

# Depth feature extraction-based deep ensemble learning framework for high frequency futures price forecasting

Jujie Wang\*, Yu Chen, Shuzhou Zhu, Wenjie Xu

School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China



## ARTICLE INFO

### Article history:

Available online 26 April 2022

### Keywords:

High-frequency futures  
Price prediction  
Feature extraction  
Attention mechanism  
Ensemble learning

## ABSTRACT

Whether the change trend of futures price can be accurately analyzed and predicted is the key to the success or failure of futures trading. This paper constructs a new deep ensemble learning framework combining signal decomposition and exogenous variable feature mining for high-frequency futures price prediction, which consists of depth feature extraction (DFE), long short-term memory optimized by attention mechanism (ALSTM) and Light gradient boosting machine (LightGBM). In the depth feature extraction stage, based on multi-scale entropy (MSE) and Savitzky-Golay filter (SG filter), an improved denoising variational mode decomposition (VMD) is proposed to extract the fluctuation characteristics of futures price signal and eliminate the interference of complex components. To avoid the collinearity redundancy of high-dimensional exogenous variables, an enhanced dimensionality reduction method combining Spearman correlation analysis and stacked autoencoder (SAE) is designed to ensure the simplicity and correlation of input factors. In the prediction phase, ALSTM is adopted as a base predictor for constructing point prediction model by the DFE results, which can focus on learning more important data features. Finally, LightGBM, which has excellent effect in the field of ensemble learning, is used to integrate the base prediction results to obtain the final results. The actual closing price data of three representative futures varieties in China's futures market are selected to verify the accuracy of the proposed framework. Compared with other benchmark models, this developed framework has better futures closing price prediction performance.

© 2022 Elsevier Inc. All rights reserved.

## 1. Introduction

After the outbreak of COVID-19 in 2019, it swept the globe in a fleeting time and became a typical "black swan" incident. The market environment has changed greatly, which has had a serious impact on the economy and the financial industry [1]. The uncertainty of the time, mode and form of economic recovery threatens consumers and producers [2]. As a financial derivative, futures play an indispensable role in price discovery, hedging and risk control. During the epidemic, more investors try to make use of futures for risk hedging. Therefore, the research on futures price is of particular importance. Due to the special "T+0" trading mechanism, futures can be bought and sold unlimited times in the same day. Therefore, high-frequency trading is immensely popular in the futures market and is an important branch of quantitative investment. High frequency data contains a lot of information, which can help traders and investors make decisions quickly [3]. However,

how to predict high-frequency futures with nonlinear and random fluctuations through turbulent market noise is still an urgent problem for researchers to explore.

Reviewing the related studies, the forecasting models for financial time series can be broadly classified into three types: statistical models, artificial intelligence models and hybrid models. Researchers often use statistical models to analyze or predict the volatility of futures prices. Evans [4] found a strong correlation between U.S. economic news announcements and intraday price volatility of futures. Stoll et al. [5] found that the transmission of trading signals is time-sensitive, and the accuracy of predicting price movements using 5-minute high-frequency futures data is higher than that of 10-minute high-frequency data. Brooks [6] used ARCH models to study futures price volatility and found that volatility varies in unevenness. Bunnag [7] used a GARCH model to predict oil futures prices and calculate the optimal oil portfolio weights. Huang et al. [8] estimated and predicted the price volatility of four agricultural product futures using GARCH and EGARCH, proving that GARCH can achieve effective estimation. Although traditional statistical models such as GARCH can achieve effective results in predicting financial time series, it needs to be based

\* Corresponding author.

E-mail address: [jujiewang@126.com](mailto:jujiewang@126.com) (J. Wang).

on complete statistical assumptions [9]. In the rapidly changing financial market, financial time series often have complex characteristics, such as nonlinearity and nonstationarity. The traditional statistical model cannot effectively deal with the nonlinear characteristics, and the effectiveness and accuracy of its prediction results are difficult to be further improved.

In view of the inherent defects of traditional statistical models in predicting nonlinear and non-stationary time series data, researchers turn their attention to artificial intelligence (AI) models with adaptive characteristics and learning ability [10]. Different from statistical models, artificial intelligence models do not need to meet statistical assumptions and can capture the internal laws and nonlinear information of data. Artificial intelligence models such as artificial neural network (ANN) [11] and recurrent neural network (RNN) [12] have shown excellent performance in financial data prediction. For example, Kulkarni and Haidar [13] used ANN to build a short-term prediction model of crude oil price to accurately predict the future market trend. Hajabotorabi et al. [14] developed a novel wavelet-based deep recurrent neural network (RNN) for crude oil futures price prediction. Their results prove that the developed model has a crucial betterment in the error measurement of crude oil futures time series pricing. As a common variant of RNN, long short-term memory (LSTM) neural network overcomes the gradient disappearance and gradient explosion problems of RNN in processing long sequence data [15]. Relevant studies have proved that LSTM has been widely used in the field of financial time series, such as carbon finance [16], stocks [17], and has penetrated the field of futures forecasting. For example, Zhang et al. [18] constructed a novel LSTM model to predict the price of energy futures at different time intervals. The experimental results show that the volatility of energy futures price decreases with the increase of time interval. Although the AI model shows superior performance in predicting financial time series, considering the uncertainty and complex volatility of the real financial market, a single AI model cannot adapt to the data of distinctive characteristics and modes, and it is difficult to meet the needs of researchers for prediction accuracy [19]. To further improve the effectiveness and accuracy of prediction, researchers began to develop hybrid models to obtain better prediction performance.

Relevant research shows that, the existing hybrid models are mainly divided into two types: one is to combine different prediction models, and the other is to combine data processing technology and prediction models. Among them, the hybrid model based on decomposition integration technology has proved to be an effective tool for financial time series prediction [20,21]. Decomposition integration technology decomposes complex time series into several subsequences with simpler structure through signal decomposition algorithm. Then, proper prediction models are constructed according to different data characteristics. Finally, the prediction results of all subsequences are integrated to obtain the final prediction results. Zhu et al. found that the decomposition of carbon price can find the intrinsic features of carbon price under different modes, effectively tap the fluctuation trend of carbon price and improve the prediction performance [22]. Liu and Long [23] use empirical wavelet transform (EWT) to decompose the stock market closing price series into several subsequences and use LSTM to predict them, respectively. The experimental results show that the hybrid framework can be used for the analysis and research of financial data. Yu et al. [24] used empirical mode decomposition (EMD) to decompose the crude oil price into several subsequences and realized effective prediction using adaptive linear neural network. Although these methods can achieve effective prediction results, they all have unavoidable shortcomings. The decomposition performance of wavelet transform (WT) depends on the artificially selected wavelet basis function and the choice

of decomposition level. Although EMD can realize adaptive decomposition based on data-driven, it also has inherent problems such as endpoint effect. To overcome these shortcomings, variational modal decomposition (VMD) was proposed in 2014. VMD is an adaptive technology, which has faster convergence speed, stronger mathematical theory support, and can suppress mode aliasing effect [25]. Liu et al. [26] proposed a new decomposition integration framework with VMD and ANN for product futures price forecasting, which showed superior performance in whether one-step or multi-step prediction. Some of the previous studies usually forecast each subseries after decomposition, and there is often a similar complexity between different subseries. In addition, there is a large amount of noise in financial time series, and the treatment of noise is often neglected. This not only increases the computational complexity of the model, but also may lead to the accumulation of prediction errors. Therefore, the effective processing of decomposed subsequences stays to be further studied.

To further improve the prediction accuracy of the mixed model, researchers began to introduce exogenous variables into the prediction model. Based on the data decomposition of carbon price, Hao and Tian [27] introduce a variety of exogenous variables of carbon price and use Max relevance min redundancy algorithm to select several factors that have a great impact on carbon price. Using the historical data of carbon price and exogenous variables as input variables, the extreme learning machine (ELM) has achieved good prediction performance and effectively predicted the fluctuation trend of carbon price. When considering external variable inputs, data redundancy needs to be considered, which may increase the complexity of each prediction model. Therefore, it is not enough to perform feature selection on exogenous variables. It is necessary to reduce redundant information and to mine the intrinsic features of exogenous variables. There are many feature extractors used to capture high-level features of data, such as principal component analysis (PCA) [28], least absolute shrinkage and selection operator (LASSO) [29], etc. However, the most of these methods are shallow architecture with a single hidden layer. In contrast, deep networks with multiple hidden layers usually have more powerful representation ability for complex data. Therefore, deep learning techniques like stacked autoencoder (SAE) [30] and deep belief network (DBN) [31] are gradually come into public view in data processing and have been adopted in many areas. Though these methods can efficiently capture the high-level features of the raw inputs, they cannot guarantee the removal of factors that are not related to the target data. To solve this problem, it is necessary to select features before feature extraction. Granger causality test [32] and correlation test [33] have been proved to be helpful in previous studies.

From some of the previous studies, it can be found that although the hybrid model based on decomposition integration strategy has excellent prediction performance and robustness, there are still some drawbacks. First, some researchers often ignore the complexity of the similarity between different decomposition subsequences, and modeling each subsequence separately undoubtedly increases the complexity of model computation. Second, the noise present in the financial time series needs to be dealt with, which will affect the validity and accuracy of the prediction results. Third, the selection of exogenous variables alone is not sufficient; redundant information in the data will affect the accuracy of the prediction and mining the intrinsic feathers of the exogenous variables is necessary. Fourth, after obtaining the prediction results of each decomposed subseries, some of the current studies are mainly limited to linear integration, i.e., accumulating the predicted values to obtain the final prediction results. However, the simple linear integration method may lead to a decrease in prediction accuracy due to problems such as er-

ror accumulation [34]. And the nonlinear integration method can further explore the potential features between subseries to further improve the prediction accuracy. At present, there are two common nonlinear ensemble learning algorithms: bootstrap aggregating (bagging) represented by random forest and boosting represented by adaptive boosting (Adaboost) [35] and gradient boosting decision tree (GBDT) [36]. Extreme gradient boosting (Xgboost) is a special GBDT, which has the characteristics of high efficiency and flexibility. However, its high space complexity and large memory consumption make it too expensive to process large-scale data. Light gradient boosting machine (LightGBM) [37] which is based on histogram, can accelerate the training speed of GBDT model without damaging the accuracy and avoid the defect of Xgboost [38,39].

Motivated by the above discussions, a novel information fusion ensemble framework for high-frequency futures price forecasting is established. First, the original closing data is decomposed into several subseries using the VMD algorithm. Second, the Savitzky-Golay (SG) filter is used to reduce the noise in different subsequences. Third, multi-scale entropy (MSE) is used to measure the complexity of different subsequences, and subsequences with similar complexity are reconstructed into new subsequences. Meanwhile, Spearman correlation analysis combined with SAE is used to capture the intrinsic features of exogenous variables. Fourth, the attention mechanism-optimized LSTM model is developed to predict the reconstructed subsequences. Finally, LightGBM is introduced to nonlinearly integrate the prediction results of the ALSTM model to obtain the final prediction results. The main contributions of this paper are summarized below.

