

Using equation (19) (20), all IMF components' T and R values are calculated, where the first five R values are shown in Table Ⅱ.

From Table Ⅱ, it can be seen that when k=3, $1 \leq R_{k-1} \leq 3$, i.e., the first 3rd order IMF component is the noise component.

The IMF components containing noise are reconstructed with the remaining components after wavelet threshold noise reduction, i.e., the pre-processed noise reduced closing price data is obtained, as shown in Fig. 7.

Table. 2 R-value

K value	2	3	4	5	6
R value	3.894	1.501	2.418	1.752	3.478

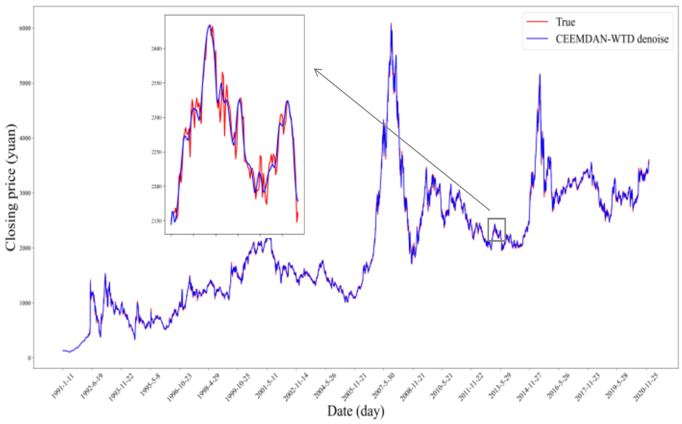

Figure. 7 Comparison of closing prices before and after noise reduction with CEEMDAN-WTD

2.5 Closing Price Prediction Results for the CEEMDAN-WTD-BiLSTM-Transformer Model

In order to verify the effectiveness of the CEEMD-WTD-BiLSTM-Transformer model on stock closing price prediction, this paper uses a total of 7335 points of SSE stock data from January 11, 1991, to January 13, 2021, as the original sample for prediction, dividing the training set and test set according to the ratio of 8:2 and using the maximum-minimum normalization method to dimensionless the data. In addition, a sliding window was constructed with a time step of 5, i.e., the stock data of the previous 5 days was used as input for the time-series rolling prediction. The pre-processed data is fed into the model, and the timing features are first extracted using a Bi-LSTM recurrent neural network, then the features are extracted using a 3-layer encoder, and finally, a fully-connected layer is added to transform the data dimension to 1-dimensional for the stock closing price prediction output. The model parameters are set in Table Ⅲ.

2.6 Comparative analysis of predictive models

To further evaluate the performance of the models on the stock closing price prediction problem, the LSTM, Transformer, BiLSTM-Transformer model, and CEEMDAN-BiLSTM-Transformer model were selected for cross-sectional comparison. The comparison of the prediction results of different models is shown in Fig. 8, from which it can be seen that the degree of coincidence between the predicted curves of several models and the actual curve is from low to high: Transformer, LSTM, BiLSTM-Transformer, CEEMDAN-BiLSTM-Transformer, CEEMDAN-WTD-BiLSTM-Transformer, CEEMDAN-WTD-BiLSTM-Transformer has the best prediction results in terms of overall and local similarity to the real curve.

Four commonly used error evaluation metrics are used to evaluate models: Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

Mean Absolute Percentage Error (MAPE) is a statistical measure of predictive accuracy ranging from 0 to positive infinity, with 0% indicating a perfect model and greater than 100% indicating a poor model.

$$MAPE = \frac{1}{n} \sum_{n}^{i=1} \left| \frac{\hat{y}_f - y_i}{y_i} \right| \times 100\% \tag{33}$$

Mean Squared Error (MSE) is the expected value of the squared difference between the estimated value of a parameter and the actual value of the parameter, which evaluates the degree of variability of the data. The smaller the MSE value, the better the accuracy of the prediction model.

$$MSE = \frac{1}{n} \sum_{n}^{i=1} (\hat{y}_i - y_i)^2 \tag{34}$$

Mean Absolute Error (MAE) is the average of the absolute errors, which better reflects the actual situation of the error in the predicted values.

