



A Stock Price Forecasting Model Integrating Complementary Ensemble Empirical Mode Decomposition and Independent Component Analysis

Youwei Chen^{1,2} · Pengwei Zhao¹ · Zhen Zhang³ · Juncheng Bai¹ · Yuqi Guo¹

Received: 24 June 2022 / Accepted: 1 September 2022
 © The Author(s) 2022

Abstract

In recent years, due to the non-stationary behavior of data samples, modeling and forecasting the stock price has been challenging for the business community and researchers. In order to address these mentioned issues, enhanced machine learning algorithms can be employed to establish stock forecasting algorithms. Accordingly, introducing the idea of “decomposition and ensemble” and the theory of “granular computing”, a hybrid model in this paper is established by incorporating the complementary ensemble empirical mode decomposition (CEEMD), sample entropy (SE), independent component analysis (ICA), particle swarm optimization (PSO), and long short-term memory (LSTM). First, aiming at reducing the complexity of the original data of stock price, the CEEMD approach decomposes the data into different intrinsic mode functions (IMFs). To alleviate the cumulative error of IMFs, SE is performed to restructure the IMFs. Second, the ICA technique separates IMFs, describing the internal foundation structure. Finally, the LSTM model is adopted for forecasting the stock price results, in which the LSTM hyperparameters are optimized by synchronously utilizing the PSO algorithm. The experimental results on four stock prices from China stock market reveal the accuracy and robustness of the established model from the aspect of statistical efficiency measures. In theory, a useful attempt is made by integrating the idea of “granular computing” with “decomposition and ensemble” to construct the forecasting model of non-stationary data. In practice, the research results will provide scientific reference for the business community and researchers.

Keywords Stock price forecasting · Complementary ensemble empirical mode decomposition · Independent component analysis · Long short-term memory

Abbreviations

ANNs	Artificial neural networks
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
CEEMD	Complementary ensemble empirical mode decomposition
CS	CEEMD and SE
CSA	Cuckoo search algorithm
DA	Directional accuracy
ENN	Elman neural network
ELM	Extreme learning machine
EMD	Empirical mode decomposition
GARCH	Generalized autoregressive conditional heteroskedasticity
GM	Grey model
GRNN	Generalized regression neural network
ICA	Independent component analysis
IMFs	Intrinsic mode functions
LSTM	Long short-term memory

✉ Youwei Chen
 chenyouwei@163.com
 Pengwei Zhao
 pwzhao@mail.xidian.edu.cn
 Zhen Zhang
 zhen.zhang@dlut.edu.cn
 Juncheng Bai
 JunchBai@163.com
 Yuqi Guo
 gyq_5706@163.com

- ¹ School of Economics and Management, Xidian University, Xi'an 710126, China
- ² School of Economics and Management, Xi'an University of Posts and Telecommunications, Xi'an 710061, China
- ³ School of Economics and Management, Dalian University of Technology, Dalian 116024, China

LSSVM	Least-squares support vector machine
MAE	Mean absolute error
MAPE	Mean absolute percentage error
PSO	Particle swarm optimization
RBFN	Radial basis function network
RMSE	Root mean square error
SE	Sample entropy
SSA	Singular spectral analysis
SVM	Support vector machine
VMD	Variational mode decomposition
WNN	Wavelet neural network
WOA	Whale optimization algorithm
WPT	Wavelet packet transform
WT	Wavelet transform

1 Introduction

With the continuous development of economic and financial markets, there are more and more factors affecting stock market transactions. In addition to the basic market factors, national tax politics, macro-economy, financial conditions and investors' irrational psychology and behavior factors have different and interrelated effects on stock prices. Financiers, researchers, and the government have considered stock index price forecasting, which is highly challenging because the stock market is mainly a dynamic, unstable, chaotic system [1, 2]. Therefore, to accurately forecast the stock price, this paper aims to enhance the model's predictive impact.

All kinds of stock price prediction models have appeared in recent decades. At present, the existing models of forecasting stock price are generally categorized into statistical approaches, artificial intelligence technologies, and hybrid models [3]. The popular statistical methods, such as autoregressive moving average (ARMA) [4], autoregressive integrated moving average (ARIMA) [5], and generalized autoregressive conditional heteroskedasticity (GARCH) [6], provide a better performance in a relatively stable stock market [5, 7]. For instance, Kristjanpoller and Michell [8] employed different GARCH models to predict the stock market fluctuations. The Markov Switching method was then conducted to specify the states of external effects. Finally, the experimental results indicated that the GARCH model with external factors could improve the forecasting accuracy. Pai and Lin [7] utilized the ARIMA model to capture the linear patterns. The data results revealed that the ARIMA model could utilize the unique strength of stock price trend forecasting.

Due to the nonlinear behavior and fast variations of stock prices, the statistical models' ability to forecast the stock price is limited [9–11]. With the rapid growth of artificial intelligence, several efficient artificial intelligence

approaches with strong robustness and fault tolerance have been presented for predicting the stock price in the literature, such as artificial neural networks (ANNs) (Elman neural network (ENN) [3], generalized regression neural network (GRNN) [12], radial basis function network (RBFN) [13] and wavelet neural network (WNN) [3], fuzzy logic methods [14], support vector machine (SVM) [15, 16], least-squares support vector machine (LSSVM) [17], and extreme learning machine (ELM) [18]. For example, Wang et al. [3] utilized the ENN, which can process dynamic data, to forecast four types of global stock indices. The experimental results indicated that ENN could improve prediction accuracy with fewer hidden neurons. Paiva et al. [15] established a unique decision-making model using the SVM and mean-variance approach for portfolio selection. Firstly, SVM was employed to forecast the stock exchange index. Next, the mean-variance method was conducted to allocate the investment funds. This fusion approach extended the theoretical usage of machine learning to propose a potentially practical method for stock price prediction.

Although various stock price predicting models have been established in the literature, the mentioned models are not generic. Notably, the statistical methods suffer from poor extrapolation ability, narrow prediction scale, and significant dependency on data. They can only be employed for linear data and are inappropriate for fluctuating data and noise [19, 20]. Although the grey model (GM) is appropriate for data with an exponential-type trend, it cannot describe the undulated series trend [21, 22]. The ANNs' initial random weights and thresholds can influence the forecasting accuracy, increasing the forecasting model's instability [19]. Scholars integrated traditional models to develop a hybrid forecasting model regarding the mentioned drawbacks, attaining the desired effect and becoming a typical mainstream [23]. For instance, Zolfaghari and Gholami [5] established a hybrid model by combining the financial factors and ARIMAX-GARCH family models for forecasting the stock price index. Compared with the benchmark models, the results reflected an enhancement in the hybrid model's stock index prediction. Kristjanpoller et al. [6] combined ANFIS and GARCH techniques to specify individual effects on any stock index and used an ANN framework to enhance the stock price's prediction performance. The results indicated that the mentioned approach could better estimate the stock index fluctuations.

The artificial intelligence method is prevalent in the stock field, but it is unstable, and parameter initialization can affect its results [24–26]. Furthermore, to solve the instability problem of the ANNs, various optimization algorithms, such as particle swarm optimization (PSO) [3, 27], firefly algorithm [28], whale optimization algorithm (WOA) [29, 30], and an enhanced version of the cuckoo search algorithm (CSA) [31], have been utilized to optimize the ANNs' initial

weight and threshold. Chandar [32] employed the NN optimized by genetic algorithms and PSO for forecasting the intraday stock price. The results demonstrated that the PSO-BPNN model could maximize forecasting precision while estimating the intraday stock price. Similarly, Das et al. [33] modified the ELM by the crow search algorithm to obtain the optimum weights to forecast the short-term stocks' closing price. High precision was obtained while validating the established model on several actual markets.

While obtaining accurate prediction results, it is also a challenge to find the stock data's trend characteristics and original signals. The hybrid prediction models are established to solve the mentioned problems. In order to distinguish and derive the primary characteristics of the stock price data, the hybrid model has been combined with various decomposition methods like the wavelet transform (WT) [5], wavelet packet transform (WPT) [34], singular spectral analysis (SSA) [35], variational mode decomposition (VMD) [36], empirical mode decomposition (EMD) [37], and complementary ensemble empirical mode decomposition (CEEMD) [3]. However, these decomposition methods produce too many IMFs, increasing the time cost and accumulating more errors after multiple stacking. Fusing IMFs with similar frequency to reduce the time cost and cumulative error is a valuable issue that can improve the decomposition methods.

Based on the mentioned analysis above, there are two disadvantages in the existing stock price prediction methods. On the one hand, the internal signal of the original data can not be revealed. On the other hand, the IMFs obtained by decomposition method have cumulative error. Therefore, aiming at solving these problems, this paper proposes a new prediction framework of stock price based on the theory of "granular computing" and the idea of "decomposition and integration". First, the collected unstable stock price data are processed by decomposition method to obtain the subsequences with different characteristics. Then the complexity measurement method is used to calculate the entropy of the subsequences. Meanwhile, for the sake of alleviating the cumulative error, the sequences with similar entropy are merged. Next, the essential characteristics of stock price data is revealed. Finally, the neural network improved by the optimization algorithm is utilized to forecast the merged subsequences, respectively. Finally, the forecasting results are integrated. Specifically, the stock price forecasting model based on ICA and CEEMD is established to reveal the primitive characteristics of stock price data and construct a more efficient hybrid model than the traditional ones. The established model comprises four units: a data preprocessing unit composed of CEEMD and SE (CS), a reconstitution and analysis unit, an optimization unit, and a prediction unit. The data preprocessing unit and reconstitution and analysis

unit present a decomposition technique to overcome the deficiencies of traditional decomposition techniques for attaining further enhancements. Meanwhile, to reveal the stock price data's primary features, the ICA is performed to separate the independent component from the IMFs. The Long short-term memory (LSTM) optimized by the PSO algorithm is then employed for forecasting the IMFs. In the end, final prediction values are achieved by integrating the forecasting results of all IMFs.

The current study is arranged as the following. Section 2 expresses the CEEMD, SE, ICA, and LSTM approaches. The established stock price prediction model is given in Sect. 3. Section 4 provides the experimental results. The conclusion and future research are presented in Sect. 5.

2 Preliminaries

The current section presents the comparative methods of the established model, including the CEEMD and ICA.

2.1 CEEMD

EMD decomposition has been extensively utilized to transform the complex initial series signal with nonlinear and nonstationary characteristics into a class of fluctuating modes described with IMFs, assuming that each signal can be a mix of IMFs series. There exist two conditions for any IMF: the extrema and zero-crossings should have similar values throughout the dataset, and the summation of the peak and valley values should be zero. The second condition indicates that the positions of upper and lower envelopes are symmetrically distributed along with a zero axis. Nevertheless, modal aliasing may be caused by the EMD approach, resolved through noise-based analysis. In terms of the average frequency distribution feature of Gaussian white noise added to the series, the signal has continuity on various scales, avoiding modal aliasing to a certain extent [38]. In order to prevent the mode mixing in EMD, the EEMD is further enhanced using a random Gaussian white noise incorporated with the original series data [39]. Yeh et al. [40] extended EMD to EEMD and CEEMD. Time series can be reestablished to eliminate the residual auxiliary noise and attain a more complete decomposed result through the CEEMD. For a specific original time series, the CEEMD algorithm is described as follows:

- (1) N pairs of Gaussian white noise containing positive and negative signs are added to the original signal, and the resultant $2N$ signal set can be described as:

$$\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} X \\ n \end{bmatrix} \quad (1)$$

where X and n stand for the original signal, and the Gaussian white noise, respectively. M_1 and M_2 stand for the signals obtained after adding or subtracting the original signal to or from the noise, respectively.