- (1) Different intrinsic features hidden in the closing price and exogenous variables are considered. In this paper, different deep feature extraction models are constructed.
- (2) VMD and SG filter effectively decompose the futures closing price series and reduce the noise in them. MSE reconstructs the subsequences with similar complexity and improves the computational efficiency of the prediction model.
- (3) The feature extraction model based on Spearman correlation analysis and SAE effectively captures the intrinsic features of exogenous variables to further improve the stability and accuracy of prediction.
- (4) The LSTM model is improved by adding the attention mechanism, which can give more attention to notable features, realize the effective extraction of short-term patterns and avoid the loss of long-term information.
- (5) LightGBM was introduced into the financial field, its application scope was expanded, and the effectiveness and universality of it are confirmed.
- (6) An information fusion ensemble framework is constructed for high-frequency futures price forecasting, which has superior prediction performance and robustness. In practical application, it can supply decision support for high-frequency trading.

The rest of this paper is structured as follows: In Section 2, the methodology and principles involved are presented. The process and diagram of the proposed framework are displayed in Section 3. The empirical results and discussion are introduced in Section 4. At last, conclusions and future directions are shown in Section 5.

## 2. Methodology

The specific principles of the methods and technical models involved in this study will be presented in this section.

### 2.1. Variational mode decomposition

VMD is a novel adaptive non-recursive decomposition technology proposed by K. Ragomietretskiy [40]. Wang et al. [41] used SVD and VMD to denoise the nonstationary signals of roller bearings, and the results proved that VMD can effectively decompose the nonstationary data and extract the effective information, and the method is not a black-box model and is interpretable. In recent years, VMD has been widely used in non-stationary data processing and prediction, such as fault diagnosis [42], wind power prediction [43] and PM<sub>2.5</sub> concentration prediction [44].

The input signal  $x(t)$  will be decomposed into a plurality of variational modes which are commonly known as the Intrinsic Mode Function (IMF), each of them with center frequency and limited bandwidth. The variation problem can be displayed as follows:

$$\min_{\{u_k\}, \{w_k\}} \left\{ \sum_{k=1}^k \left\| \partial_t[(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-j w_k t} \right\|_2^2 \right\} \text{ s.t. } \sum_{k=1}^k u_k = x(t) \quad (1)$$

where  $k$  is the number of modes;  $\{u_k\}, \{w_k\}$  ( $k \in [1, K]$ ) denote the sets of all modes and corresponding center frequencies, respectively;  $\delta(t)$  is Dirac distribution;  $*$  represents the convolution operator.

Then the quadratic penalty term  $\gamma$  and Lagrange multiplier  $\varsigma$  are taken into account to relieve constraint. The formula is constructed as Eq. (2).

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \varsigma) = & \gamma \sum_{k=1}^k \left\| \partial_t[(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-j w_k t} \right\|_2^2 \\ & + \left\| x(t) - \sum_{k=1}^k [u_k(t)] \right\|_2^2 \\ & + [\varsigma(t), x(t) - \sum_{k=1}^k u_k(t)] \end{aligned} \quad (2)$$

During the next process, the alternate direction method of multipliers (ADMM) is referenced as a common method to update  $u_k$ ,  $w_k$  and  $\varsigma$ , and thus find out the saddle point of the Lagrangian function in Eq. (2). The specific iterative steps are as follows:

- (1) Initialize the values of  $\{u_k^1\}, \{w_k^1\}, \{\varsigma^1\}$  to 0, at the same time,  $k$  is set to the number of times to be resolved.
- (2) Update  $\ddot{u}_k^{n+1}$  and  $w_k^{n+1}$  by following Eq. (3) and (4).

$$\ddot{u}_k^{n+1}(w) = \frac{\ddot{x}(w) - \sum_{i \neq k} \ddot{u}_i(w) + \frac{\ddot{\varsigma}(w)}{2}}{1 + 2\gamma(w - w_k)^2} \quad (3)$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\ddot{u}_k(w)|^2 dw}{\int_0^\infty |\ddot{u}_k(w)|^2 dw} \quad (4)$$

Where  $\ddot{u}_k^{n+1}, \ddot{x}(w), \ddot{u}_i(w), \ddot{\varsigma}(w)$  indicate  $u_k^{n+1}(w), x(w), u_i(w), \varsigma(w)$  after the Fourier transform, respectively;  $n$  denotes the number of iterations.

- (3) Update  $\mathcal{L}(w)$  in the non-negative frequency interval, the calculation process is as follows:

$$\ddot{\varsigma}^{n+1}(w) = \ddot{\varsigma}^n(w) + \xi \ddot{x}(w) - \sum_{k=1}^k \ddot{u}_k^{n+1}(w) \quad (5)$$

where  $\xi$  stands for the iterative factor.

- (4) Set a condition to stop iteration. If the condition in Eq. (6) is reached, then stop the iteration process, otherwise it returns to step (b) to calculate again. Eventually, the  $k$  modes are obtained.

$$\frac{\sum_{k=1}^k \|\ddot{u}_k^{n+1} - \ddot{u}_k^n\|_2^2}{\|\ddot{u}_k^n\|_2^2} < \rho \quad (6)$$

where  $\rho$  is the preset evaluation accuracy.

The number of modal components  $k$  needs to be determined before VMD decomposition. If the value of  $k$  is too large, problems such as mode mixing and noise generation will occur, and if the value of  $k$  is too small, it will be under-decomposed and difficult to extract enough effective information. In this paper, the number of  $k$  is selected by observing and analyzing the center frequencies of different modal components.

## 2.2. Multi-scale entropy

While the traditional sample entropy only responds to the characteristics of a signal at a certain scale, the multiscale entropy algorithm can evaluate the complexity of a signal at different time scales. The more complex the signal is, the more noise it contains, and the larger the multiscale entropy value [45].

For complex futures closing price data, after VMD decomposition, each IMF has a different level of complexity due to the presence of noise. Therefore, multi-scale entropy is used to quantify the nonlinear and non-stationary features in the signal [46]. For a given data series  $\chi = \{\chi_1, \chi_2, \dots, \chi_l\}$ , its MSE is calculated by the following steps:

(1) Handle the raw data series with coarse-grained algorithm and a new series can be obtained.

$$\left\{ \begin{array}{l} g_j^{(\psi)} = \frac{1}{\psi} \sum_{i=(j-1)\psi+1}^{j\psi} \chi_i \\ 1 \leq j \leq \frac{N}{\psi} \end{array} \right. \quad (7)$$

where  $\psi$  is the scale factor;  $j \in [1, N]$ ;  $N = \frac{1}{\psi}$  represents each length of the time series after coarse-grained algorithm.

(2) Construct a m-dimentional vector  $G^{(\psi)}(i)$  based on Eq. (7).

$$G^{(\psi)}(i) = [g^{(\psi)}(i), g^{(\psi)}(i+1), \dots, g^{(\psi)}(i+m-1)] \quad (8)$$

among which  $i = 1, 2, \dots, N-m+1$ .

(3) Define  $d(G^{(\psi)}(i), G^{(\psi)}(k))$  is the distance between  $G^{(\psi)}(i)$  and  $G^{(\psi)}(j)$ .

$$\left\{ \begin{array}{l} d(G^{(\psi)}(i), G^{(\psi)}(j)) = \max \left| g_{i+k}^{(\psi)} - g_{j+k}^{(\psi)} \right|, k = 0, 1, \dots, m-1 \\ 1 \leq i, j \leq \frac{N}{\eta-m}, i \neq j \end{array} \right. \quad (9)$$

(4) Set the threshold  $u$ , if  $d(G^{(\psi)}(i), G^{(\psi)}(k)) < u$ , define  $[G^{(\psi)}(i), G^{(\psi)}(k)]$  as the matching vector pairs in  $m$  dimension. Use  $T$  to represent the total number of matching vector pairs in  $m$  dimension. Mark the ratio of  $T$  to  $N-m+1$  as  $C_i^{\psi,m}(u)$ . Therefore, the average value of  $C_i^{\psi,m}(u)$  can be defined as follows:

$$C^{\psi,m}(u) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \frac{T}{N-m+1} \quad (10)$$

(5) Increase the dimension to  $m+1$  and repeat the above steps to obtain  $C^{\psi,m+1}(u)$ . At this moment, the MSE can be defined as:

$$\text{MSE}(\chi, \psi, m, u) = -\ln \frac{C^{\psi,m+1}(u)}{C^{\psi,m}(u)} \quad (11)$$

According to the above calculation process, it can be seen that the calculation result of MSE is closely related to the pattern dimension  $m$ , the threshold  $u$ , the data length  $N$  after coarse granulation, and the scale factor  $\psi$ . The size of the mode dimension  $m$

is related to the length of the data, and the threshold  $u$  is set too small to make the statistical information missing, and too large to make the results sensitive to noise. Referring to related literature and experimental data,  $m$  is set to 2 and  $u$  is set to 0.15. For the scale factor  $\psi$ , it can be chosen in the range of 1 to 20 [47].

## 2.3. Savitzky-Golay filter

SG Filter is a type of digital filter methods which utilizes convolution operation to smooth the original time-domain signals [48]. The algorithm can maintain the shape and length of the signals while removing the micro noise and exhibits excellent performance in the field of removing high frequency noise. The futures high frequency closing price data samples tend to be noisy, which will affect the validity and accuracy of the prediction results [49].

The main idea of SG Filter is to use local least-squares polynomial approximation to fit the raw signals, thereby achieving data smoothing. The polynomial is shown as Eq. (12), with the number of input samples set to  $2Z+1$ . At the same time, the center point is considered at  $c_p = 0$ .

$$p(c_p) = \sum_{k=0}^S a_k c_p^k \quad (12)$$

where  $S$  is the power of the polynomial, and  $S \leq 2Z+1$ .

Then, the description of mean-squared approximation error  $\varepsilon$  and the solution of the polynomial are displayed in Eq. (13). The square approximation error is utilized to smooth the local signals.

$$\varepsilon_S = \sum_{c_p=-Z}^Z (p(c_p) - \chi(c_p))^2 = \sum_{c_p=-Z}^Z \left( \sum_{k=0}^S a_k c_p^k - \chi(c_p) \right)^2 \quad (13)$$

## 2.4. Spearman correlation analysis

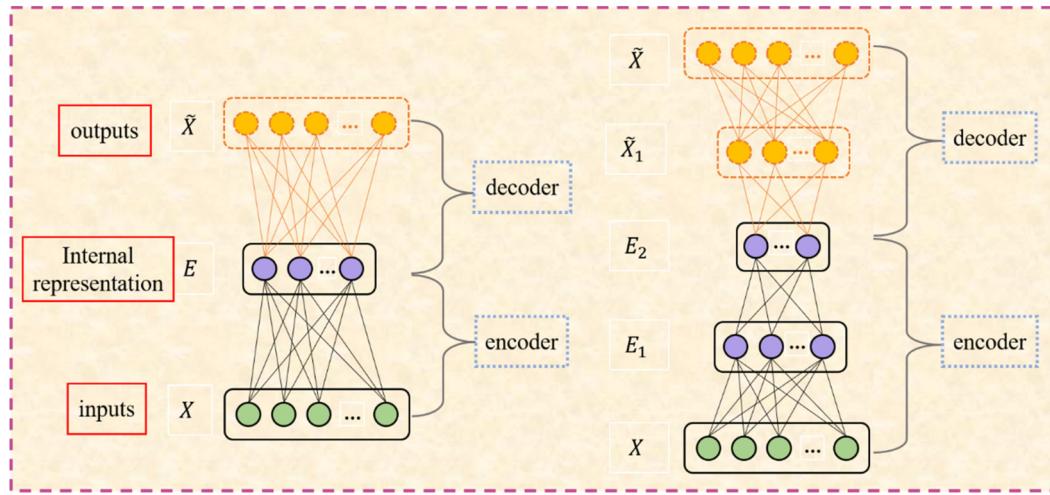
The Spearman correlation analysis is a non-parameter indicator to measure the dependent relationships between two variables. It is not affected by outliers and is suitable for nonlinear variables with complex relationships. It does not assume that the data points obey normal distribution or satisfy any definite relationship. Suppose that the paired labeled data are  $\{X_H, Y_H\} = \{(x_1, y_1), \dots, (x_h, y_h)\}$ , where  $x_j = [x_{j(1)}, x_{j(2)}, \dots, x_{j(d)}]$ ,  $j \in [1, h]$ . Assume the collection set of the  $d^{\text{th}}$  input variable of the labeled data is  $X_{H(d)}$ . Then,  $X_{H(d)} = \{x_{1(d)}, x_{2(d)}, \dots, x_{h(d)}\}$ . The Spearman correlation coefficient  $\rho$  of the  $d^{\text{th}}$  input data is displayed as Eq. (14).

$$\left\{ \begin{array}{l} \rho_d(X_{H(d)}, Y_H) = 1 - \frac{6 \sum c_{j(d)}^2}{h(h^2-1)} \\ c_{j(d)} = r(x_{j(d)}) - r(y_j) \end{array} \right. \quad (14)$$

among them,  $r(x_{j(d)})$  and  $r(y_j)$  are the ranks of each variable in the input data set  $X_{H(d)}$  and  $Y_H$ , respectively.  $c_{j(d)}$  represents the difference between the two ranks of each variable.