Table. 3 Model Parameters

Parameter name	Parameter	Parameter name	Parameter
Number of BiLSTM neurons	64	Loss function	MSE
Number of multi heads h	5	Activation functions	sigmoid
Weighting Matrix Dimensions d_k	64	Optimizers	Adam
Number of encoders	3	Number of training rounds	1000
Encoder input size d_model	300	batch	512
dropout	0.2	Time step	5
Number of fully connected layer nodes	128	Training set: Test set	8:2

$$MAE = \frac{1}{n} \sum_{n}^{i=1} |\hat{y}_i - y_i| \tag{35}$$

The coefficient of determination, R^2, is the linear correlation between the predicted and observed values. r 2 values range from 0 to 1, with an r 2 value of 1 indicating a perfect correlation.

$$R^2 = 1 - \frac{\sum_{n}^{i=1} \left(\hat{y}_i - y_i\right)^2}{\sum_{n}^{i=1} \left(\bar{y}_i - y_i\right)^2} \tag{36}$$

The smaller the value of MAPE, RMSE, and MAE, the smaller the error and the better the prediction. The larger the value, the better the model predicts.

After the training was completed, the experimental results statistics of each model on the test set of SSE index stock data error evaluation index results are shown in Table Ⅳ.

Table. 4 Comparison Of Closing Price Forecast Results Of Alternative Models

	MAPE/%	MSE/yuan	MAE/yuan	R^2
LSTM	1.16	47.2146	36.3730	0.9565
Transformer	1.04	44.5092	34.6062	0.9678
BiLSTM-Transformer	0.96	40.4435	31.5811	0.9704
CEEMDAN-BiLSTM-Transform	0.87	36.7114	26.6604	0.9837
CEEMDAN-WTD-BiLSTM-Transformer	0.62	25.4659	19.1641	0.9924

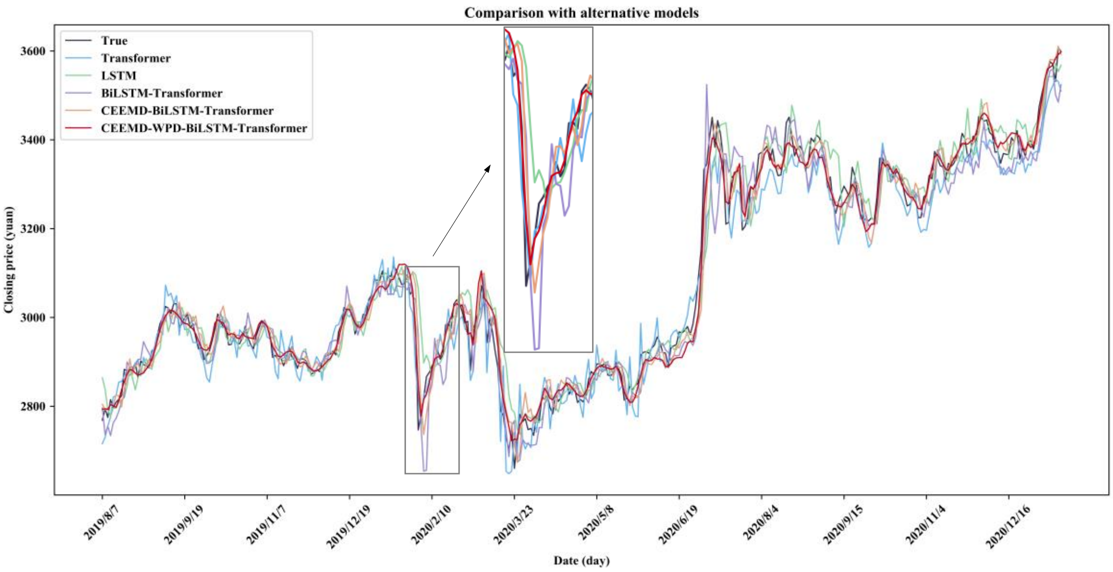

Figure. 8 Comparison of the Prediction Results of Different ModelFrom the prediction results of each model, it can be seen that

(1) The BiLSTM-Transformer neural network forecasting model gets some improvement in stock closing price forecasting accuracy over both LSTM and Transformer neural network forecasting models, which also proves that BiLSTM-Transformer, as an improved neural network model of LSTM and Transformer, can achieve good performance in the forecasting of stock time-series data.