- (2) The EMD is applied to the target signals to achieve a class of IMF components for any time signal. IMF_{ij} stands for the i th component of the j th IMF.
- (3) The results of any IMF are generated after taking an average of the entire ensembles, as described in the following:

$$IMF_j = \frac{1}{2N} \sum_{i=1}^{2N} IMF_{ij} \quad (2)$$

Now, the final signal $Y(t)$ can be generated after applying CEEMD as:

$$Y(t) = \sum_{j=1}^k IMF_j(t) + res \quad (3)$$

where res is the residual term.

2.2 ICA

As a signal separating technique, the ICA approach has been presented by Comon [41]. The ICA aims to extract source signals s_i , ($i = 1, 2, \dots, n$) from the input signal $Y(t)$. If $Y(t)$ is a linear combination of s_i , ($i = 1, 2, \dots, n$), it can be represented with the following matrix form

$$Y = AS = \sum_{i=1}^n a_i s_i^T \quad (4)$$

where $Y = [y_1, y_2, \dots, y_n]^T$ stands for the input signal vector, $A = [a_1, a_2, \dots, a_n]$ is the original signal vector, which is a mixing matrix, and $A \in R^{n \times n}$, where a_i is the i th column of A . Generally, we do not know A . $S = [s_1, s_2, \dots, s_n]$ is the source signal vector, while s_i and s_j are mutually independent. The ICA aims to obtain the separation matrix M , where $M \in R^{n \times n}$ such that

$$X = MY = MAS \quad (5)$$

where X stands for the estimated value of S . Since ICA includes the uncertainties of variances and order, y_i is not necessarily an estimation of s_i . The following three assumptions can be considered: the first one is that the mixing matrix A is full rank, the second is that source signals are mutually independent, and the third is that

there is at most an individual source signal with a Gaussian distribution.

3 CEEMD and ICA-Based Stock Price Forecasting Model

The current section adopts a decomposition-hybridization algorithm based on the CEEMD, SE, ICA, PSO, and LSTM methods to forecast the stock price, abbreviated to CS-ICA-PSO-LSTM. The following subsection presents various performance measures to attain the forecasting accuracy of the established model.

It is challenging for traditional approaches to efficiently obtain the measured data's dynamic features due to the stock price's nonlinear and non-stationary features. Zhu [42] indicated that nonlinear and non-stationary problems could be solved using the data envelopment analysis approach. Accordingly, inspired by the research work of [42], we constructed CS-ICA-PSO-LSTM forecasting model of stock price. We call four-tuple (U, E, I, F) a hybrid forecasting system, where $U = \{x_1, x_2, \dots, x_n\}$ is a finite nonempty set, E is the value of sample entropy, $I = \{IMF_1, IMF_2, \dots, IMF_n, Res\}$ is the decomposition results, $F = \{f_1, f_2, \dots, f_m\}$ refers to the clustering results. The established model can be categorized into the following steps.

3.1 Decomposition and Reconstruction of Time Series of Stock Price Based on CEEMD and SE

Influenced by various information, the time series of stock price has strong nonstationary and irregularity. To solve this problem, the theory of "granular computing" and "decomposition and ensemble" is introduced in this paper. The mode decomposition approach is adopted for decomposing the stock price into IMFs. As a self-adaption decomposition approach, EMD [37] can decompose the measured data into various IMFs. Boundary impact and mode mixing are the main shortcomings of the decomposed IMFs. In order to resolve the mentioned difficulties, the stock price is decomposed into various IMFs using the CEEMD [40]. Based on CEEMD, the raw time series of stock price can be decomposed into different components. Because CEEMD decomposition method transforms complex time series into relatively stable and regular IMFs, the hidden information is easy to mine. The detailed process and definition are presented as following:

Definition 3.1 Let (U, E, I, F) be a hybrid forecasting system. Assume ε denote the decomposition error. The relation of U and I is defined as following:

$$x_i - IMF_{ij} \leq \varepsilon \quad \text{or} \quad X - IMF_{j=1}^m \leq \varepsilon (i = 1, 2, \dots, n) \quad (6)$$

However, too many IMFs lead to high time costs and cumulative errors. The current paper employs the SE to measure the stock price complexity. Then, the IMFs with the same complexity are fused. Sample entropy (SE) presented by Richman et al. [43] is utilized to verify the data series complexity. Compared with approximate entropy, the advantage of SE is that it much less dependent on the data length and has a good immunity to the noise interference. The SE has two benefits: On the one hand, the length of the input data does not affect its calculation results. On the other hand, the SE is not influenced by the loss of the input data. Even though two-thirds of the input data remain, it cannot significantly influence the computations. It can fully describe the signal entropy and includes three crucial parameters: the similarity tolerance r , adjusted to $0.2STD$, where STD describes the series standard deviation, the embedding dimension m , chosen as 2, and the size of the time series [44]. In order to implement the SE value, the following steps should be performed:

- (1) $\mathbf{Y} = (y(1), y(2), \dots, y(n))$ is reconstructed into the following matrix form:

$$\tilde{\mathbf{Y}} = \begin{bmatrix} y(1), y(2), \dots, y(n-m+1) \\ y(2), y(3), \dots, y(n-m+2) \\ \vdots \\ y(m), y(m+1), \dots, y(n) \end{bmatrix} \quad (7)$$

- (2) The distance between vectors $\mathbf{y}(i)$ and $\mathbf{y}(j)$, denoted by $d[\mathbf{y}(i), \mathbf{y}(j)]$, is described as:

$$d[\mathbf{y}(i), \mathbf{y}(j)] = \max(|\mathbf{y}(i+k) - \mathbf{y}(j+k)|) \quad (8)$$

$$(1 \leq k \leq m-1; 1 \leq i \neq j \leq n-m+1)$$

where k stands for the step length.

- (3) Considering a threshold r , if B_i is a number for a specific $\mathbf{y}(i)$ satisfying the condition $d[\mathbf{y}(i), \mathbf{y}(j)] \leq r$, the ratio $B_i^m(r)$ can be obtained as:

$$B_i^m(r) = \frac{1}{n-m+1} B_i \quad (9)$$

- (4) Calculate the mean $B^m(r)$ of $B_i^m(r)$ as:

$$B^m(r) = \frac{1}{n-m} \sum_{i=1}^{n-m} B_i^m(r) \quad (10)$$

- (5) Update m and iterate steps (1) to (3) to derive $B^{m+1}(r)$ as:

$$B^{m+1}(r) = \frac{1}{n-m} \sum_{i=1}^{n-m} B_i^{m+1}(r) \quad (11)$$

- (6) Obtain the SE value as:

$$E(m, r) = \lim_{n \rightarrow \infty} \left\{ -\ln \left[\frac{B^m(r)}{B^{m+1}(r)} \right] \right\} \quad (12)$$

The SE values determine the series' nonstationary degree; that is, the series complexity increases with the increase of the SE value.

3.2 Independent Component Extraction and Prediction

The revelation of source signals is the fundamental stage for establishing a stock price forecasting model. As a commonly used approach to determine the measured data's underlying parameters, ICA has been introduced into the revelation of source signals [41]. It has excellent efficiency in grasping the nonlinear and non-stationary signals' internal properties. As a signal separating technique, the ICA approach has been proposed by Comon [41]. As a powerful data analysis tool in recent years, the hidden factors or components from multivariate (multidimensional) statistical data can be found by the ICA, which is considered as an extension of principal component analysis and factor analysis. For the problem of blind source separation, ICA refers to an analysis process of separating or approximately separating the source signal without knowing the source signal, noise and mixing mechanism.

Definition 3.2 Given a hybrid forecasting system (U, E, I, F) , let $S = [s_1, s_2, \dots, s_n]$ be the source signal vector, M denote the separation matrix, A represent vector, and Y refer to the separating signal. For any $X \in U$, the following relation holds:

$$X = MY = MAS \quad (13)$$

Finally, the LSTM is modeled on the residual ICs to establish a stock price prediction model. Notably, the LSTM's prediction accuracy depends on determining the hyperparameters. In order to attain accurate forecasting results, the PSO [45] algorithm is adopted to find the optimum values of the LSTM hyperparameters. The PSO, first presented by Eberhart et al. [46] in 1995, is one of swarm intelligent optimization algorithms. This algorithm originated from the study of bird predation. The simplest and most effective way of finding edibles is to search for the nearest surrounding area where the birds have found food. Inspired by birds (particles) flocks looking for food, PSO employs the shared data between the particles to determine the global optimal solution. After initializing the particles, an iterative procedure is adopted to find the global optimal solution. In any iteration, the i th particle employs the particle tracking format at the optimal position of individual p_{best} and

swarms g_{best} to update its current position p_i and velocity v_i , respectively. This means that the v_i and p_i are updated. See [47, 48] to better understand the PSO algorithm and its applications.

Consider a D -dimensional search space. The i th particle denotes a D -dimensional vector $X_i = [x_{i1}, x_{i2}, \dots, x_{iD}]$. It also represents the location of the i th particle in the D -dimensional space or a potential solution to the problem. In every iteration k , the particle changes its speed V_i and location according to the personal and global extremum with the following updating rule:

$$V_i(k+1) = \omega V_i(k) + c_1 r_1 (P_i^{best}(k) - Y_i(k)) + c_2 r_2 (P_g^{best}(k) - Y_i(k)), \quad (14)$$

$$Y_i(k+1) = Y_i(k) + V_i(k), \quad (15)$$

where ω refers to the weight, c_1 and c_2 describe the acceleration factors, r_1 and r_2 stand for the random number distributed in the range $[0, 1]$.

LSTM, as a refined version of the RNN structure, was employed to resolve the gradient vanishing problems of RNN [49]. The LSTM employs a gating approach to control adding or removing the cells' state information. The cell state is equivalent to an information transmission path and is implemented immediately on the whole chain. The gate framework selectively allows the information to pass. Figure 1 describes the structure of an individual LSTM module.