## 2.5. Stacked auto-encoder

Auto-encoder (AE) is an unsupervised automatic learning method proposed in the early stage for reducing the dimension of high-dimensional data. In 2006, Hinton and Salakhutdinov [50] proposed the concept of SAE by stacking multiple AEs, which can effectively learn the characteristics of unlabeled data. It is trained



**Fig. 1.** The principle of AE and SAE training.

layer by layer by employing the greedy learning algorithm to extract advanced features from input data. Therefore, it can capture a more abstract and complex representation of the original input data. The structures of basic AE and SAE with double hidden layers are exemplified in Fig. 1.

The AE takes input variables  $X$  and maps them to a potential presentation space  $E$  using a neural network with single layer. The encoder transforms  $X$  into  $E$  by Eq. (15).

$$E = \sigma(W_E X) \quad (15)$$

where  $\sigma$  indicates a sigmoid function;  $W_E$  is the weight matrix between input and output neurons.

Then, the hidden representation  $E$  is mapped to a reconstructed vector  $\tilde{X}$  in the decoding layer by decoding function  $D(E) = \sigma(W'_E E)$ . The output  $\tilde{X}$  is approximately equal to the input  $X$ .

SAE consists of an input layer, multiple hidden layers and an output layer. From the first hidden layer, AE is employed for unsupervised training. Each hidden layer will become the input layer of the next hidden one. The process is repeated until there are no more hidden layers. Finally, the latent spatial representation of the last hidden layer is the dimensionality reduction result of SAE.

## 2.6. Long short-term memory neural network

Recurrent Neural Network (RNN) considers the time series information contained in the data and has a time dimension. As a variant of RNN, LSTM can well solve the gradient explosion problem and make up for the RNN's difficulty in processing long-term dependence on information.

The LSTM neural network has a control gate mechanism, which consists memory cell state  $C_t$ , input gates  $I_t$ , output gates  $O_t$ , and forget gates  $F_t$ . When the data is transferred into the LSTM,  $F_t$  determines the degree of passage of the information at the previous moment based on the output  $h_{t-1}$  of the previous moment and the current input  $x_t$ . Afterwards, the sigmoid function  $\sigma$  is adopted in  $O_t$  to determine which values are to be updated, and then new candidate values  $\tilde{C}_t$  are generated by the tanh layer. The final output  $h_t$  of the LSTM is determined by  $O_t$  and  $C_t$ . The main equations of LSTM are as follows:

$$F_t = \sigma(W_\kappa \cdot [h_{t-1}, x_t] + b_\kappa) \quad (16)$$

$$I_t = \sigma(W_\varphi \cdot [h_{t-1}, x_t] + b_\varphi) \quad (17)$$

$$\tilde{C}_t = F_t \cdot C_{t-1} + I_t \cdot \tanh(W_\tau \cdot [h_{t-1}, x_t] + b_\tau) \quad (18)$$

$$O_t = \sigma(W_\vartheta \cdot [h_{t-1}, x_t] + b_\vartheta) \quad (19)$$

$$h_t = O_t \cdot \tanh(C_t) \quad (20)$$

among them,  $W_k, b_k (k = \kappa, \varphi, \tau, \vartheta)$  respectively denotes the weight matrices and bias vectors of each gate and memory cell.

To explain the structure of the LSTM more clearly, the specific diagram is shown in Fig. 2.

## 2.7. Attention mechanism

Attention mechanism is a resource allocation mechanism that simulates human visual mechanism. Its purpose is to allocate more resources to areas with high correlation under limited computing power, so as to achieve more detailed information that requires to be focus more importance on and restrain other useless information [51]. The model with attention mechanism can pay more attention to the influence weight of different information on the prediction results and carry out more intelligent and selective learning for different inputs. The attention mechanism architecture is displayed in Fig. 3.

Supposing that the m-dimensional input vectors are  $\{h_i\}$  ( $i = 1, 2, \dots, k$ ) and the environment vector  $v_i$  can be obtained based on  $h_i \cdot v_i$  is the weighted average of the previous states, the specific calculation is shown in Eq. (21).

$$v_i = \sum_{i=1}^k a_i h_i \quad (21)$$

where  $a_i$  represents the attention weight added by the state. In order to measure the degree of influence of input information on the output,  $s_i$  which indicates the degree of correlation between  $v_i$  and  $h_i$  is introduced.

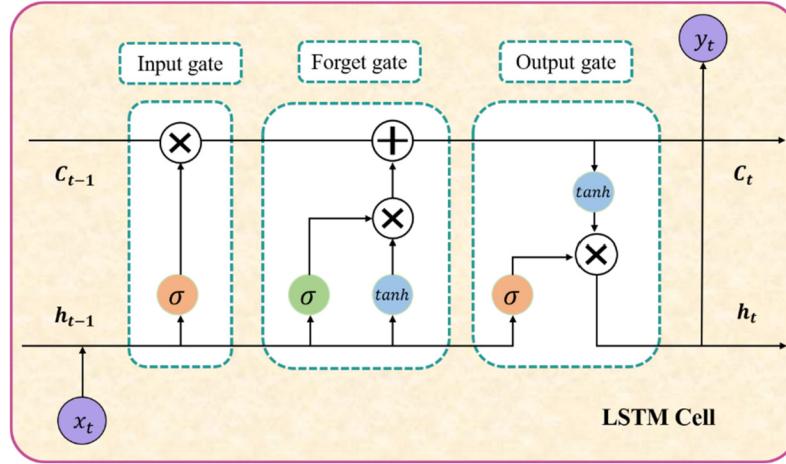
$$s_i = \tanh(W^T h_i + b_i) \quad (22)$$

Finally, the softmax function is utilized to normalize  $s_i$  to obtain  $a_i$ .

$$a_i = \text{softmax}(s_i) = \frac{e^{s_i}}{\sum j e^{s_i}} \quad (23)$$

## 2.8. LightGBM algorithm

LightGBM is a new GBDT (Gradient Boosting Decision Tree) algorithm based on histogram. Compared with the traditional GBDT, it utilizes Gradient-based One-Side Sampling (GOSS) and Exclusive



**Fig. 2.** The concrete structure of the LSTM.

Feature Bundling (EFB) technology. It transforms traversal samples into traversal histograms, which greatly reduces the time complexity and make support efficient parallel training. The optimized feature parallel and data parallel methods are used to improve the computing speed [52]. When the amount of data is quite large, the voting parallel strategy can also be adopted. Nowadays, LightGBM is widely adopted in distinct types of data mining tasks, such as classification, regression, sorting, etc.

Given a training dataset  $D = \{(x_i, y_i)\}, i \in [1, n]$ , the purpose of LightGBM is to look for an approximation  $\hat{Q}(x)$  to a certain function  $Q(x)$  which can minimize the expected value a specific loss function  $Loss(y, Q(x))$  as Eq. (24).

$$\hat{Q} = \arg \min [E_{y,D} Loss(y, Q(x))] \quad (24)$$

LightGBM incorporates a quantity of  $T$  regression trees  $\sum_{t=1}^T Q_t(D)$  to estimate the final model, which can be described in Eq. (25).

$$Q_T(D) = \sum_{t=1}^T Q_t(D) \quad (25)$$

The regression trees are represented as  $r_{p(x)}$ ,  $p \in \{1, 2, \dots, m\}$ , where  $m$  is the number of the tree leaves,  $p$  denotes the rules of the tree and  $r$  is the sample weight of the leaf nodes. For detailed steps of GOSS and EFB, please refer to reference [53] where the principle and process of LightGBM algorithm are clearly explained.

### 3. Proposed framework

The information fusion integrated forecasting framework proposed in this paper mainly consists of three modules: Depth feature extraction module, forecasting module and nonlinear integration module. The multi-source data are comprehensively and effectively screened and utilized in the feature extraction module, which lays the foundation for the next module. The collected data can be studied from multiple perspectives, information contained in the futures closing price and the impact of exogenous variables on the closing price are taken into consideration at the same time. The detailed structure of the framework is displayed in Fig. 3.

(1) In dealing with the closing price of futures, VMD is introduced to decompose the original data to obtain multiple IMFs. The information complexity of each IMF is evaluated through MSE, and the MSE threshold is set to reconstruct the characteristics of IMFs. IMFs with similar complexity are divided into one category. Before

reconstruction, the IMFs are smoothed by SG filter to avoid the interference of noise data to the data analysis process. At last, the reconstructed subsequence  $I = \{I_1, I_2, \dots, I_n\}$  are obtained.

(2) In the part of exogenous variable processing, firstly, Spearman correlation analysis is used to quantify the correlation between influencing factors and closing price. The factors with high correlation are retained. Secondly, the SAE with two hidden layers is introduced to further capture the intrinsic features of relevant factors, so as to simplify the input of prediction model and improve learning efficiency and prediction accuracy. Finally, the multi-dimensional features of the floor trading data and technical indicators are extracted as  $E = \{E_1, E_2, E_3\}$ .