(2) Compared with the BiLSTM-Transformer model, the CEEMDAN-BiLSTM-Transformer stock prediction model reduces each error by 9.38%, 9.23%, 15.58% and improves R^2 by 1.25%,

respectively, proving that the CEEMDAN decomposition can effectively improve the accuracy of the model, making the model superior in solving the stock closing price prediction problem with a certain superiority.

(3) The CEEMDAN-WTD-BiLSTM-Transformer model outperforms the other four models in the four error evaluation indicators of MAPE, RMSE, MAE, and R^2. Its MAPE, RMSE, and MAE values are the lowest, at 0.62%, 25.4659, 19.1641, respectively. It is reduced by 28.74%, 30.63%, 28.12%, respectively, compared with CEEMD-biLSTM-Transform, and R^2 increased by 0.88%, which indicates that Wavelet threshold noise reduction can further optimize the model so that the CEEMD-WTD-BiLSTM-Transformer model is optimal in terms of prediction accuracy, stability, and generalization, and can predict the future closing price of the SSE well.

2.7 Interval forecasts

A hybrid sliding kernel density estimation method was used to fit a probability density function using the forecast error data to further improve the forecast accuracy. Probability theory is applied to obtain stock closing price prediction intervals with specified confidence probabilities, and the reliability and accuracy of the interval predictions are evaluated using predict interval coverage probability (PICP) and predicted average interval width (PINAW), which are defined in equation (37) and equation (38) respectively. The results of the combined interval prediction at 80%, 90%, and 95% confidence levels are shown in Fig.9, and the corresponding prediction interval coverage and average bandwidth of the prediction interval are shown in Table Ⅴ, which shows that the confidence interval contains almost all actual values, indicating that the prediction interval has the advantages of both high reliability and accuracy.

The predicted interval coverage λ and the predicted interval average bandwidth w are defined as follows:

$$\lambda = \frac{1}{N} \sum_{N}^{i=1} k^{(\alpha)}, \tag{37}$$

$$w = \frac{1}{NR} \sum_{N}^{i=1} \left(\hat{U}(x_i) - \hat{L}(x_i) \right). \tag{38}$$

Where: k is a Boolean quantity, if the predicted target value t_i is included in the upper and lower bounds of the interval prediction, then $k=1$, otherwise $k=0$; R is the range of variation of the predicted target value; $\widehat{U}(x_i)$ and $\widehat{L}(x_i)$ are the upper and lower bounds of the interval derived from different estimation methods, respectively. N is the number of predicted samples.

Table. 5 Predicted Interval Coverage and Average Bandwidth

Confidence /%	PICP/%	PINAW/yuan
80%	94.84	0.0572
90%	96.36	0.0725
95%	99.70	0.0856

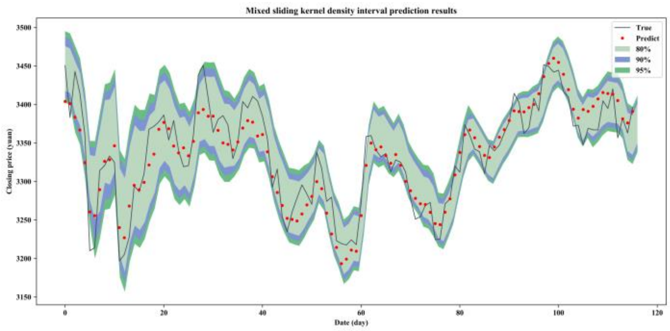

Figure. 9 Interval Forecast Results

At 80%, 90%, and 95% confidence probabilities, the interval coverage is 94.84%, 96.36%, and 99.70%, respectively, and the bandwidth indicators are 0.074 5 and 0.101 2. The PICP indicator meets the corresponding confidence level, but the PINAW indicator is also good, i.e., the prediction method has the advantages of both high reliability and good accuracy.

3. Conclusion

This paper proposes a stock price range prediction method based on the CEEMDAN-WTD-BiLSTM-Transformer model for stock closing price prediction. This paper selects SSE index stock data for experiments, and the experimental results show that the method can effectively predict stock prices.