In the LSTM structure, C_{t-1} and h_{t-1} represent the cell state and output of the upper layer. y_t refers to the new input. \tanh is the generated new cell value. Three σ symbols are the input, output, and forgetting gates, respectively. These three gates cooperate to control the whole cell state. The renewal process of each cell is as follows:

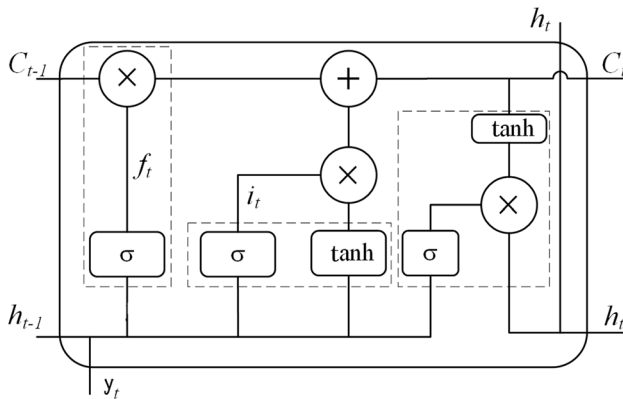


Fig. 1 The Structure of LSTM

- (1) Let W_f and b_f represent the weights and bias. The following formula determines all information to be abandoned in the forgetting gate:

$$f_t = \sigma(W_f[h_{t-1}, y_t] + b_f) \quad (16)$$

- (2) The following formula calculates all information to be updated in the input gate:

$$\begin{aligned} i_t &= \sigma(W_i[h_{t-1}, y_t] + b_i) \\ h_t &= \tanh(W_C[C_{t-1}, y_t] + b_C) \\ C_t &= f_t * C_{t-1} + i_t^* h_t \end{aligned} \quad (17)$$

- (3) The following formula updates the cell's output gate:

$$\begin{aligned} S_t &= \sigma(W_S[h_{t-1}, y_t] + b_S) \\ h_t &= S_t * \tanh(C_t) \end{aligned} \quad (18)$$

Based on the above discussion, the current study starts with the mode decomposition method improved by the sample entropy, followed to reveal the source signals of independent components (ICs). In the end, the neural network optimized by heuristic algorithm is adopted to produce the prediction results. The detailed flow chart is illustrated in Fig. 2, which depicts the framework of the established model. Considering various goals, three components can be categorized. From the decomposition unit aspect, the original information X can be decomposed into a range of IMFs through the CEEMD model (the subscript n stands for the number of IMFs rather than the sifting times in the CEEMD algorithm, while the residual function Res is considered a new IMF for convenience). Taking into consideration that too many IMFs will accumulate errors, the SE is used to reconstruct IMFs. Next, the ICA is conducted to reveal the independent components of IMFs. Finally, the LSTM optimized by PSO is utilized to forecast the all independent components. These parameters of all algorithms are listed in Table 1.

3.3 Evaluation Criteria

In order to assess the accuracy of prediction results, five measures are chosen as evaluation criteria: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), directional accuracy (DA) prediction statistic, and coefficient of determination (R^2) [50–52]. These criteria are defined as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\% \quad (19)$$

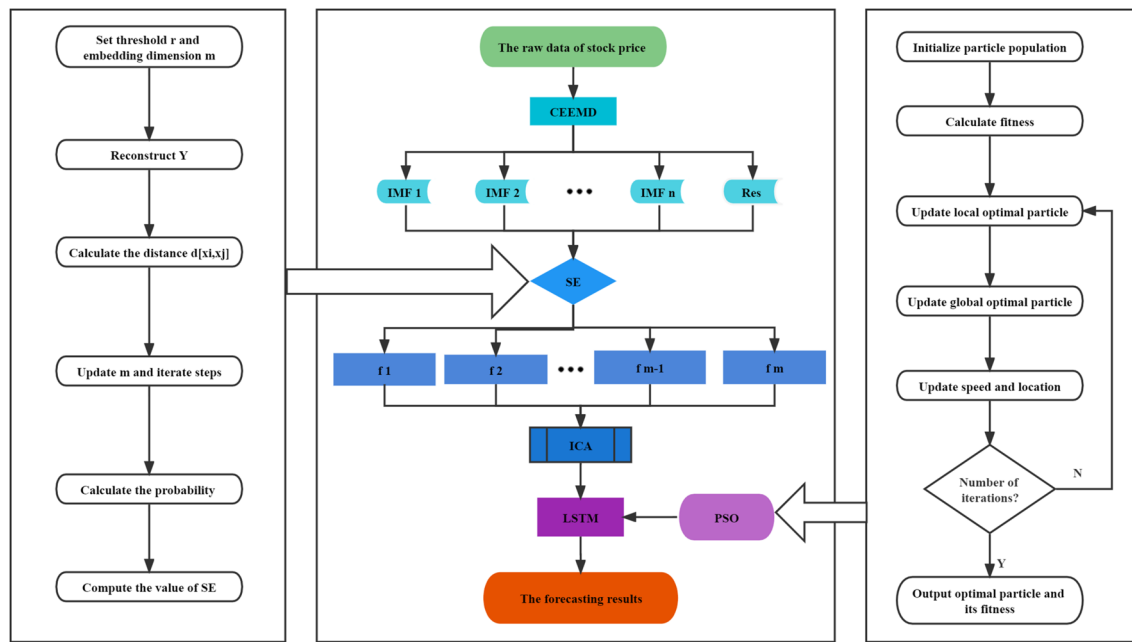


Fig. 2 Framework of the Proposed Forecasting Model

Table 1 The parameter setting of these methods

Algorithm	Parameter	Value	Algorithm	Parameter	Value
CEEMD	ensemble member	200	LSTM	Hidden layer	120
	SD	0.4		Batch size	100
SE	Embedding dimension	2		Activation function	<i>tanh</i>
	Similarity tolerance	0.2SD		Dropout	0.2
PSO	Number of iterations	100		Epochs	100
	Population size	50		Learning rate	0.001

Note: The SD means standard deviation

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (20)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (21)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (\bar{Y}_i - Y_i)^2} \quad (22)$$

The former four measures can be utilized to verify the forecasting accuracy, and the DA describes another essential measure to predict the movement orientation [53, 54]. *MAPE* is the ratio of the absolute value of the error to the actual value. The value of *MAPE* is $[0, \infty]$. 0 denotes a perfect prediction model, while greater than 100% indicates a poor prediction result. *MAE* and *RMSE* are performed to assess

the error between the predicted value and the real value. Its range is $[0, +\infty)$. The greater the error, the greater the value of *MAE*. When the predicted value is completely consistent with the real value, it is equal to 0. In other words, the forecasting model is perfect. The numerator of R^2 represents the sum of the square difference between the real value and the predicted value, similar to the *RMSE*. Its denominator represents the sum of the square difference between the real value and the mean.

$$DA = \frac{1}{n} - \sum_{i=1}^n d_i \quad (23)$$

where $d_i = \begin{cases} 1, & (Y_i - Y_{i-1})(\hat{Y}_i - \hat{Y}_{i-1}) \geq 0 \\ 0, & \text{otherwise} \end{cases}$; n stands for the whole number of data points; Y_i and \hat{Y}_i describe the actual and predicted values of the i th term, respectively.

MAPE, MAE, and RMSE are employed to determine the difference between the predicted and actual values. The lower values of the mentioned measures indicate a superior prediction accuracy. Nevertheless, better prediction accuracy can be obtained for higher values of the DA and R^2 . In addition, Shannon entropy is utilized to measure the complexity of neural networks, which is expressed as the following:

$$S_e = -k \sum_{i=1}^N p_i = \frac{D(i)}{\sum_{i=1}^N D(i)} \log p_i = \frac{D(i)}{\sum_{i=1}^N D(i)}. \quad (24)$$

where N denotes the number of nodes, and $D(i)$ represents the degree of i th node, $k = 1.38 \times 10^{-23}$ denotes the Boltzmann constant.

4 Experiments and Discussions

This section is organized as follows. Data sources are introduced in subsection 4.1. The decomposition, reconstruction, and ICA results are presented in subsections 4.2 and 4.3. Finally, the prediction accuracy of different models is compared in subsection 4.4.

4.1 Data

The current work aims to predict the stock price. Based on the turnover and total market value, the Shanghai Stock Exchange (SSE) 50 index is chosen, with the features of a large scale, relatively stable, and sufficient liquidity [55]. Thus, the current work randomly chooses four stocks in the SSE 50 index as candidate assets. The ticker symbols of four stocks are “SH600518”, “SH600519”, “SH600999”, and “SH601988”. The statistical description about the 4 stocks is presented as following:

Data from March 19, 2001 to March 16, 2021 are collected and divided into a training set and a testing set. The training set ranges from March 19, 2001 to April 17, 2017 and the testing set is from April 18, 2017 to March 16, 2021. The ratio of training set and testing set is 8:2.

4.2 The CEEMD Decomposition and Reconstruction Results

At first, CEEMD is employed for decomposing the stock price’s raw series into several IMFs. The CEEMD decomposition results are presented in Figs. 3, 4, 5, 6 and 7. The SE values for all IMFs obtained by CEEMD decomposition are presented in Table 1. In order to

alleviate the calculation cost, improve the modeling speed, and prevent overfitting problems, a combination of IMFs with the same sample entropy values is employed. For “SH600518”, the sample entropy values of the first three IMFs (IMF1, IMF2, and IMF3) are more than the others, indicating a complicated, volatile, but inconspicuous pattern. In contrast, the last three IMFs and the residue (IMF10, IMF11, IMF12, and Res) are much lower with an evident trend with slight complexity and volatility. Thus, the mentioned components can be combined to form the following new Intrinsic Mode Functions (NIMFs): high-frequency sequence NIMF1 (IMF1-3), medium-frequency sequence NIMF2 (IMF4-6), low-frequency sequence NIMF3 (IMF7-9), and trend sequence NIMF4 (IMF10-12, Residual). The variation trend of the mentioned four NIMFs is relatively normal and simple, suitable for further deriving the variation characteristics of any IMF and learning the predictive model. The same reconstruction model is chosen for the other three stocks. The detailed reconstruction results are presented in Fig. 5.

4.3 The ICA Results

The separation unit is another step of the established model. We start with separating ICs. The ICA model is applied to the IMFs to attain ICs. Moreover, the ICs are found by revealing the source signal. Finally, the ICs with source signals are considered as the input of the LSTM for the construction of the forecasting model.