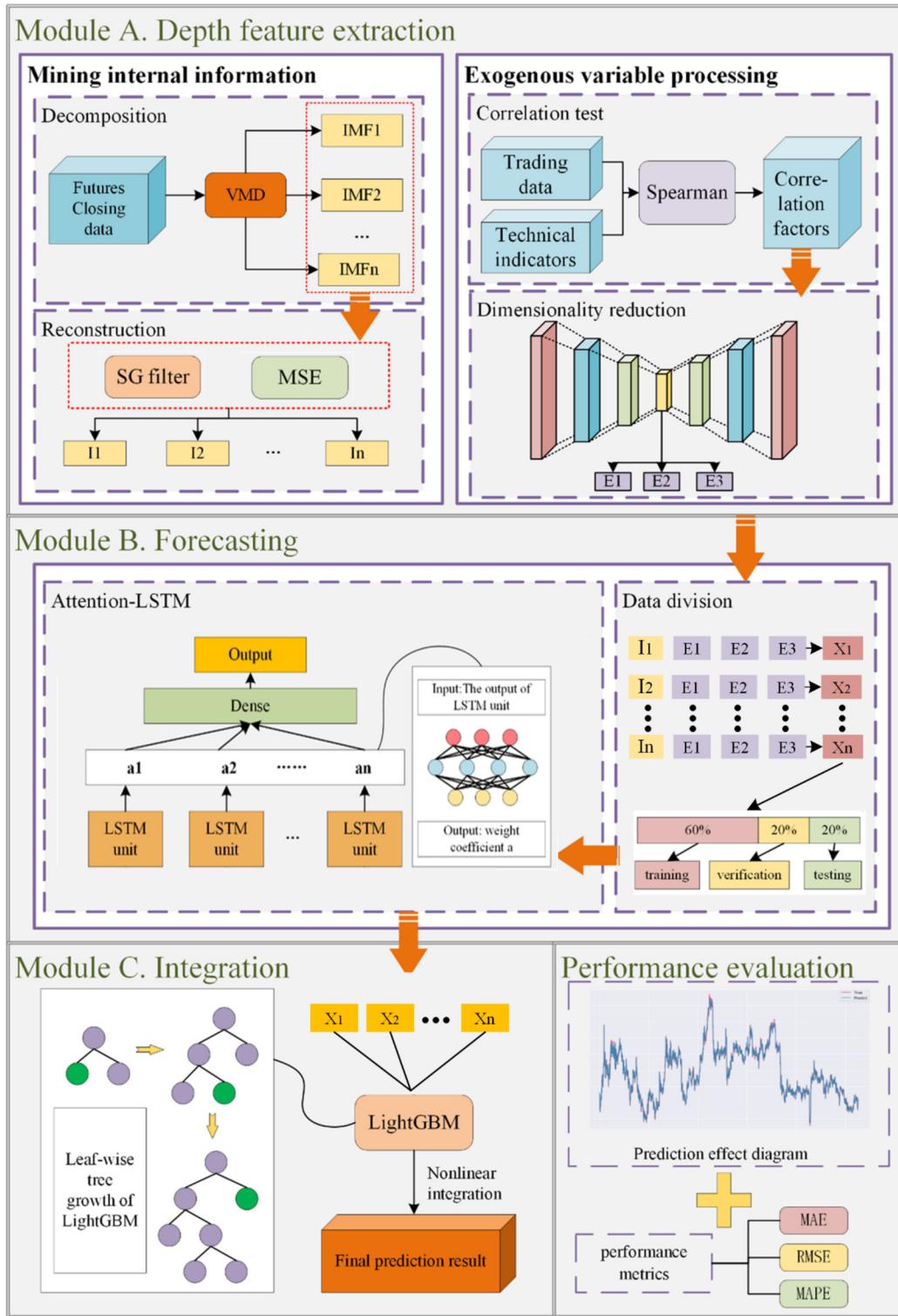
(3) After feature extraction, the data  $I$  and  $E$  will be used as input variables for the prediction model, which are divided into training set, validation set and test set according to the ratio of 6:2:2, respectively. The prediction target is the closing price data of the next 5 minutes.

(4) Then, appropriate prediction models are built for the two parts of the data, respectively. Different from the hard attention mechanism, this paper does not directly restrict the input information, but assigns the learning weight according to the importance difference of each input sequence. Since the location of the information contained in each sequence cannot be determined, the weights trained by neural network are adopted to weight the global input features in space or channel, for the sake of concentrating on the information in a specific spatial region. After the attention mechanism optimization, the LSTM assigns different attention weights to the characteristics of different relevance degrees, so as to avoid the problem of information overload and improve the computational efficiency. With this step, the predicted value of each reconstructed subsequence is obtained.

(5) Nonlinear integration is the final module. The predicted values of each reconstructed subsequence are nonlinearly integrated using the LightGBM algorithm to obtain the final prediction.

### 4. Simulation and discussion

In this section, in order to fully test the feasibility of the information fusion framework proposed in this paper for high-frequency futures closing price prediction, the 5-minute high-frequency closing prices of three representative futures in the Chinese futures market, namely, CSI 300 stock index futures, rebar futures and apple futures (referring to a type of fruit), are selected as benchmark data. In addition, several models from excellent papers in the field of financial time series prediction are introduced for comparative experiments. The selected data and prediction effectiveness evaluation criteria are described in detail below, and



**Fig. 3.** The proposed deep learning based information fusion integrated forecasting framework.

then the experimental procedure is given. At last, the experimental results are shown and discussed.

#### 4.1. Data collection and pre-processing

##### 4.1.1. Data pre-processing

This paper selects the closing prices of three main futures contracts for experiments, i.e., the contracts with the largest volume, are the most actively traded contracts in futures, and the data span

the period 2021/01/01 to 2021/12/31. Considering that contracts with different expiration dates will hardly ever trade at the same price, this could lead to problems with time series jumps if the next master contract is used directly to splice with it as it enters close to the delivery date. As shown in Fig. 4(A), if the data of different apple futures main contracts are directly spliced, the futures closing prices on the rolling days may jump significantly, which will undoubtedly have an impact on the subsequent modeling predictions.

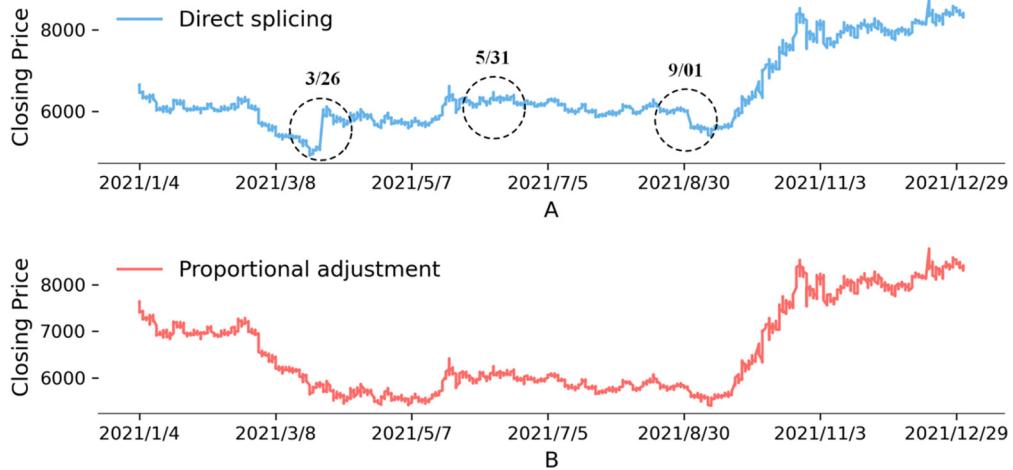


Fig. 4. Rolling adjustment of main contract data.

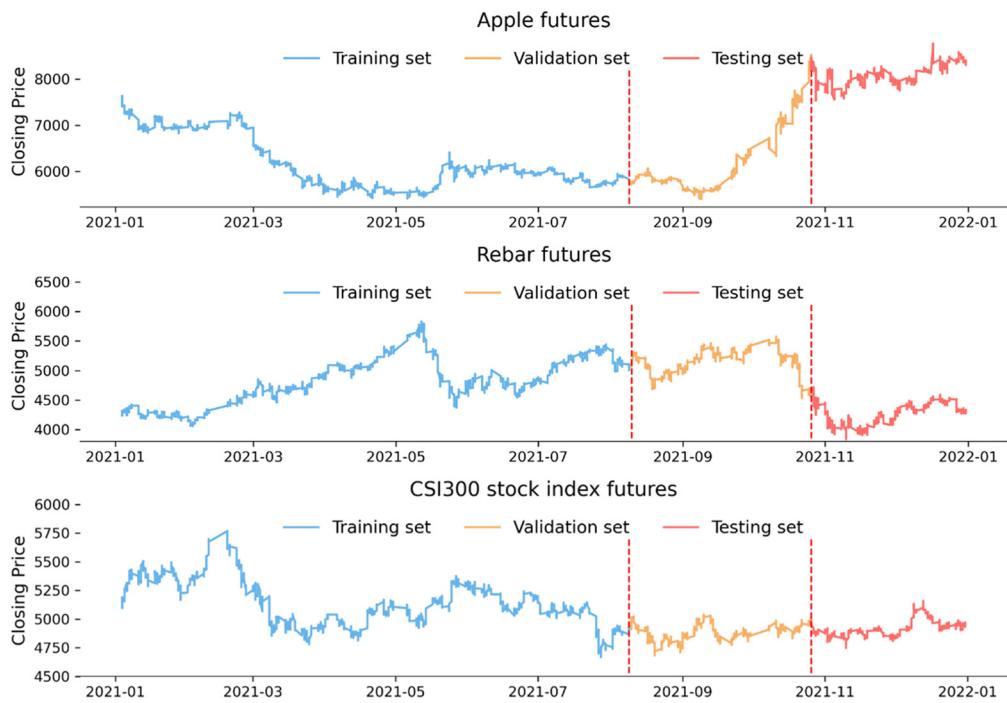


Fig. 5. The diagram of the trend of datasets and the sample division.

To ensure data continuity, this paper uses the proportional adjustment method to make rolling adjustments to futures prices on futures contract adjustment dates. The proportional adjustment method uses the proportion of the new contract price to the old contract price for relative adjustment. Taking apple futures as an example, January futures expired on March 25 and the price on the rolling date was RMB 5,131 per ton, while the opening price of March futures on March 26 was RMB 6,074. This means that all prices prior to the rolling date need to be multiplied by  $6074/5131$ , about 1.1838 times. The adjusted data for the main apple futures contract is shown in Fig. 4(B), which is more continuous.

Continuous contract data are obtained after rolling adjustment of the main contracts of the three futures. The first 60% of the observed data of each data set is used as the training set, the next 20% of the data is the validation set, and the last 20% portion is the test set. Fig. 5 plots the trend and division of the experimental data set, and the statistical characteristics of the data are shown

**Table 1**  
Statistical characteristics of the closing data of three futures.

	Max	Min	Mean	Median	Standard deviation
Apple	8776.00	5400.00	6553.26	6085.94	944.13
Rebar	5828.00	3836.00	4768.07	4768.96	431.479
CSI 300	5769.55	4666.80	5041.29	4981.58	185.49

in Table 1, including the maximum, minimum, mean, median and standard deviation.

At the same time, in the selection of exogenous variables, the relevant floor trading data and technical indicators utilized in this paper are shown in Table 2.

#### 4.1.2. Feature scaling

When dealing with multidimensional feature problems, it is necessary to scale the feature data in order to eliminate the differences in magnitude and order of magnitude between them. This

**Table 2**  
Selected exogenous variables.

Exogenous variables	Explanation
Open	The trading price of the first futures contract per 5 minutes
High	The maximum trading price of a futures contract per 5 minutes
Low	The minimum trading price of a futures contract per 5 minutes
Volume	The number of futures contracts traded in 5 minutes
SMA	Simple moving average.
WMA	Weighted moving average.
EMA	Exponential moving average.
MACD	Moving average convergence and divergence.
ATR	Average true range.
SAR	Stop and reverse index.
RSI	Relative strength index.
ROC	Rate of change.
CCI	Commodity channel index.
OBV	On balance volume.

helps the neural network to converge faster and improve the performance of the prediction model. In this paper, we use classical normalization to implement feature scaling by limiting the data range to [0,1] with the following equation:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (26)$$

where  $x$  is the original data and  $x'$  is the normalized data.