(1) The model removes noise from the stock closing price by joint CEEMDAN-wavelet threshold de-noising, effectively removing the noise signal and improving the model prediction accuracy.

(2) A Bi-LSTM recurrent neural network is used to replace the position-coding in the original Transformer model. It uses the time-series relationship between the forward and backward time series; the self-attentiveness mechanism is retained, enabling the model to quickly obtain global time-series information and realize parallel computation, which increases computational efficiency while effectively solving the long-distance dependence problem and possessing more vital feature extraction capability.

(3) Compared with other models, the CEEMDAN-WTD-BiLSTM-Transformer model performs better than the other four models on the four error evaluation indicators of MAPE, RMSE, MAE, and R^2, and its values of MAPE, RMSE, and MAE are all the lowest. They are 0.62%, 25.4659, and 19.1641, respectively, which are 28.74%~46.55%, 30.63%~46.06%, 28.12%~47.31% lower than other models, respectively; the value of R^2 is the highest, which is 0.9894, which is very close to 1, and compared with other models, an improvement of 0.88%~3.62% is achieved.

(4) Combining the distribution characteristics of the Shanghai Composite Index closing price and its prediction error, the mixed sliding kernel density method is used to predict the Shanghai Composite Index closing price interval. A prediction interval of the Shanghai Composite Index closing price that considers reliability and accuracy is generated as an extension of the deterministic prediction, which provides decision-makers with richer forecast information.

(5) The CEEMDAN-WTD-BiLSTM-Transformer model is proved to be optimal in prediction accuracy, model stability, and generalization ability, and can adapt to some anomalous data sets and learn to learn the characteristics of irregular data changes, and fit the closing price of the SSE more reasonably. The CEEMDAN-WTD-BiLSTM-Transformer model has many parameters and a long-running time. Therefore, how to reduce the model parameters and improve the computational efficiency while ensuring the effectiveness of the model will be the main research direction in the future.

References

[1] Cheng Zhaozhi. Research on financial time series forecasting based on deep learning. University of Electronic Science and Technology, 2021. doi: 10.27005/d.cnki.gdzku. 2021.004845.

[2] Yang W, Han A, Wang Shouyang. Research on financial time series forecasting based on interval-type data. Journal of Systems Engineering,2016,31(06): 816-830.DOI: 10.13383/j.cnki.jse.2016.06.010.

[3] Yan H, Ouyang H. Financial time series prediction based on deep learning. Wireless Personal Communications, 2018, 102(2): 683-700.

[4] Hochreiter S, Schmidhuber J. Long short-term memory.Neural Computation, 1997, 9(8): 1735-1780.

[5] Moghar A, Hamiche M. Stock Market Prediction Using LSTM Recurrent Neural Network. Procedia Computer Science, 2020, 170:1168-1173.

[6] Peng Y, Liu YH, Zhang RF. Modeling and analysis of stock price prediction based on LSTM. Computer Engineering and Applications, 2019, 55(11):4.

[7] Zhang Y, Chu G, Shen D. The Role of Investor Attention in Predicting Stock Prices: The Long Short-term Memory Networks Perspective. Finance Research Letters, 2020, 38(2):101484.

[8] SuiJinCheng. Research on Stock Forecasting based on LSTM Neural Network. Qingdao: Qingdao University of Science and Technology, 2020.

[9] He Yiyue, Li Ping, Han Jinbo. Research on stock market index forecasting modeling based on CEEMDAN‑LSTM. Statistics and Information Forum, 2020, 35(6): 34 -45.

[10] Yan Hongju. Deep learning-based integrated forecasting of financial time series data [J]. Statistics and Information Forum, 2020, 35(4): 33 -41.

[11] Li Yafeng, Wang Hongbo, Li Chen, Wang Fuhao, Liu Meng, Luo Jingjing. An LSTM futures investment strategy incorporating attention mechanism. Computer System Applications,2021,30(08): 22-30.DOI: 10.15888/j.cnki.csa.008110.

[12] Li Q, Liu N., Yan B. S., Song U. An improved EMD algorithm based on Cubic Hermite. Machine Tools and Hydraulics,2021,49(21):186-190.