The four ICs are separated through the ICA from the recombination modes. Figs. 8, 9, 10, 11 describe the final results obtained with the enhanced CS-ICA method. Based on Eq. (13), four separation matrixes are derived.

$$M_{SH600518} = \begin{pmatrix} 0.2845 & 0.2116 & 0.4625 & -0.8127 \\ -0.1737 & 0.5012 & 0.7049 & 0.4708 \\ 0.1273 & 0.8347 & -0.5341 & -0.0420 \\ 0.9342 & -0.0849 & 0.0630 & 0.3408 \end{pmatrix} \quad (25)$$

$$M_{SH600519} = \begin{pmatrix} 0.5367 & 0.3560 & -0.4494 & 0.6191 \\ -0.6338 & -0.0869 & 0.1976 & 0.7428 \\ 0.5566 & -0.4717 & 0.6363 & 0.2506 \\ 0.0206 & 0.8020 & 0.5951 & -0.0470 \end{pmatrix} \quad (26)$$

$$M_{SH600999} = \begin{pmatrix} -0.3628 & -0.0012 & -0.2276 & 0.9036 \\ 0.9289 & 0.0146 & -0.0119 & 0.3700 \\ -0.0328 & 0.9648 & 0.2555 & 0.0525 \\ -0.0672 & -0.2625 & 0.9396 & 0.2093 \end{pmatrix} \quad (27)$$

$$M_{SH601988} = \begin{pmatrix} 0.9990 & -0.0078 & -0.0004 & 0.0438 \\ -0.0044 & 0.6854 & 0.6910 & 0.2294 \\ 0.0149 & 0.7281 & -0.6518 & -0.2142 \\ 0.0417 & 0.0010 & 0.3141 & -0.9485 \end{pmatrix} \quad (28)$$

$$\hat{x}(t)_{SH600519} = \sum_{i=1}^4 a_i s_i(t) = 1.0624s_1 + 0.2197s_2 + 0.9718s_3 + 1.3707s_4 \quad (30)$$

The following formula can evaluate the stock price series. The mixing matrix A is the inverse of M , where a_i stands for the sum of the i th column.

$$\hat{x}(t)_{SH600999} = \sum_{i=1}^4 a_i s_i(t) = 0.3120s_1 + 1.3016s_2 + 1.2400s_3 + 0.8192s_4 \quad (31)$$

$$\hat{x}(t)_{SH600518} = \sum_{i=1}^4 a_i s_i(t) = 0.1459s_1 + 1.5032s_2 + 0.3859s_3 + 1.2531s_4 \quad (29)$$

$$\hat{x}(t)_{SH601988} = \sum_{i=1}^4 a_i s_i(t) = 1.0346s_1 + 1.6014s_2 - 0.1222s_3 - 0.5917s_4 \quad (32)$$

Table 2 The statistical properties of stocks

Stock	Max	Min	Mean	SD	Kurtosis	Skewness
SH600518	50.09	22.31	39.58	28.95	2.06	− 0.32
SH600519	1357	27.37	782.10	112.95	1.05	0.07
SH600999	39.45	31.25	35.63	34.12	2.06	− 0.53
SH601988	6.98	3.92	4.72	4.84	1.62	0.04

Note: The SD means standard deviation

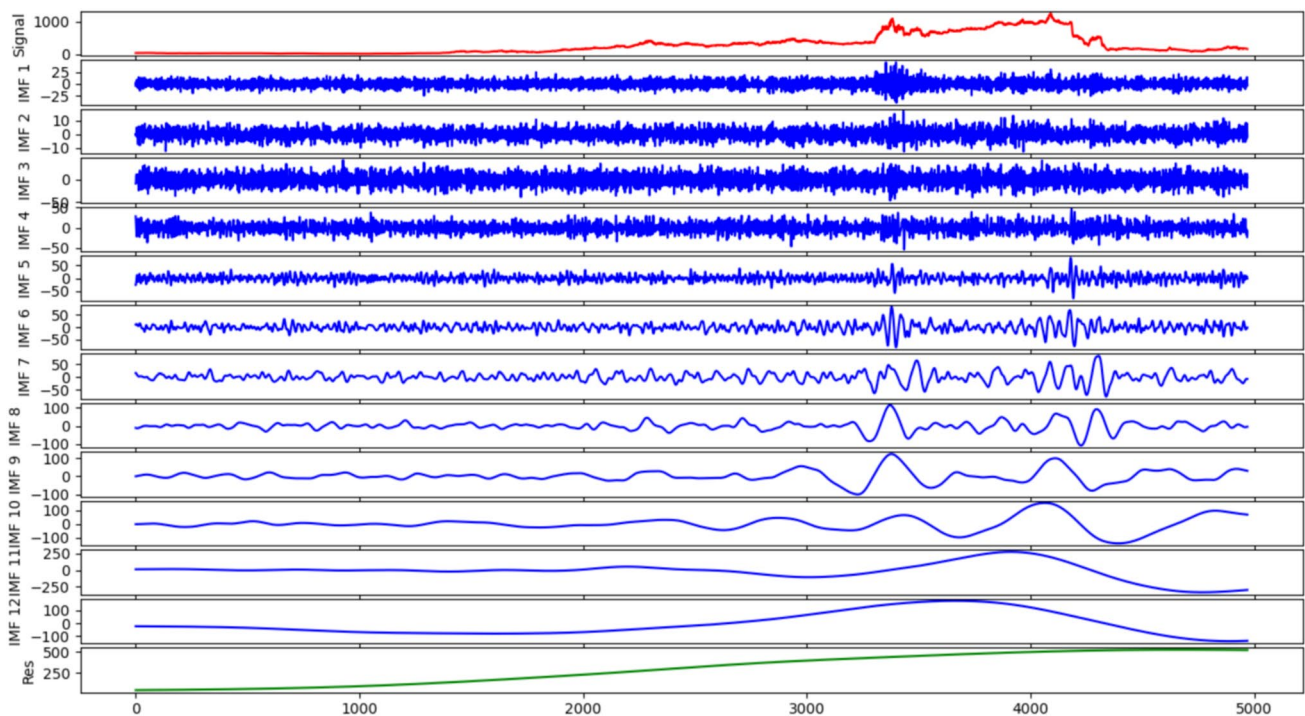


Fig. 3 The decomposition results of CEEMD for SH600518

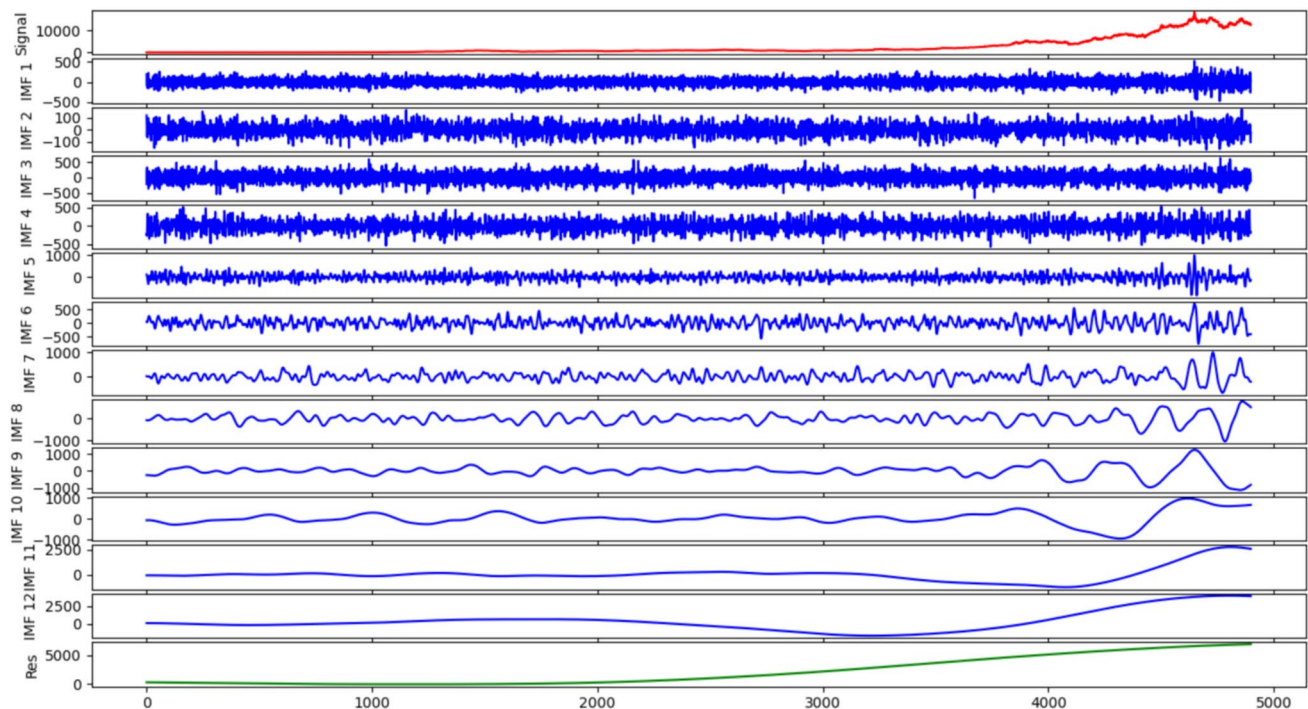


Fig. 4 The decomposition results of CEEMD for SH600519

4.4 Comparing the Prediction Accuracy

In order to verify the accuracy of the established model, the current work employs some benchmark models for comparison, including ARIMA, BP, SVM, LSTM, CNN, PSO-LSTM, CS-PSO-LSTM, and CS-ICA-PSOLSTM (CS refers to the CEEMD optimized by SE). The prediction results of the established model are compared with the other benchmark models in Tables 4, 5, 6 and 7. The following results can be obtained:

- (1) The LSTM is superior to ARIMA and BP, as described in Table 2. Compared with ARIMA and BP for SH600518, the considerable decrease in the MAE index in the 1-step forecasting is 89.57% and 49.21%. As shown in Table 3, for MAPE in the SH600519, the decline is 51.25% and 24.98% in the 2-step prediction, respectively. As presented in Table 4, the MAE, RMSE, and MAPE in the LSTM model are significantly less

than in the ARIMA model in the 3-step prediction. Compared with ARIMA, the LSTM model decreases 72.84%, 71.35%, and 72.92% in MAE, RMSE, and MAPE for 3-step prediction, respectively. The results reveal that deep learning can efficiently mine the long-term dependence of stock series. From the experiment involving different single models, we can get that LSTM play a positive role in forecasting performance of stock price.

- (2) The PSO-LSTM model nonlinear benchmark is constructed to evaluate the PSO model's efficiency for stock price forecasting. PSO-LSTM gives higher prediction performance. As shown in Table 4, compared with LSTM at 2-step forecasting for the SH600999, the MAE, RMSE, and MAPE values of PSO-LSTM for SH600999 are 45.07%, 35.91, and 43.15%, respectively. Hence, it has been verified that the PSO model is more suitable for stock forecasting. Compared with the ARIMA model, it is evident that

Table 3 The values of sample entropy of IMFs

Stock	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10	IMF11	IMF12	Res
SH600518	1.0684	1.2164	1.2370	0.7954	0.4568	0.2497	0.0937	0.0308	0.0147	0.0073	0.0019	0.0017	0.0002
SH600519	1.1187	1.2235	1.2542	0.7771	0.4692	0.2927	0.1245	0.0496	0.0159	0.0068	0.0013	0.0015	0.0005
SH600999	0.9266	1.1397	1.2183	0.6538	0.3948	0.1674	0.0575	0.0301	0.0151	0.0094	0.0042	–	0.0006
SH601988	0.9114	1.1415	1.2435	0.6939	0.4284	0.2061	0.0828	0.0397	0.0089	0.0098	0.0029	–	0.0009