#### 4.2. Evaluation of prediction accuracy

In consideration of the difficulty of measuring the prediction effect of the proposed framework comprehensively by a single evaluation standard, three common indicators are employed as regression performance indexes: root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The smaller the value of each error index, the more accurate the prediction is. The particular formulas are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^k (\hat{y}_i - y_i)^2}{k}} \quad (27)$$

$$MAE = \frac{1}{k} \sum_{i=1}^k |\hat{y}_i - y_i| \quad (28)$$

$$MAPE = \frac{1}{k} \sum_{i=1}^k \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (29)$$

where  $k$  denotes the number of values of the testing set,  $\hat{y}$  is the ultimate forecasting result,  $y$  represents the true value of the testing set.

#### 4.3. Experimental process

##### 4.3.1. Depth feature extraction

In this part, the closing price, floor trading data and technical indicators are comprehensively and effectively screened and utilized. The internal implied characteristics of the closing price and the impact of exogenous variables on the target data are deeply mined. Considering there is correlation between transaction data and technical indicators, they are combined for feature extraction to eliminate redundant information.

##### (a) VMD-based data decomposition

**Table 3**  
MSE values of IMFs.

Scale factor	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
1	0.0028	0.0959	0.3530	0.5212	0.5804	0.6018
3	0.0083	0.3177	0.5938	0.6318	0.8254	1.0746
5	0.0139	0.4716	0.6659	0.8515	1.1157	0.6431
mean	0.0083	0.2951	0.5375	0.6682	0.8405	0.7732

**Table 4**  
MSE value of IMF after denoising.

Scale factor	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
1	0.0027	0.0955	0.3523	0.5202	0.5779	0.6012
3	0.0082	0.3172	0.5901	0.6139	0.7151	0.9026
5	0.0138	0.4710	0.6589	0.8191	1.0705	0.5123
mean	0.0082	0.2946	0.5338	0.6511	0.7878	0.6720

To effectively deal with complex volatile futures closing price series, this paper first uses the VMD algorithm to decompose them into several simpler, predictable components to extract the intrinsic information in the futures closing price series. Firstly, the VMD decomposition is performed by observing and analyzing the center frequencies of different modal components to determine the most appropriate number of decompositions  $k$ . Taking apple futures as an example, when  $k$  is 2, 3, 4, 5 and 6, there is no confounding of center frequencies. When  $k$  is 7, the center frequencies are mixed, so the apple futures closing price data is decomposed into 6 modal components. Similarly, the closing price data of rebar and CSI 300 futures are decomposed into 6 and 5 modal components, respectively.

Taking apple futures as an example, the decomposition results of apple futures closing price series are shown in Fig. 6(A).

##### (b) Signal Denoising and Reconstruction Based on MSE and SG Filter

After decomposing the futures closing price series into several IMFs using VMD, each IMF still contains complex noise information, which undoubtedly affects the subsequent predictive modeling. The MSE algorithm can measure the complexity of the sequence. In this paper, MSE is used to calculate the multiscale entropy values and mean values of each order IMF component at different scales, with  $m$  set to 2,  $u$  set to 0.15, and the scale factors  $\psi$  chosen to be 1, 3, 5. Taking apple futures as an example, the MSE values of each IMF component are shown in Table 3.

It can be found that the MSE values of different IMFs vary, and the higher the complexity of the IMF and the more noise it has, the larger its MSE value is. In order to reduce the noise in the data and improve the accuracy of prediction, this paper uses SG filter to comb the data for noise reduction. After several experiments, it was determined that effective denoising could be achieved when the sliding window width of SG was set to 19 and the polynomial fitting order was set to 3. Taking apple futures as an example, each IMF component after denoising is shown in Fig. 6(B), and its MSE values are shown in Table 4.

The results demonstrate that the MSE values, i.e., the complexity of each IMF is reduced to some extent after the SG filter denoising. In addition, to reduce the computational complexity of the prediction model, this paper uses the mean MSE value as a reconstruction indicator, where IMFs with similar mean MSE values have similar complexity and are reconstructed into a new component. Taking apple futures as an example, IMFs can be reconstructed into three new components  $I = \{I_1, I_2, I_3\}$ , where  $I_1 = IMF_1$ ,  $I_2 = IMF_2$ ,  $I_3 = IMF_3 + IMF_4 + IMF_5 + IMF_6$ . Ultimately, the reconstructed apple futures closing price series is shown in Fig. 6(C). The results of decomposition, denoising and reconstruction of the rebar futures and CSI 300 futures data are shown in Fig. 7 and Fig. 8.

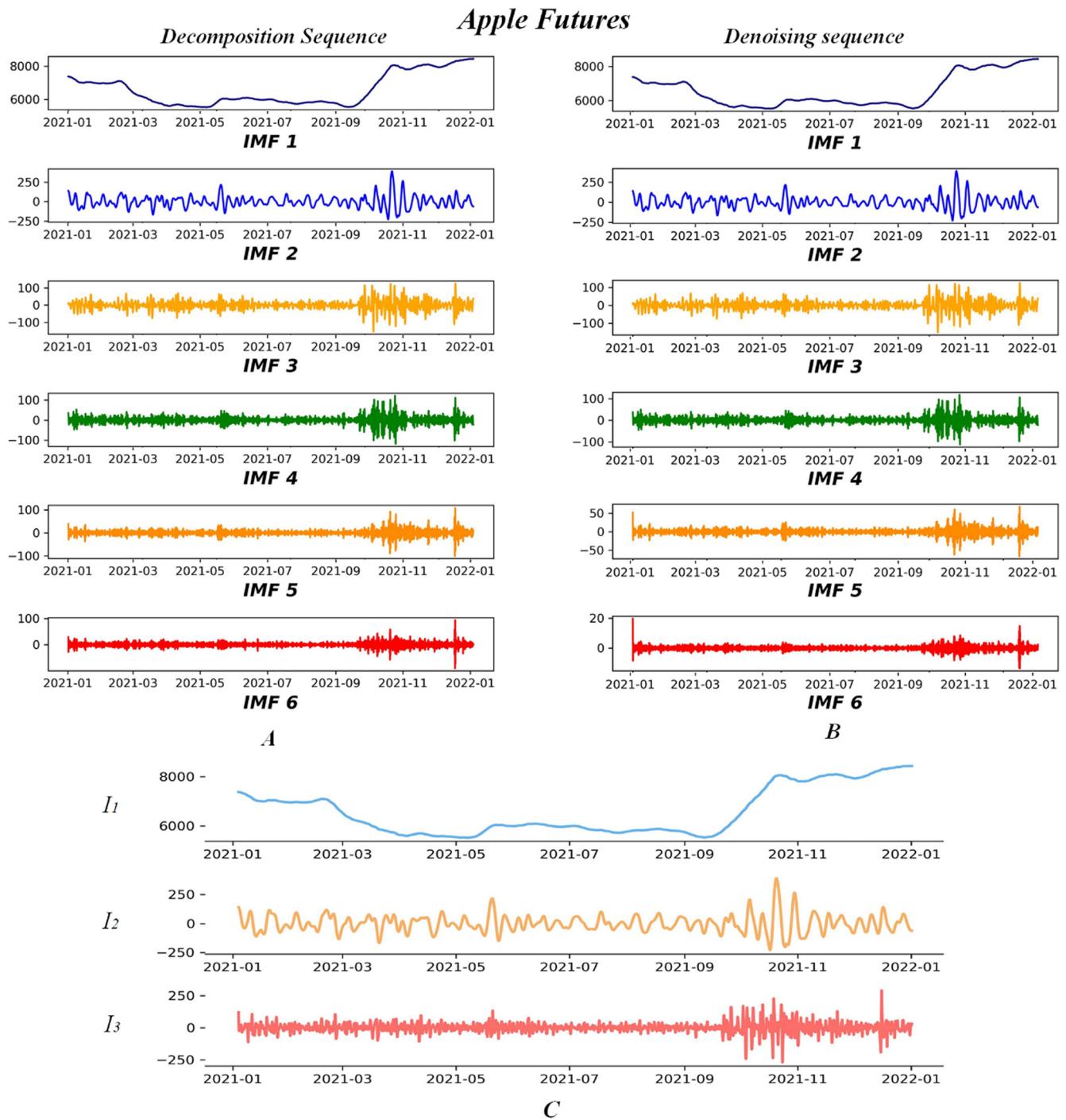


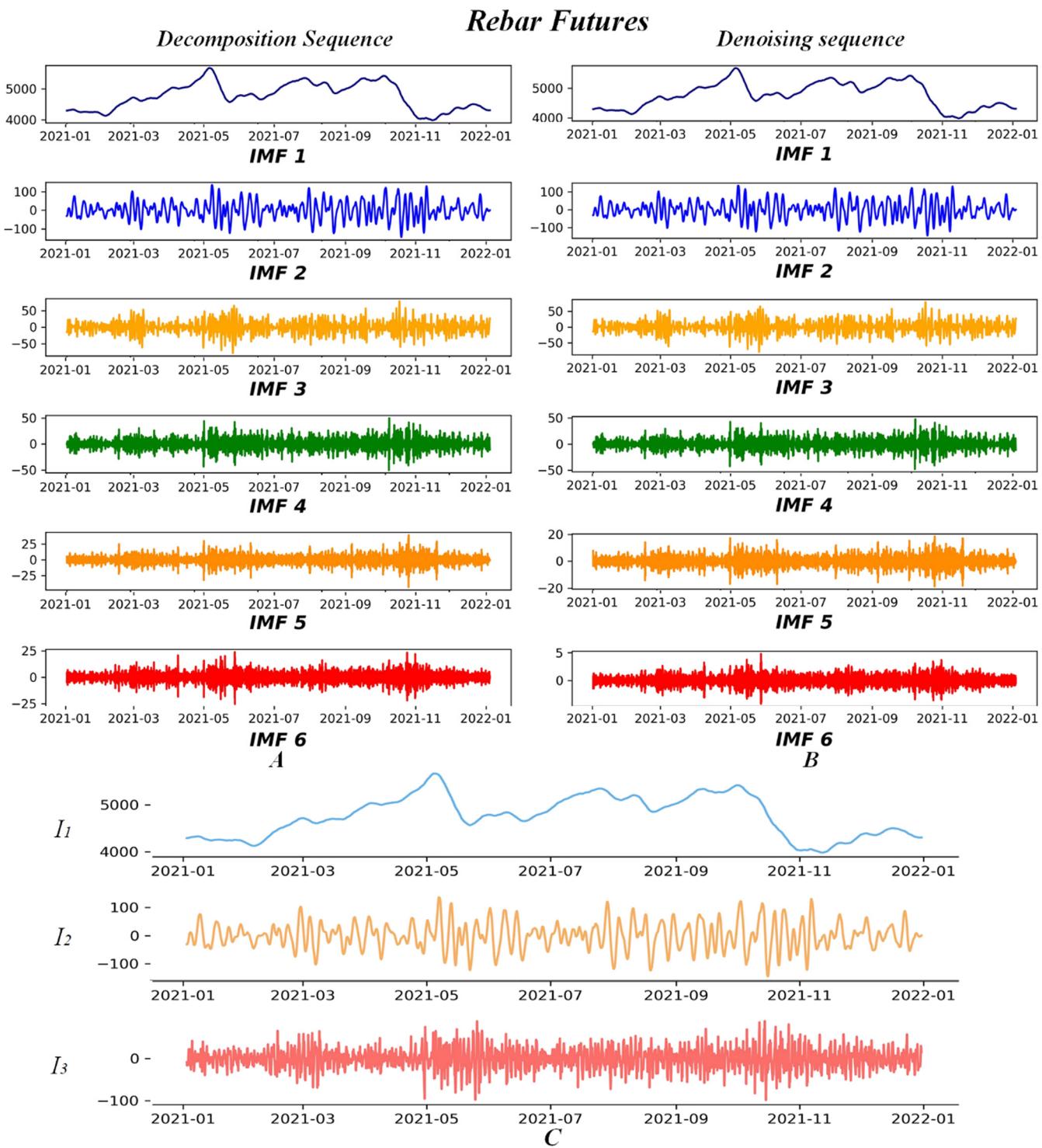
Fig. 6. The process of decomposition, denoising and reconstruction of apple futures data.