[13] TANG B, DONG S, SONG T. Method for eliminating mode mixing of empirical mode decomposition based on the revised blind source separation. Processing, 2012,92(1):248-258.

[14] Xue F, Sun XW, Dong Zixue, Yang HR, Wang HY. Research on EEMD-based data noise reduction algorithm. Mechanical Engineering and Automation,2021(05):9-11.

[15] ZHANG Jianwei,HOU Dove,TOU Zhenlei,ZHANG Yina,MA Ying. Noise reduction method for vibration signals of drainage structures based on CEEMDAN and SVD. Vibration and Shock, 2017,36(22): 138-143.DOI: 10.13465/j.cnki.jvs.2017.22.022.

[16] Zhang Jianwen,Liu Yang,Zhang Dapeng,Zhang Huanyu. A new method of de-noising based on CEEMDAN with wavelet adaptive thresholding Electrical Measurement and Instrumentation,2018,55(10):14-18+33.

[17] Bie Fengfeng,Zhang Shijia,Pei Junfeng,Xu Pengqing. Research on rolling bearing fault pattern recognition method based on wavelet thresholding and CEEMDAN joint de-noising. Mechanical Design and Manufacture,2020(02): 68-71.DOI: 10.19356/j.cnki.1001-3997.2020.02.017.

[18] Zhao K,Zheng S-X. Research on noise reduction of wind turbine data based on improved EMD and wavelet thresholding method. Journal of Shanghai Electric Power University,2020,36(02):136-140.

[19] Cai Cai, ZHAO Xiaotao, HU Xianeng, HUANG Xianghai, CHEN Huiming.Research on the combined CEEMDAN-wavelet thresholding method for ball mill cylinder vibration signal de-noising. Mechanical Science and Technology,2020,39(07): 1077-1085.DOI: 10.13433/j.cnki.1003-8728.20190241.

[20] Zhu Jianjun,Zhang Zhetao,Kuang Cuilin,Pan Jiabao. A reliable wavelet de-noising quality evaluation index. Journal of Wuhan University (Information Science Edition),2015,40(05): 688-694.DOI:10.13203/j. whugis20130417.

[21] Lu Tieding,Qian Wenlong,He Xiaoxing,Le Ying. An improved EMD method for determining the partitioned IMF components. Geodesy and Geodynamics,2020,40(07): 720-725.DOI: 10.14075/j.jgg.2020.07.012.

[22] Volker D. Rhythm and Speech Rate: A Variation Coefficient for ΔC. Language and language-processing, 2006(1):231-241.

[23] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017): 5998-6008.

[24] Jielin Qiu and Jiacheng Zhu and Michael Rosenberg and Emerson Liu and Ding Zhao. Optimal Transport based Data Augmentation for Heart Disease Diagnosis and Prediction. arXiv preprint arXiv:2202.00567, 2022.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.

[26] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. arXiv preprint arXiv:1607.06450, 2016.

[27] SuiJinCheng. Research on Stock Forecasting based on LSTM Neural Network Qingdao: Qingdao University of Science and Technology, 2020.

[28] Yang XY, Zhang YF, Ye TZ, et al. Probabilistic interval prediction of wind power combination based on plain Bayesian. High Voltage Technology, 2020(3).

[29] Tan, Meng Yu. Financial data prediction and classification based on support vector machine regression and learning. [Master's thesis]. Xi'an University of Electronic Science and Technology, 2014

[30] T. Huang & L. Yu, Application of SDAE-LSTM model in financial time series forecasting. Computer Engineering and Applications, 2019. 55(01): pp. 142-148.

[31] Kara Y, Boyacioglu M A, Baykan Ö K.Predicting direction of stock price index movement using artificial neural networks and support vector machines:The sample of the Istanbul Stock Exchange.Expert systems with Applications,2011,38(5):5311-5319.

[32] Patel J,Shah S,Thakkar P,et al.Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques.Expert Systems with Applications,2015,42(1):259-268.

[33] Armano G,Marchesi M,Murru A. A hybrid genetic-neural architecture for stock indexes forecasting. Information Sciences,2005,170(1):3-33