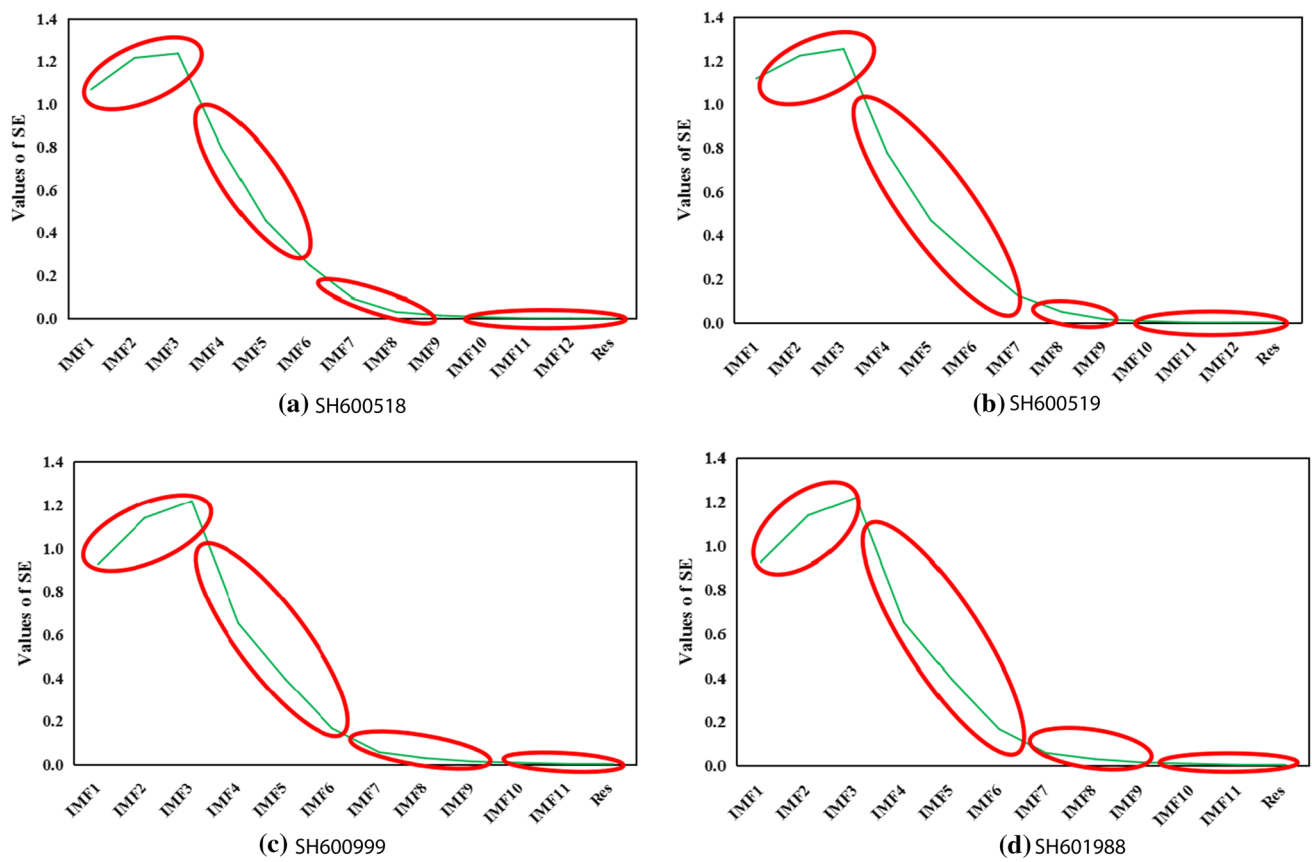


Fig. 5 The results of IMF reconstruction

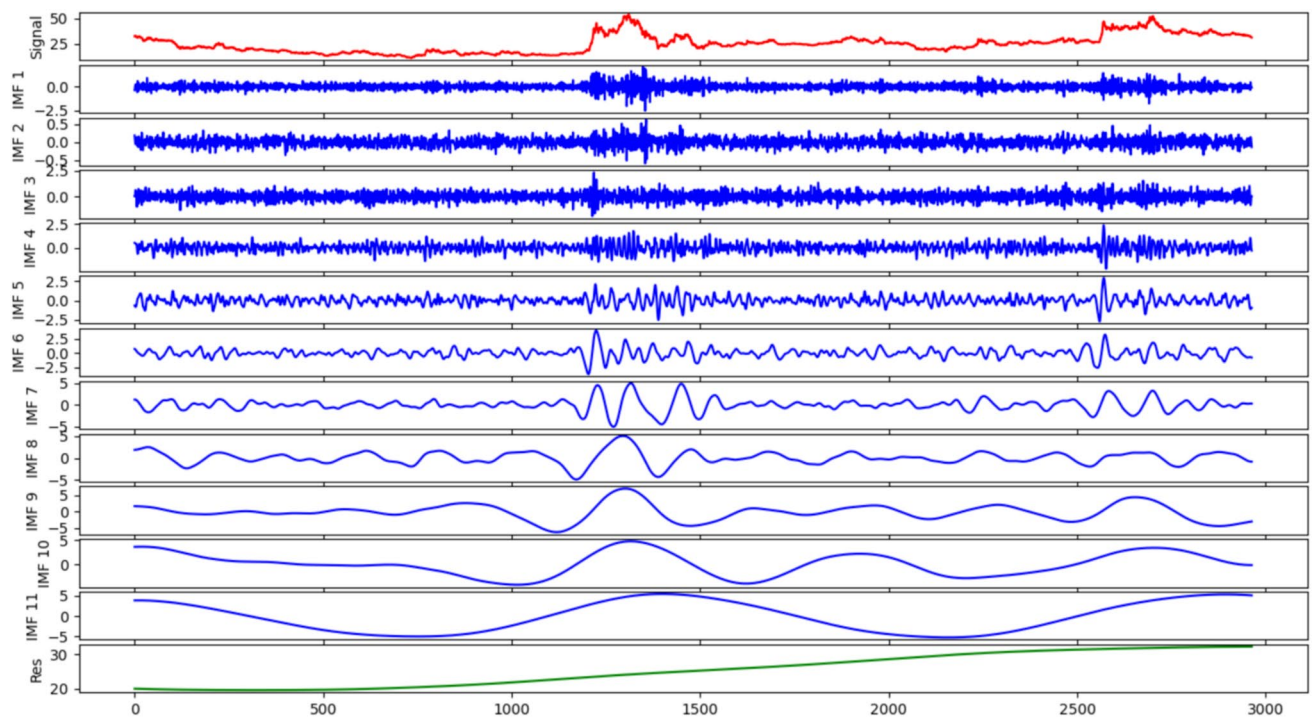


Fig. 6 The decomposition results of CEEMD for SH600999

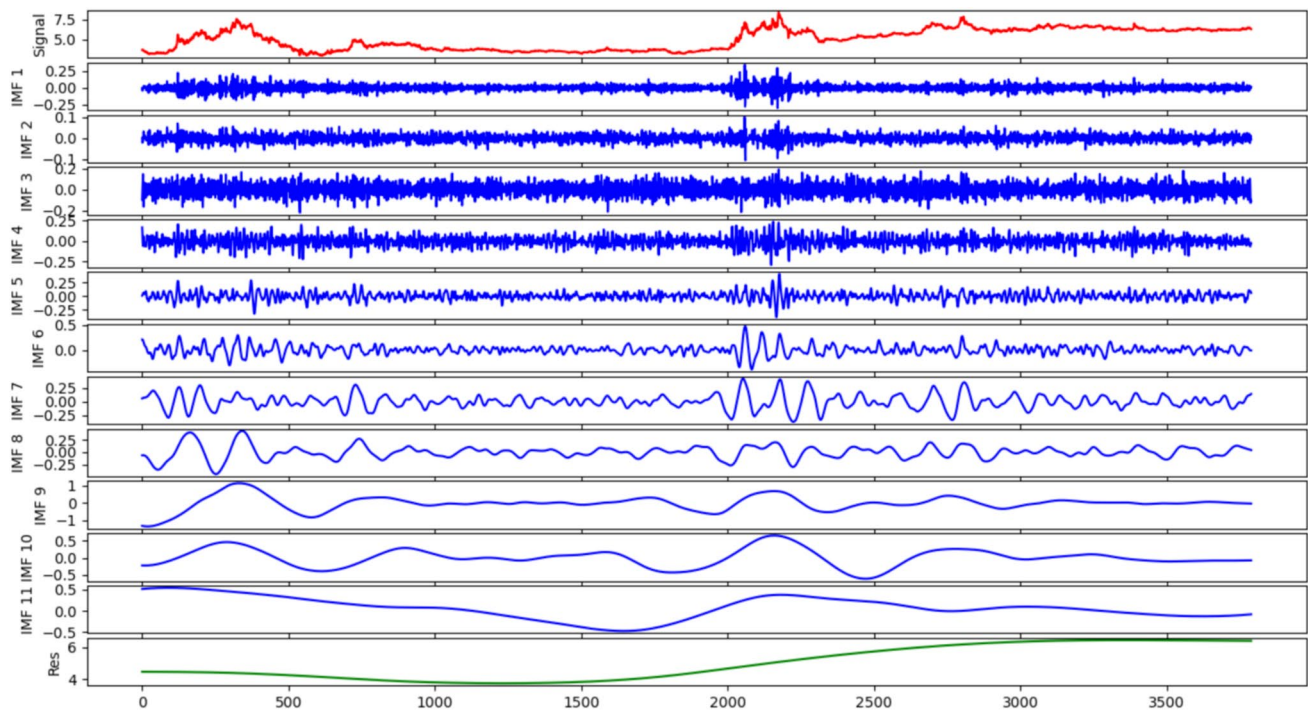


Fig. 7 The decomposition results of CEEMD for SH601988

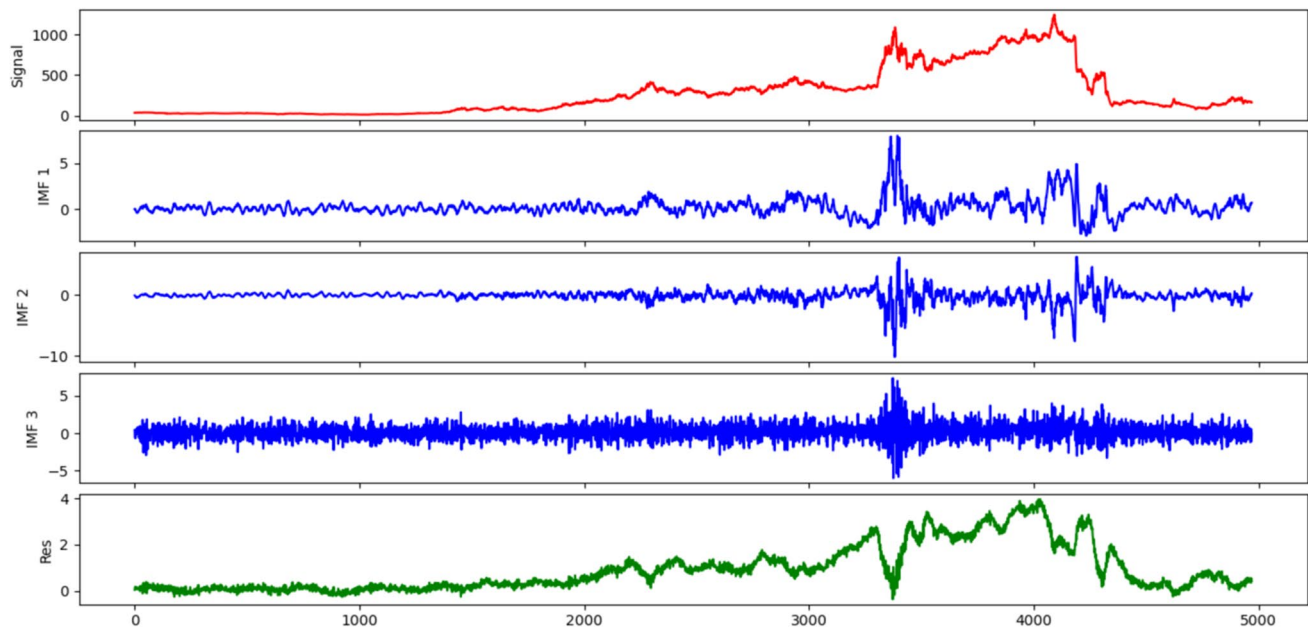


Fig. 8 The results of ICA for SH600518

the LSTM model provides better efficiency in 1-step prediction. For instance, Table 5 shows that the MAE, RMSE, MAPE, and DA values of the PSO-LSTM are decreased by 18.33%, 70.91%, 74.14%, and 0.09% than ARIMA in 1-step prediction, respectively. According to

these mentioned analysis, the PSO approach providing optimal parameters for LSTM achieves a better forecasting performance than the LSTM model.