### (c) Feature extraction of exogenous variables

Due to SAE can only learn the high-level features of data and cannot guarantee the correlation between input factors and target data, Spearman correlation test is introduced before SAE feature engineering for nonlinear data processing. The correlation test results are shown in Table 5. The range of correlation coefficient is  $[-1, 1]$ . The closer it approaches to 0, the more it represents that there is no correlation between the two variables. In this experiment, the variables with correlation coefficient  $>0.5$  or coefficient  $<-0.5$  were selected as strong correlation variables. It is clear from the table that there is a strong correlation between price factors, average indices, SAR of technical indicators and closing

prices during the 5-minute period, so they are retained as inputs to SAE.

After Spearman correlation test, eight exogenous variables were selected. In order to further extract more abstract and effective intrinsic features from the exogenous variables and reduce the computational complexity of the model. In this paper, we use SAE to further extract the intrinsic features from the 8 exogenous variables. After extensive experiments, the structure of SAE in this experiment is set as follows: two hidden layers, and the nodes of the first and second layers are set to 5 and 3, respectively. All hidden layers are trained by random batch gradient descent method. Sigmoid is chosen as the activation function of each layer to improve the convergence speed of the network. Next, Adam is used



**Fig. 7.** The process of decomposition, denoising and reconstruction of rebar futures data.

as the optimizer and the mean absolute error function is utilized as the loss function. Set the batch size of each layer to 256. The default number of iterations is set to 5000.

Considering that the deep network structure may introduce overfitting problems, the early stop method is introduced, which has better performance than regularization in many application cases. The early stop method sets a given number of iterations, called *patience*. The training process is terminated if the average absolute error of the model on the validation set is not reduced after *patience* iterations. Through experimentation and validation,

the optimal patience is set to 80. Taking the exogenous variables of apple futures as an example, the intrinsic features of the eight exogenous variables are highly extracted into three new features  $\{E_1, E_2, E_3\}$ , and the extracted features are shown in Fig. 9.

#### 4.3.2. Forecasting by ALSTM

After the depth feature extraction module, the reconstructed subsequences and the depth extracted features are used as input variables to construct the corresponding prediction models. First, the input is processed into the required 3D structure for ALSTM.

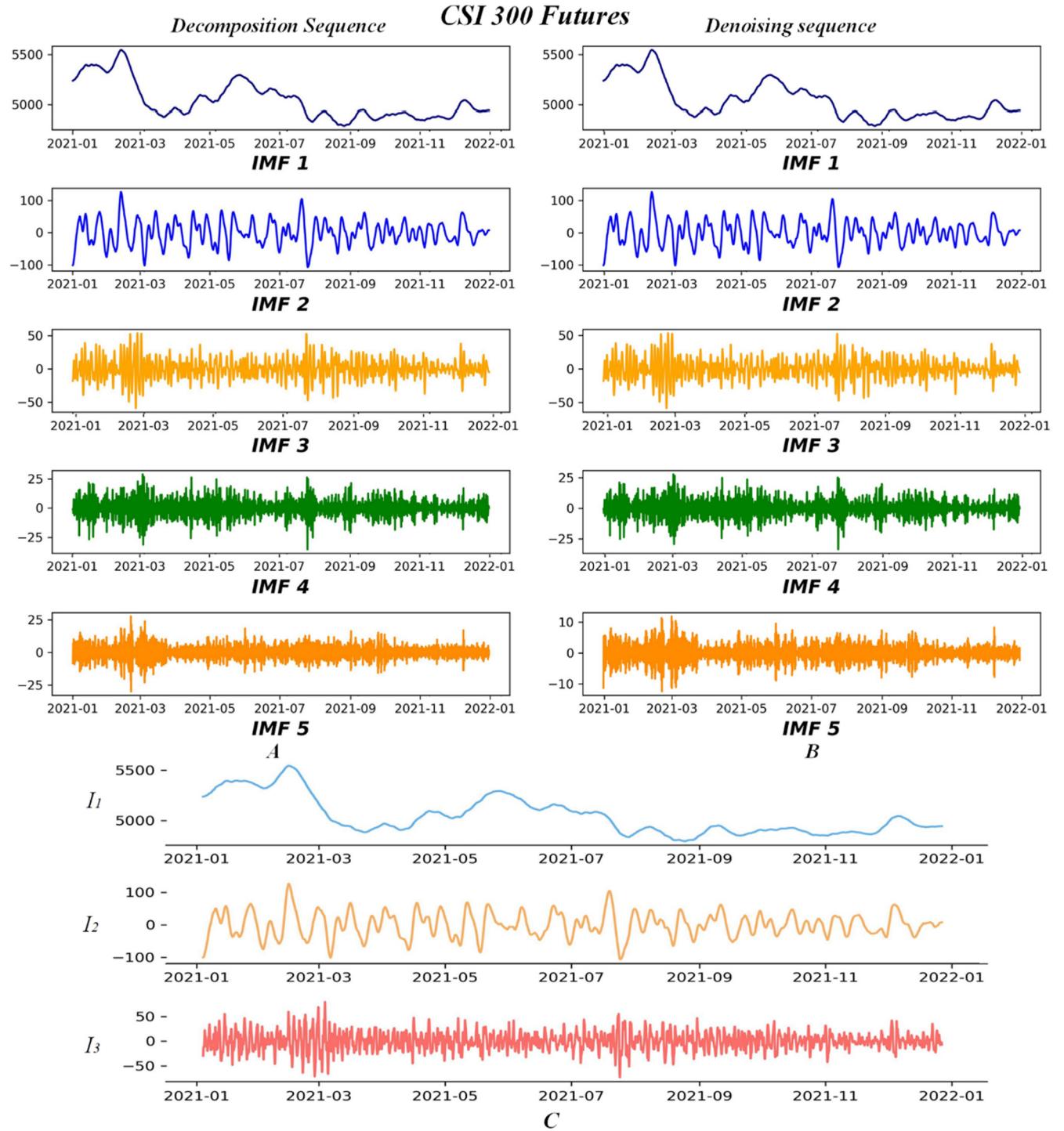


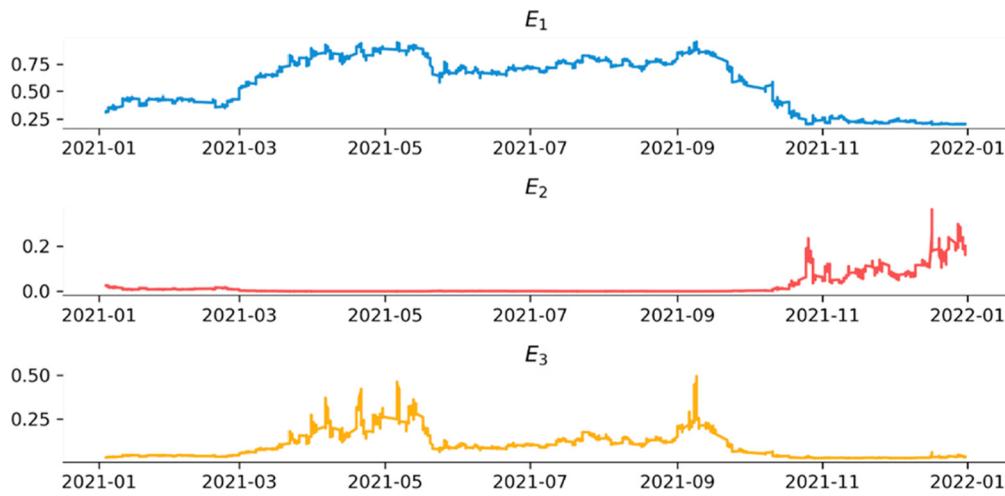
Fig. 8. The process of decomposition, denoising and reconstruction of CSI 300 futures data.

Then the input is processed through the LSTM layer, the output of the hidden layer is fused through the attention layer, and finally mapped to the output space through the perceptual layer.

Note that the learning rate has an important effect on accelerating the convergence rate and improving the accuracy of the model, the learning rate decreasing algorithm is introduced in the model constructing. Set the initial learning rate as 0.05. If the verification loss does not decrease after 10 training epochs, the learning rate will be halved and the minimum learning rate is 0.001. Meanwhile, the over fitting problem caused by complex network structure is also worthy of attention. Thus, the regularization strategies are ap-

plied to avoid the problem, chiefly including Dropout mechanism and early stopping method. Dropout mechanism randomly discards the neurons of hidden layers, preventing over reliance on some local features. The early stopping method is same as that described in Section 4.3.1 (c).

Meanwhile, the other hyper-parameters adjustment is realized through substantial experiments and priori experience: look back is set to 4, batch size equals to 128, neurons is set to 64, activation is 'sigmoid', and optimizer is 'Adam', the loss function is 'MSE', the upper limit of epochs is 1000.



**Fig. 9.** SAE extracts the intrinsic characteristics of exogenous variables of apple futures.

**Table 5**  
The results of Spearman correlation test.

	Apple	Rebar	CSI300
Open	0.999470	0.999658	0.998697
High	0.999750	0.999837	0.999414
Low	0.999745	0.999851	0.999360
Volume	0.022822	-0.064881	-0.031402
SMA	0.999318	0.999551	0.998228
WMA	0.99962	0.999756	0.999035
EMA	0.999539	0.999702	0.998829
MACD	-0.017352	-0.009586	0.008572
ATR	0.585609	0.179287	0.305897
SAR	0.993671	0.995712	0.984250
RSI	0.079383	0.086480	0.124043
ROC	0.040764	0.037245	0.074151
CCI	0.041481	0.045590	0.078812
OBV	0.639645	0.925446	0.815198

**Table 6**  
The parameters setting of LightGBM.

Parameter name	Explanation	Parameter value
Feature fraction	The proportion of feature selection for tree building	0.8
Num leaves	Number of leaf nodes	511
Max depth	Maximum depth of the tree	11
Bagging fraction	The proportion of sample selection for tree building	0.8
Num iterations	Maximum number of iterations	75
Learning rate	Learning rate of training model	0.05

#### 4.3.3. Nonlinear ensemble learning by LightGBM

After obtaining the predicted values of each reconstructed subsequence by the ALSTM model, the predicted values of all reconstructed subsequences are finally integrated. For example, the outputs of the predicted values of the three reconstructed subsequences  $\{I_1, I_2, I_3\}$  of the apple futures closing price dataset are pooled together as the input variables of the LightGBM algorithm and fitted to obtain the final prediction results. In order to improve the generalization of the model, a grid search algorithm was used to find the optimal parameters. Table 6 shows the selection of the important parameters involved.