(3) Compared with PSO-LSTM, CS-PSO-LSTM provides higher prediction accuracy. For example,

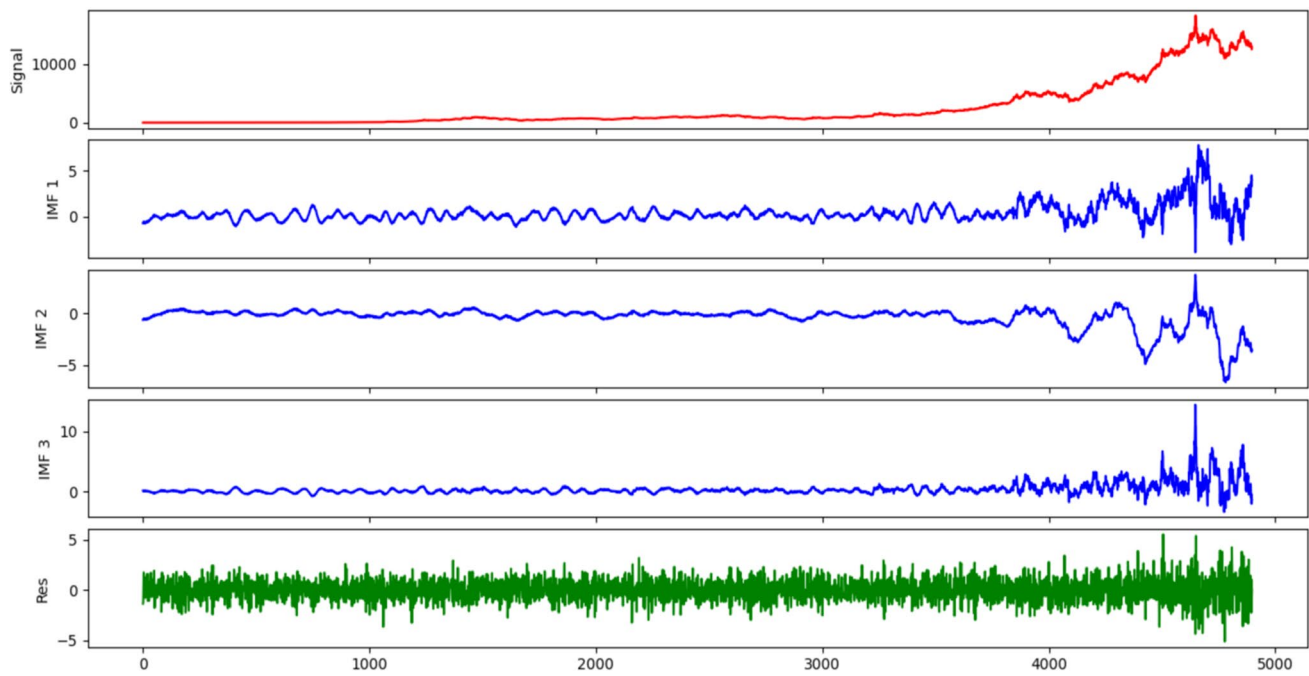


Fig. 9 The results of ICA for SH600519

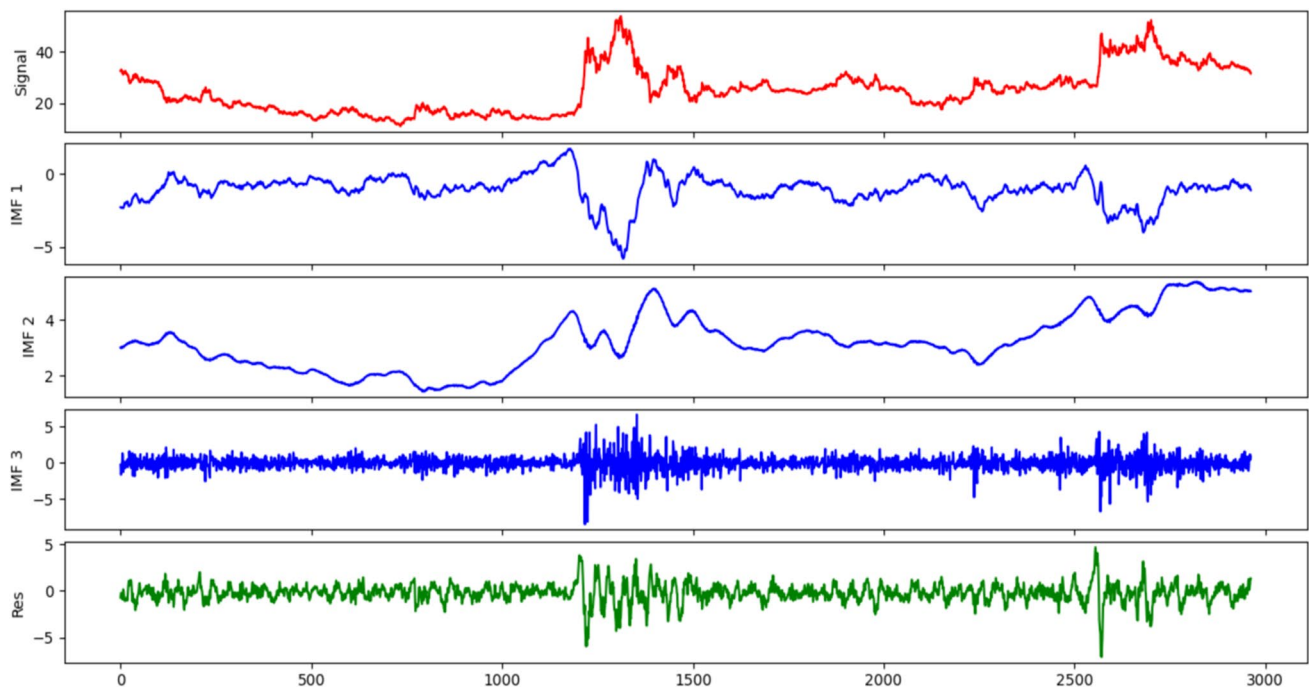


Fig. 10 The results of ICA for SH600999

the MAE, RMSE, and MAPE of CS-PSO-LSTM at 1-step forecasting are increased by 41.67%, 31.25%, and 36.11% for SH601988, respectively. The results indicate that the CEEMD algorithm optimized by SE is an improved decomposition algorithm. Compared

with the ARIMA, BP, and LSTM, the increase in MAPE of the CS-PSO-LSTM model for SH601988 in the 3-step forecasting is 80.60%, 78.37%, and 50.18%, respectively, demonstrating the positive influence of the CS on the ENN model. As a result, the CS-PSO-

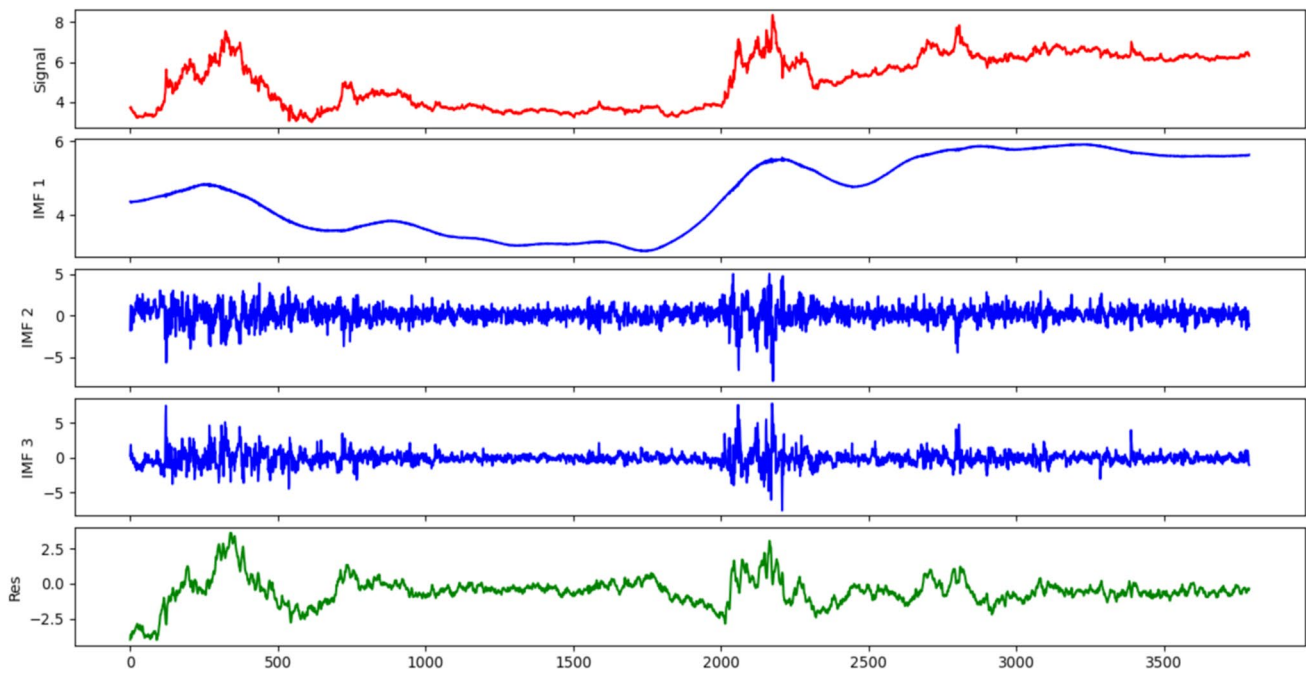


Fig. 11 The results of ICA for SH601988

LSTM approach has made great progress compared with the PSO-LSTM model in the field of stock price forecasting.