To show the experimental results more visually, the final prediction results and evaluation indicators of the framework proposed in this paper are shown in Fig. 10. As can be seen in Fig. 10, although some values do not achieve fully accurate predictions, the

hybrid model effectively predicts the overall trend of futures. The experimental results prove that the framework can effectively predict the closing price of the futures and provide more reference information for high-frequency futures trading.

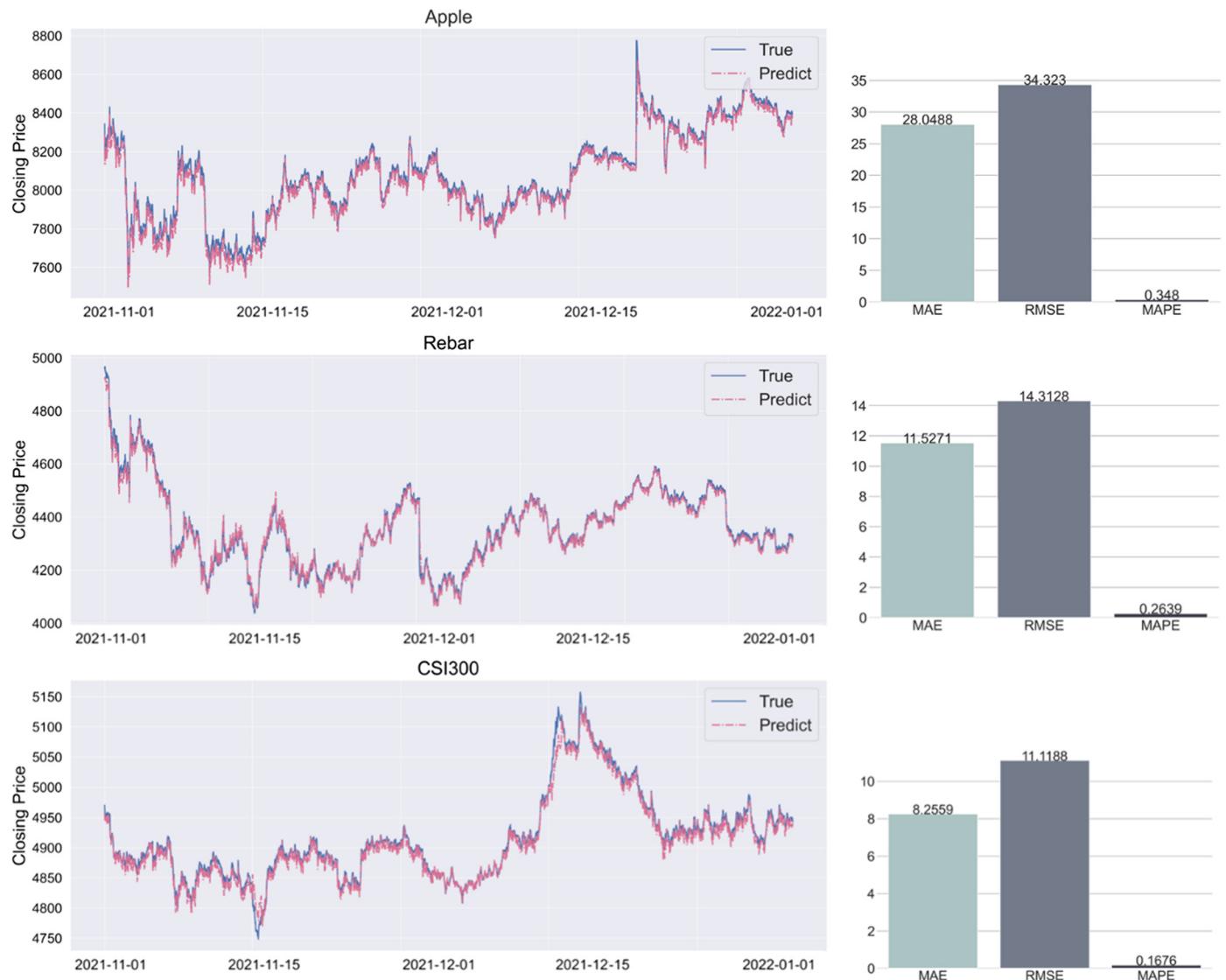
#### 4.4. Comparative experiment

To further validate the effectiveness of the forecasting framework proposed in this paper, we construct six benchmark models for comparative experiments, which are chosen to be typical of the best results in the field of financial time series forecasting. In addition, the parameters of the benchmark models are set by the methods introduced in the related literature. The forecasting performance and computational complexity (computation time) of the comparison models are shown in Table 7.

It is obvious from Table 7 that the prediction performance of the single model is slightly weaker than that of the hybrid model. Although multilayer perceptron neural network (MLP) and LSTM can learn some features of the data, the nonlinearity and high complexity of the unprocessed data make the prediction performance unsatisfactory. In addition, LSTM is more suitable for time series forecasting in finance than MLP. The memory mechanism of LSTM also makes it more computationally complex, and its computation time is about twice that of MLP.

Meanwhile, the prediction performance of VMD-bidirectional gated unit neural network (BiGRU) and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)-LSTM using the decomposition integration technique is significantly improved. Taking apple futures data as an example, the RMSE, MAE and MAPE of VMD-BiGRU are reduced by 30.84%, 27.62% and 52.87%, respectively, compared with LSTM. Therefore, the decomposition integration technique can effectively improve the prediction accuracy of the model. At the same time, modeling prediction for all subsequences after decomposition also substantially increases the computational complexity of the model. Its computation time is about 10 times that of the LSTM model.

More, VMD-LSTM-support vector machine (SVR) introduces nonlinear integration technique, which effectively improves the accuracy of prediction. In terms of MAPE, it is reduced by 71.11%, 53.30%, and 66.82%, respectively, compared to VMD-BiGRU on the three data sets. Therefore, the advantages of the nonlinear integration technique are significant. CEEMDAN-sample entropy (SE)-LSTM-RF uses SE to recombine decomposition sequences of similar complexity on the basis of the decomposition nonlinear integration technique, and its computational complexity is reduced to some extent, which reduces the risk of computational



**Fig. 10.** Comparison diagram of prediction results and true values.

error accumulation. As a result, its computation time for the three datasets is reduced by 24.63%, 21.73% and 25.27%, respectively, compared with VMD-LSTM-SVR, and its prediction performance is improved.

The prediction performance of the framework proposed in this paper outperforms these six benchmark models. From the overall comparison results, although CEEMDAN-SE-LSTM-RF has good prediction accuracy and stability, it is still difficult to surpass the method developed in this paper. In this paper, based on the decomposition integration strategy, SG filter is introduced to reduce the noise present in the sequence, and MSE is used to reconstruct the decomposition sequence with similar complexity, which reduces the complexity of prediction and improves the accuracy of prediction. In addition, this paper introduces exogenous variables and extracts the intrinsic features in the exogenous variables to reduce the impact of redundancy of exogenous variables on the prediction model. Finally, the final prediction results are obtained by using LightGBM nonlinear integration. The comparative experimental results prove that the framework proposed in this paper is an effective tool for high-frequency futures closing price prediction and can provide an effective reference for investors' trading.

## 5. Conclusion

In this paper, a novel information fusion synthesis framework is developed for 5-minute high-frequency futures closing price forecasting. DFE-ALSTM-LightGBM integrates three modules, which includes depth feature extraction, forecasting and integration. In this paper, closing price, floor trading data and technique indexes are selected as fundamental data, among them, floor trading data and technique indexes are combined as exogenous variables for analysis. In the first module, VMD algorithms supported by MSE, and SG filter are developed to handle closing prices, as well as enhanced dimensionality reduction methods including Spearman correlation analysis and SAE to extract intrinsic features of exogenous variables. In the forecasting module, ALSTM is adopted to gain more comprehensive information and allocate reasonable weight features with different degrees of importance. Finally, in the integration module, LightGBM is employed to aggregate the forecast values and produce the ultimate futures closing price prediction results. The simulation results show that the proposed framework can achieve the predictions with higher accuracy and better robustness compared with other latest forecasting methods. Therefore, it is a powerful tool for high frequency futures closing price forecasting.

**Table 7**

The prediction effect of the comparison models and proposed framework.

Model	Futures	RMSE	MAE	MAPE (%)	Computation time (minutes)
MLP [54]	Apple	85.2341	65.2131	9.2326	1.6
	Rebar	54.8722	46.4365	7.7564	1.2
	CSI300	42.5322	39.8644	6.1255	1.3
LSTM [55]	Apple	80.3412	63.1092	8.0234	2.8
	Rebar	50.0122	44.6752	7.1255	2.1
	CSI300	39.4451	34.2351	5.9822	2.2
VMD-BiGRU [56]	Apple	55.5612	45.1244	3.7812	21.4
	Rebar	39.1991	34.8749	2.1015	18.1
	CSI300	35.6215	27.9124	2.6414	18.5
CEEMDAN-LSTM [57]	Apple	57.1294	49.9851	3.1242	22.4
	Rebar	38.1293	32.5831	2.8713	19.7
	CSI300	35.6871	30.9831	2.9814	19.4
VMD-LSTM-SVR [58]	Apple	44.2355	37.0456	1.0923	20.3
	Rebar	27.9864	19.9742	0.9814	18.4
	CSI300	23.8752	16.5418	0.8764	18.2
CEEMDAN-SE-LSTM-RF [59]	Apple	39.0987	32.1131	0.8765	15.3
	Rebar	20.8761	14.9865	0.5341	14.4
	CSI300	18.5991	12.8731	0.4421	13.6
The proposed framework	Apple	34.3230	28.0488	0.3480	16.3
	Rebar	14.3128	11.5271	0.2639	15.2
	CSI300	11.1188	8.2559	0.1676	15.1

Meanwhile, how to establish optimal proper forecast models can be the topic of further study. Realizing automatic parameter selection through optimization algorithm and taking more market factors into consideration are the main research directions. With the innovation of technology and the maturity of methods, new algorithms will be explored in the future research.

### CRediT authorship contribution statement

**Jujie Wang:** Conceptualization, Writing – review & editing. **Yu Chen:** Methodology, Software, Writing – original draft. **Shuzhou Zhu:** Investigation, Resources, Validation. **Wenjie Xu:** Investigation, Resources, Validation.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 71971122 and 71501101).

### References

- [1] J.W. Goodell, COVID-19 and finance: agendas for future research, *Finance Res. Lett.* 35 (2020) 101512.
- [2] I. Sifat, A. Ghafoor, A. Ah Mand, The COVID-19 pandemic and speculation in energy, precious metals, and agricultural futures, *J. Behav. Exp. Finance* 30 (2021) 100498.
- [3] Y. Liu, Z. Niu, M.T. Suleman, L. Yin, H. Zhang, Forecasting the volatility of crude oil futures: the role of oil investor attention and its regime switching characteristics under a high-frequency framework, *Energy* 238 (2022) 121779.
- [4] K.P. Evans, Intraday jumps and US macroeconomic news announcements, *J. Bank. Finance* 35 (2011) 2511–2527.
- [5] H.R. Stoll, R.E. Whaley, The dynamics of stock index and stock index futures returns, *J. Financ. Quant. Anal.* 25 (1990) 441–468.
- [6] R. Brooks, Power arch modelling of the volatility of emerging equity markets, *Emerg. Mark. Rev.* (2007) 124–133.
- [7] T. Bunnag, Hedging petroleum futures with multivariate GARCH models, *Int. J. Energy Econ. Policy* 5 (2015) 105–120.
- [8] W. Huang, Z. Huang, M. Matei, T. Wang, Price volatility forecast for agricultural commodity futures: the role of high frequency data, *J. Econ. Forecast.* (2012) 83–103.
- [9] B. Zhu, S. Ye, P. Wang, J. Chevallier, Y. Wei, Forecasting carbon price using a multi-objective least squares support vector machine with mixture kernels, *J. Forecast.* 41 (2022) 100–117.
- [10] P. Jiang, R. Li, N. Liu, Y. Gao, A novel composite electricity demand forecasting framework by data processing and optimized support vector machine, *Appl. Energy* 260 (2020) 114243.
- [11] B. Zhu, A novel multiscale ensemble carbon price prediction model integrating empirical mode decomposition, genetic algorithm and artificial neural network, *Energies* 5 (2012) 1–16.
- [12] C. Tian, Y. Hao, Point and interval forecasting for carbon price based on an improved analysis-forecast system, *Appl. Math. Model.* 79 (2020) 126–144.
- [13] S. Kulkarni, I. Haidar, Forecasting model for crude oil price using artificial neural networks and commodity futures prices, *Comput. Sci.* (2009).
- [14] Z. Hajabotorabi, F.F. Samavati, F.M. Maalek Ghaini, A. Shahmoradi, Multi-WRNN model for pricing the crude oil futures market, *Expert Syst. Appl.* 182 (2021) 115229.
- [15] L. Hui, L. Fang, C. Dihuang, A novel wind DC microgrid energy management strategy based on LSTM forecast model, *J. Phys. Conf. Ser.* 1871 (2021) 12014.
- [16] Y. Huang, X. Dai, Q. Wang, D. Zhou, A hybrid model for carbon price forecasting using GARCH and long short-term memory network, *Appl. Energy* 285 (2021) 116485.
- [17] W. Lu, J. Li, Y. Li, A. Sun, J. Wang, A CNN-LSTM-based model to forecast stock prices, *Complexity* (2020) (2020) 1–10.
- [18] L. Zhang, J. Wang, B. Wang, Energy market prediction with novel long short-term memory network: case study of energy futures index volatility, *Energy* 211 (2020) 118634.
- [19] J. Wang, Q. Cui, X. Sun, A novel framework for carbon price prediction using comprehensive feature screening, bidirectional gate recurrent unit and Gaussian process regression, *J. Clean. Prod.* 314 (2021) 128024.
- [20] Y. Li, S. Wang, Y. Wei, Q. Zhu, A new hybrid VMD-ICSS-BiGRU approach for gold futures price forecasting and algorithmic trading, *IEEE Trans. Comput. Soc. Syst.* 8 (2021) 1357–1368.
- [21] J. Wang, J. Cao, S. Yuan, M. Cheng, Short-term forecasting of natural gas prices by using a novel hybrid method based on a combination of the CEEMDAN-SE- and the PSO-ALS-optimized GRU network, *Energy* 233 (2021) S1828348590X.
- [22] B. Zhu, P. Wang, J. Chevallier, Y. Wei, Carbon price analysis using empirical mode decomposition, *Comput. Econ.* 45 (2015) 195–206.
- [23] H. Liu, Z. Long, An improved deep learning model for predicting stock market price time series, *Digit. Signal Process.* 102 (2020) 102741.

- [24] L. Yu, S. Wang, K.K. Lai, An EMD-Based Neural Network Ensemble Learning Model for World Crude Oil Spot Price Forecasting, Springer Berlin Heidelberg, 2008.
- [25] J. Zhu, P. Wu, H. Chen, J. Liu, L. Zhou, Carbon price forecasting with variational mode decomposition and optimal combined model, *Phys. A, Stat. Mech. Appl.* 519 (2019) 140–158.
- [26] W. Liu, C. Wang, Y. Li, Y. Liu, K. Huang, Ensemble forecasting for product futures prices using variational mode decomposition and artificial neural networks, *Chaos Solitons Fractals* 146 (2021) 110822.
- [27] Y. Hao, C. Tian, A hybrid framework for carbon trading price forecasting: the role of multiple influence factor, *J. Clean. Prod.* 262 (2020) 120378.
- [28] M. He, Y. Zhang, D. Wen, Y. Wang, Forecasting crude oil prices: a scaled PCA approach, *Energy Econ.* 97 (2021) 105189.
- [29] D. Bredin, C. O'Sullivan, S. Spencer, Forecasting WTI crude oil futures returns: does the term structure help?, *Energy Econ.* 100 (2021) 105350.
- [30] A. Ali, F. Yangyu, S. Liu, Automatic modulation classification of digital modulation signals with stacked autoencoders, *Digit. Signal Process.* 71 (2017) 108–116.
- [31] Y. Zhang, J. Le, X. Liao, F. Zheng, Y. Li, A novel combination forecasting model for wind power integrating least square support vector machine, deep belief network, singular spectrum analysis and locality-sensitive hashing, *Energy* 168 (2019) 558–572.
- [32] E. Torun, T. Chang, R.Y. Chou, Causal relationship between spot and futures prices with multiple time horizons: a nonparametric wavelet Granger causality test, *Res. Int. Bus. Finance* 52 (2020) S1177327463.
- [33] Y. Yin, Model-free tests for series correlation in multivariate linear regression, *J. Stat. Plan. Inference* 206 (2020) 179–195.
- [34] G. Liao, T. Tsao, Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting, *IEEE Trans. Evol. Comput.* 10 (2006) 330–340.
- [35] L.T. Bui, V. Truong Vu, T.T. Huong Dinh, A novel evolutionary multi-objective ensemble learning approach for forecasting currency exchange rates, *Data Knowl. Eng.* 114 (2018) 40–66.
- [36] Q. Gu, Y. Chang, N. Xiong, L. Chen, Forecasting Nickel futures price based on the empirical wavelet transform and gradient boosting decision trees, *Appl. Soft Comput.* 109 (2021) 107472.
- [37] S. Ben Jabeur, R. Khalfaoui, W. Ben Arfi, The effect of green energy, global environmental indexes, and stock markets in predicting oil price crashes: evidence from explainable machine learning, *J. Environ. Manag.* 298 (2021) 113511.
- [38] A. Shehadeh, O. Alshboul, R.E. Al Mamlook, O. Hamedat, Machine learning models for predicting the residual value of heavy construction equipment: an evaluation of modified decision tree, LightGBM, and XGBoost regression, *Autom. Constr.* 129 (2021) 103827.
- [39] P. Xue, Y. Jiang, Z. Zhou, X. Chen, X. Fang, J. Liu, Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms, *Energy* 188 (2019) 116085.
- [40] Y. Wang, R. Markert, J. Xiang, W. Zheng, Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system, in: *Mechanical Systems and Signal Processing*, vol. 60–61, 2015, pp. 243–251.
- [41] Q. Wang, L. Wang, H. Yu, D. Wang, A.K. Nandi, Utilizing SVD and VMD for denoising non-stationary signals of roller bearings, *Sensors (Basel)* (2021) 22.
- [42] Z. Jin, D. He, R. Ma, X. Zou, Y. Chen, S. Shan, Fault diagnosis of train rotating parts based on multi-objective VMD optimization and ensemble learning, *Digit. Signal Process.* 121 (2022) 103312.
- [43] G. Zhang, H. Liu, J. Zhang, Y. Yan, L. Zhang, C. Wu, et al., Wind power prediction based on variational mode decomposition multi-frequency combinations, *J. Mod. Power Syst. Clean Energy* 7 (2019) 281–288.
- [44] G. Xing, E. Zhao, C. Zhang, J. Wu, A decomposition-ensemble approach with denoising strategy for PM 2.5 concentration forecasting, *Discrete Dyn. Nat. Soc.* (2021) (2021) 1–13.
- [45] H. Luo, T. Qiu, C. Liu, P. Huang, Research on fatigue driving detection using forehead EEG based on adaptive multi-scale entropy, *Biomed. Signal Process. Control* 51 (2019) 50–58.
- [46] S. Zou, T. Qiu, P. Huang, X. Bai, C. Liu, Constructing multi-scale entropy based on the empirical mode decomposition (EMD) and its application in recognizing driving fatigue, *J. Neurosci. Methods* 341 (2020) 108691.
- [47] Y. Kaur, G. Ouyang, M. Junge, W. Sommer, M. Liu, C. Zhou, et al., The reliability and psychometric structure of Multi-Scale Entropy measured from EEG signals at rest and during face and object recognition tasks, *J. Neurosci. Methods* 326 (2019) 108343.
- [48] H.L. Kennedy, Improving the frequency response of Savitzky-Golay filters via colored-noise models, *Digit. Signal Process.* 102 (2020) 102743.
- [49] W. Huang, Z. Huang, M. Matei, T. Wang, Price volatility forecast for agricultural commodity futures: the role of high frequency data, *J. Econ. Forecast.* (2012) 83–103.
- [50] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 313 (2006) 504–507.
- [51] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, S. Dong, A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction, *Phys. A, Stat. Mech. Appl.* 583 (2021) 126293.
- [52] N. Wang, G. Zhang, W. Pang, L. Ren, Y. Wang, Novel monitoring method for material removal rate considering quantitative wear of abrasive belts based on LightGBM learning algorithm, *Int. J. Adv. Manuf. Technol.* 114 (2021) 3241–3253.
- [53] J. Yan, Y. Xu, Q. Cheng, S. Jiang, Q. Wang, Y. Xiao, et al., LightGBM: accelerated genomically designed crop breeding through ensemble learning, *Genome Biol.* 22 (2021) 271.
- [54] X. Fan, S. Li, L. Tian, Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model, *Expert Syst. Appl.* 2 (2015) 3945–3952.
- [55] K. Rajakumari, M. Kalyan, M. Bhaskar, Forward forecast of stock price using LSTM machine learning algorithm, *Int. J. Comput. Theory Eng.* 12 (2020) 74–79.
- [56] Q. Zhu, F. Zhang, S. Liu, Y. Wu, L. Wang, A hybrid VMD-BiGRU model for rubber futures time series forecasting, *Appl. Soft Comput.* 84 (2019) 1568–4946.
- [57] Y. Lin, Y. Yan, J. Xu, Y. Liao, F. Ma, Forecasting stock index price using the CEEMDAN-LSTM model, *N. Am. J. Econ. Finance* 57 (2021) 1062–9408.
- [58] K. Zhang, H. Cao, J. Thé, H. Yu, A hybrid model for multi-step coal price forecasting using decomposition technique and deep learning algorithms, *Appl. Energy* 306 (2022) 0306.
- [59] J. Wang, X. Sun, Q. Cheng, Q. Cui, An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting, *Sci. Total Environ.* 762 (2021) 0048.

**Jujie Wang** is an associate professor in the School of Management Science and Engineering of Nanjing University of Information Science and Technology, China. His research interests are in the areas of Signal Processing, Time Series Analysis, Machine Learning and Data Mining.

**Yu Chen** is a graduate student in the School of Management Science and Engineering of Nanjing University of Information Science and Technology, China. Her current research fields include Signal Processing, Artificial Intelligent and Machine Learning.

**Shuzhou Zhu** is a graduate student in the School of Management Science and Engineering of Nanjing University of Information Science and Technology, China. His current research fields include Signal Processing, Artificial Intelligent and Machine Learning.

**Wenjie Xu** is currently pursuing bachelor's degree in the School of Management Science and Engineering of Nanjing University of Information Science and Technology, China. His current research fields include Signal Processing, Artificial Intelligent and Machine Learning.