- (4) The ICA method can improve forecasting accuracy. As presented in Table 3, the MAE, RMSE, and MAPE of CS-ICA-PSO-LSTM in 1-step forecasting for SH600519 are 68.81%, 73.51%, and 54.03%. This performance is better than the other benchmark models. As presented in Table 3, the DA indices of CS-ICA-PSO-LSTM are inferior to ARIMA, BP, LSTM, and PSO-LSTM in terms of the 1-step forecasting performance. According to the above analyses, the performance of the CS-ICA-PSO-LSTM approach is excellent and can be accepted on economic feasibility. It is obvious to see that the forecasting results of the proposed approach with the ICA method can capture the essential characteristics of stock price data and improve the forecasting performance.
- (5) Due to the significant effect of the forecasting horizon on the developed model's stability, the prediction is divided into 1-step prediction, 2-step prediction, and 3-step prediction in terms of the forecasting horizon. When the forecasting horizon is 1-STEP for SH601988, MAPE values are 0.71%, 0.86%, and 0.99%. The forecasting accuracy degrades with the increase of the forecasting horizon.

5 Conclusion and Future Work

As a crucial research area, stock forecasting has attracted great concern owing to its potential financial advantages. Precise forecasting of stock price fluctuations significantly implies the analysis of fluctuation and source signal and establishing a suitable forecasting model. To overcome the weak performance of the existing methods, the current paper presents a hybrid model to forecast stock prices through the CEEMD, SE, ICA, PSO, and LSTM. The current study includes the following tasks. At first, the stock price is adaptively decomposed into several sequences. Further, the SE is employed to reconstruct the decomposed sequence according to complexity. The ICA model is then employed to separate the ICs describing the original data's internal formation mechanism. Finally, the IC components are chosen as the input data in the LSTM approach, while the PSO algorithm is employed for finetuning the LSTM model's hyperparameters.

The proposed ES-ICA-PSO-LSTM model comprises various models, including the CEEMD, SE, ICA, PSO, and LSTM, to forecast the stock price. The outstanding performance is attributed to the following reasons: (1) Based on the theory of "granular computing" and "decomposition and ensemble", the raw data of stock price are decomposed into different components. On the one hand, the hidden information is revealed. On the other hand, different features (trend, period, random, etc.) are classified. (2) SE is conducted to restructure the IMFs

Table 4 Forecasting performance of different models for SH600518

Models	1-Step forecasting						2-Step forecasting						3-Step forecasting					
	MAE	RMSE	MAPE	DA	R ²	E _s	MAE	RMSE	MAPE	DA	R ²	E _s	MAE	RMSE	MAPE	DA	R ²	E _s
ARIMA	530.70	564.09	324.85	0.44	-1.72	—	531.10	564.32	324.98	0.44	−1.73	—	570.07	632.77	367.18	0.44	−2.43	—
BP	108.93	157.77	51.25	0.44	0.79	1.67	135.98	172.91	66.13	0.49	0.74	1.67	531.49	564.55	325.10	0.44	−1.73	1.67
SVM	68.73	80.29	40.37	0.44	0.82	3.25	80.79	120.36	51.38	0.49	0.82	3.25	267.78	300.93	173.42	0.44	0.92	3.25
LSTM	55.33	77.43	22.28	0.43	0.95	3.25	66.25	82.69	31.64	0.43	0.94	3.25	132.51	147.04	70.72	0.43	0.81	3.25
CNN	43.22	62.57	15.78	0.44	0.95	3.25	45.71	70.35	20.36	0.43	0.95	3.25	94.57	100.62	53.26	0.43	0.82	3.25
PSO-LSTM	30.25	55.63	10.68	0.44	0.97	3.25	30.58	55.22	10.42	0.43	0.97	3.25	31.67	55.50	10.32	0.42	0.97	3.25
CS-PSO-LSTM	15.27	24.51	5.75	0.48	0.99	5.27	15.55	21.69	6.75	0.51	1.00	5.27	16.47	28.66	5.95	0.48	0.99	5.27
CS-ICA-PSO-LSTM	10.78	19.03	3.59	0.43	1.00	5.27	14.71	26.70	5.18	0.43	0.99	5.27	15.13	28.13	4.94	0.44	0.99	5.27

Bold values indicate the optimal results

Table 5 Forecasting performance of different models for SH600519

Models	1-Step forecasting					2-Step forecasting					3-Step forecasting							
	MAE	RMSE	MAPE	DA	R ²	E _s	MAE	RMSE	MAPE	DA	R ²	E _s	MAE	RMSE	MAPE	DA	R ²	E _s
ARIMA	6390.75	7356.94	64.19	0.49	-2.72	-	6391.37	7357.50	64.20	0.49	-2.72	-	6392.03	7358.06	64.21	0.49	-2.72	-
BP	3451.25	4489.57	30.48	0.51	-0.39	1.67	4665.02	5903.10	41.72	0.50	-1.40	1.67	4816.06	6096.75	43.04	0.50	-1.56	1.67
SVM	2938.27	3545.72	22.46	0.51	-0.39	3.25	3942.02	5203.54	36.78	0.49	-0.93	3.25	4530.52	5789.64	39.81	0.50	-1.26	3.25
LSTM	2437.52	3336.71	20.66	0.50	0.23	3.25	3601.77	4788.01	31.30	0.49	-0.58	3.25	3988.97	5231.19	34.97	0.49	-0.88	3.25
CNN	2361.47	3256.83	20.01	0.50	0.25	3.25	3052.67	4010.24	25.89	0.50	0.02	3.25	3533.29	4864.19	30.62	0.49	0.17	3.25
PSO-LSTM	2298.36	3159.53	19.39	0.49	0.31	3.25	2480.53	3393.81	21.02	0.50	0.21	3.25	2424.01	3312.90	20.57	0.49	0.25	3.25
CS-PSO-LSTM	1874.22	2656.16	15.51	0.50	0.51	5.27	2428.84	3336.75	20.50	0.49	0.23	5.27	1964.36	2801.73	16.04	0.50	0.46	5.27
CS-ICA-PSO-LSTM	584.62	703.72	7.13	0.52	0.97	5.27	2124.83	3211.70	16.70	0.50	0.29	5.27	1953.55	2698.72	16.47	0.51	0.50	5.27

Bold values indicate the optimal results

Table 6 Forecasting performance of different models for SH600999

Models	1-Step forecasting					2-Step forecasting					3-Step forecasting							
	MAE	RMSE	MAPE	DA	R^2	E_s	MAE	RMSE	MAPE	DA	R^2	E_s	MAE	RMSE	MAPE	DA	R^2	E_s
ARIMA	11.94	13.45	32.06	0.50	-3.73	-	12.15	13.63	32.67	0.50	-3.86	-	12.37	13.82	33.31	0.50	-3.99	-
BP	7.44	9.39	19.06	0.53	-1.30	1.67	7.44	9.39	19.07	0.52	-1.30	1.67	7.44	9.40	19.09	0.52	-1.31	1.67
SVM	3.41	6.21	10.36	0.53	-0.72	3.25	4.52	6.92	11.29	0.52	-0.67	3.25	5.26	7.42	13.45	0.52	-0.29	3.25
LSTM	2.70	3.73	6.86	0.54	0.64	3.25	2.84	3.62	7.44	0.52	0.66	3.25	3.36	3.96	9.02	0.51	0.59	3.25
CNN	2.05	2.90	5.16	0.53	0.71	3.25	2.15	3.02	6.02	0.52	0.80	3.25	3.02	3.57	7.09	0.52	0.80	3.25
PSO-LSTM	1.51	2.20	4.09	0.52	0.87	3.25	1.56	2.32	4.23	0.53	0.86	3.25	1.59	2.40	4.31	0.52	0.85	3.25
CS-PSO-LSTM	0.89	1.46	2.42	0.55	0.94	5.27	1.13	1.69	3.05	0.52	0.93	5.27	1.32	1.95	3.58	0.52	0.90	5.27
CS-ICA-PSO-LSTM	0.55	0.86	1.51	0.48	0.98	5.27	0.80	1.23	2.19	0.53	0.96	5.27	0.81	1.25	2.20	0.52	0.96	5.27

Bold values indicate the optimal results

Table 7 Forecasting performance of different models for SH601988

Models	1-Step forecasting					2-Step forecasting					3-Step forecasting							
	MAE	RMSE	MAPE	DA	R ²	E _s	MAE	RMSE	MAPE	DA	R ²	E _s	MAE	RMSE	MAPE	DA	R ²	E _s
ARIMA	0.45	0.55	6.96	0.40	-5.89	-	0.45	0.55	6.96	0.40	-5.90	-	0.45	0.55	6.96	0.40	-5.91	-
BP	0.29	0.30	4.47	0.40	-1.08	1.67	0.30	0.34	4.77	0.41	-1.65	1.67	0.39	0.44	6.24	0.40	-3.31	1.67
SVM	0.17	0.21	3.52	0.40	-0.62	3.25	0.19	0.27	3.81	-0.73	3.25	0.21	0.31	4.02	0.40	-0.86	3.25	
LSTM	0.13	0.16	2.09	0.40	0.44	3.25	0.15	0.20	2.35	0.42	0.12	3.25	0.17	0.19	2.71	0.40	0.18	3.25
CNN	0.13	0.16	1.93	0.40	0.44	3.25	0.15	0.18	2.04	0.45	3.25	0.17	0.21	2.58	0.40	0.38	3.25	
PSO-LSTM	0.12	0.16	1.80	0.44	0.41	3.25	0.14	0.16	2.19	0.40	0.45	3.25	0.16	0.21	2.44	0.44	0.02	3.25
CS-PSO-LSTM	0.07	0.11	1.15	0.44	0.75	5.27	0.08	0.11	1.21	0.42	0.73	5.27	0.09	0.12	1.35	0.41	0.70	5.27
CS-ICA-PSO-LSTM	0.05	0.07	0.71	0.44	0.90	5.27	0.06	0.08	0.86	0.44	0.85	5.27	0.06	0.09	0.99	0.44	0.83	5.27

Bold values indicate the optimal results

and alleviate the cumulative error. (3) The ICA technique can describe the internal foundation structure of IMFs, which is key to reveal the essence of the original signal. In theory, a useful attempt is made by integrating the idea of “granular computing” with “decomposition and ensemble” to construct the forecasting model of non-stationary data. In practice, the research results will provide scientific reference for the business community and researchers.

The stock prices affected by emergencies is difficult to accurately forecast. Future aspects are given as follows: (1) establishing an intelligent forecasting model without considering the number of models employed in the prediction approach, (2) presenting an optimum hybrid model, (3) extending the presented model to another time series like wind speed forecasting and gold price forecasting.

Acknowledgements The authors would like to thank the editors and funding for handling and supporting our paper.

Author Contributions YC completed the main work of this paper under the guidance of PZ, containing methodology and writing original draft. ZZ performed the validation and formal analysis. JB and YG implemented the editing work. All authors read and approved the final manuscript.

Funding The work was partly supported by the National Natural Science Foundation of China (No. 72071152), the Youth Innovation Team of Shaanxi Universities (No. 2019), the General Project of Statistical Science Research of National Bureau of Statistics (No. 2021LY078), the Soft Science Research Program of Shaanxi Provincial Department of Science and Technology (No. 2022KRM114), Key projects of Scientific Research Program of Shaanxi Provincial Department of Education (New think Tank of Colleges and Universities) (No. 21JT035).

Availability of data and material Not applicable.

Declarations

Conflict of Interest All authors declare that they have no conflicts of interest.

Ethical Approval Yes.

Consent to Participate Yes.

Consent for Publication Yes.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted

use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ghosh, P., Neufeld, A., Sahoo, J.K.: Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Res. Lett.* **102**, 1544–6123 (2021)
2. Chen, C., Zhou, Y., Gospodinov, N., Maynard, A., Pesavento, E.: Long-horizon stock valuation and return forecasts based on demographic projections. *Int. J. Comput. Int. Sys.* **04**, 1–27 (2022)
3. Wang, Y., Wang, L., Yang, F., Di, W., Chang, Q.: Advantages of direct input-to-output connections in neural networks: The elman network for stock index forecasting. *Inform. Sci.* **547**, 1066–1079 (2021)
4. Schatz, M., Wheatley, S., Sornette, D.: The ARMA point process and its estimation. *Econom. Stat.* **36**, 2452–3062 (2021)
5. Zolfaghari, M., Gholami, S.: A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction. *Expert Syst. Appl.* **182**, 115149 (2021)
6. Wang, L., Ma, F., Liu, J., Yang, L.: Forecasting stock price volatility: New evidence from the GARCH-MIDAS model. *Int. J. Forecast.* **36**, 684–694 (2020)
7. Pai, P., Lin, C.: A hybrid arima and support vector machines model in stock price forecasting. *Omega* **33**, 497–505 (2005)
8. Kristjanpoller, W., Michell, K.: A stock market risk forecasting model through integration of switching regime, anfis and garch techniques. *Appl. Soft. Comput.* **67**, 106–116 (2018)
9. Harel, A., Harpaz, G.: Forecasting stock prices. *Int. Rev. Econ. Finance* **73**, 249–256 (2021)
10. Ning, G., Zhou, Y.: Application of improved differential evolution algorithm in solving equations. *Int. J. Comput. Int. Sys.* **06**, 14–19 (2021)
11. Li, R., Han, T., Song, X.: Stock price index forecasting using a multiscale modelling strategy based on frequency components analysis and intelligent optimization. *Appl. Soft Comput.* **11**, 109089 (2022)
12. Bose, A., Hsu, C.H., Roy, S.S., Lee, K.C., Mohammadi-ivatloo, B., Abimannan, S.: Forecasting stock price by hybrid model of cascading multivariate adaptive regression splines and deep neural network. *Comput. Electr. Eng.* **95**, 107405 (2021)
13. Guo, X., Wang, H., Yang, F.: Thermal power financial environment risk forecast model by combined stock multi-indicators basis on rbf neural network. *AASRI Procedia* **1**, 519–524 (2012)
14. Coskun, G.T., Yalciner, A.Y.: Determining the best price with linear performance pricing and checking with fuzzy logic. *Comput. Ind. Eng.* **154**, 107150 (2021)
15. Paiva, F.D., Cardoso, R.T.N., Hanaoka, G.P., Duarte, W.M.: Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Syst. Appl.* **115**, 635–655 (2019)
16. Han, J., Zhang, T., Li, Y., Liu, Z.: Rd-nmsvm: neural mapping support vector machine based on parameter regularization and knowledge distillation. *Int. J. Mach. Learn. Cyber* **15**, 6–26 (2022)
17. Ismail, S., Shabri, A., Samsudin, R.: A hybrid model of self-organizing maps (SOM) and least square support vector machine (lssvm) for time-series forecasting. *Expert Syst. Appl.* **38**, 10574–10578 (2011)
18. Das, S., Sahu, T.P., Janghel, R.R.: Stock market forecasting using intrinsic time-scale decomposition in fusion with cluster based

- modified CSA optimized ELM. *J. King Saud University Comput. Inform. Sci.* **3**, 101421 (2021)
19. Lin, Y., Yan, Y., Xu, J., Liao, Y., Ma, F.: Forecasting stock index price using the CEEMDAN-LSTM model. *North Am. J. Econ. Finance* **57**, 101421 (2021)
 20. Tang, J., Zuo, A., Liu, J., Li, T.: Seasonal decomposition and combination model for short-term forecasting of subway ridership. *Int. J. Mach. Learn. Cyber* **13**, 145–162 (2022)
 21. Ding, S., Li, R.: Forecasting the sales and stock of electric vehicles using a novel self-adaptive optimized grey model. *Eng. Appl. Artif. Intell.* **100**, 104148 (2021)
 22. Kiran, M.S., Siramkaya, E., Esme, E., Senkaya, M.N.: Prediction of the number of students taking make up examinations using artificial neural networks. *Int. J. Mach. Learn. Cyber* **13**, 71–81 (2021)
 23. Kamara, A.F., Chen, E., Pan, Z.: An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices. *Inform. Sci.* **594**, 1–19 (2022)
 24. Han, Y., Pan, P., Lv, H., Dai, G.: A hybrid optimization algorithm for water volume adjustment problem in district heating systems. *Int. J. Comput. Int. Sys.* **54**, 15–39 (2022)
 25. Deng, S., Xiao, C., Zhu, Y., Tian, Y.: Dynamic forecasting of the shanghai stock exchange index movement using multiple types of investor sentiment. *Appl. Soft Comput.* **111**, 109132 (2022)
 26. A stock time series forecasting approach incorporating candlestick patterns and sequence similarity. *Expert Syst. Appl.* **31**, 117595, (2022)
 27. Li, G., Jing, S., Shen, Y., Guo, B.: Efficient discrete particle swarm optimization algorithm for process mining from event logs. *Int. J. Comput. Int. Sys.* **07**, 15–21 (2022)
 28. Das, S.R., Mishra, D., Rout, M.: Stock market prediction using firefly algorithm with evolutionary framework optimized feature reduction for OSELM method. *Expert Syst. Appl.* **4**, 100016 (2019)
 29. Jain, L., Katarya, R., Sachdeva, S.: Opinion leader detection using whale optimization algorithm in online social network. *Expert Syst. Appl.* **142**, 113016 (2020)
 30. Sun, G., Shang, Y., Yuan, K., Gao, H.: An improved whale optimization algorithm based on nonlinear parameters and feedback mechanism. *Int. J. Comput. Int. Sys.* **68**, 15–38 (2022)
 31. Salgotra, R., Abouhawwash, M., Singh, U., Saha, S., Mittal, N., Mahajan, S., Pandit, A.K.: Multi-population and dynamic-iterative cuckoo search algorithm for linear antenna array synthesis. *Appl. Soft Comput.* **113**, 108004 (2021)
 32. K. C. S. Hybrid models for intraday stock price forecasting based on artificial neural networks and metaheuristic algorithms. *Pattern Recogn. Lett.* **147** (2021) 124–133
 33. Das, S., Sahu, T. P., Janghel, R. R.: Stock market forecasting using intrinsic time-scale decomposition in fusion with cluster based modified CSA optimized ELM. *J. King Saud University Comput. Inform. Sci.* 124–133, (2021)
 34. Wang, Y., Li, J., Pei, Y., Ma, Z., Jia, Y., Wei, Y.: An adaptive high-voltage direct current detection algorithm using cognitive wavelet transform. *Inform. Process. Manag.* **59**, 102867 (2022)
 35. Arteche, J., García-Enríquez, J.: Singular spectrum analysis for signal extraction in stochastic volatility models. *Econom. Stat.* **1**, 85–98 (2017)
 36. Guo, W., Liu, Q., Luo, Z., Tse, Y.: Forecasts for international financial series with vmd algorithms. *J. Asian Econ.* **80**, 101458 (2022)
 37. Xu, M., Shang, P., Lin, A.: Cross-correlation analysis of stock markets using EMD and EEMD. *Physica A Stat. Mech. Appl.* **442**, 82–90 (2016)
 38. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* **454**, 903–995, (1998)
 39. Wu, Z., Huang, N.E.: Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **1**, 1–41 (2009)
 40. Yeh, J.-R., Shieh, J.-S., Huang, N.E.: Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Adv. Adapt. Data Anal.* **2**, 135–156 (2010)
 41. Hyvriinen, A., Oja, E.: Independent component analysis: algorithms and applications. *Neural Netw.* **13**, 411–430 (2000)
 42. Zhu, J.: Multi-factor performance measure model with an application to fortune 500 companies. *Eur. J. Oper. Res.* **123**, 105–124 (2000)
 43. Wang, C.: A sample entropy inspired affinity propagation method for bearing fault signal classification. *Digit. Signal Process.* **102**, 102740 (2020)
 44. Zhang, N., Lin, A., Ma, H., Shang, P., Yang, P.: Weighted multivariate composite multiscale sample entropy analysis for the complexity of nonlinear times series. *Physica A Stat. Mech. Appl.* **508**, 595–607 (2018)
 45. Liu, W., Wang, Z., Zeng, N., Yuan, Y., Liu, X.: A novel randomised particle swarm optimizer. *Int. J. Mach. Learn. Cyber* **73**, 529–540 (2021)
 46. Rashno, A., Shafipour, M., Fadaei, S.: Particle ranking: An efficient method for multi-objective particle swarm optimization feature selection. *Knowl. Based Syst.* **5**, 108640 (2022)
 47. Chakravarty, S., Dash, P.: A PSO based integrated functional link net and interval type-2 fuzzy logic system for predicting stock market indices. *Appl. Soft Comput.* **12**, 931–941 (2012)
 48. Liu, W., Wang, Z., Zeng, N., Yuan, Y., Alsaadi, F.E., Liu, X.: A pso based deep learning approach to classifying patients from emergency departments. *Int. J. Mach. Learn. Cyber* **73**, 529–540 (2021)
 49. Levantesi, S., Nigri, A., Piscopo, G.: Clustering-based simultaneous forecasting of life expectancy time series through long-short term memory neural networks. *Int. J. Approx. Reason.* **140**, 282–297 (2022)
 50. Giovannelli, A., Massacci, D., Soccorsi, S.: Forecasting stock returns with large dimensional factor models. *J. Empir. Finance* **63**, 252–269 (2021)
 51. Cheng, H., Shi, Y.: Forecasting china's stock market variance. *Pacific-Basin Finance J.* **64**, 101421 (2020)
 52. Dai, Z., Dong, X., Kang, J., Hong, L.: Forecasting stock market returns: New technical indicators and two-step economic constraint method. *North Am. J. Econ. Finance* **53**, 101216 (2020)
 53. Li, H., Bai, J., Li, Y.: A novel secondary decomposition learning paradigm with kernel extreme learning machine for multi-step forecasting of container throughput. *Physica A: Stat. Mech. Appl.* **534**, 122025 (2019)
 54. Xie, G., Zhang, N., Wang, S.: Data characteristic analysis and model selection for container throughput forecasting within a decomposition-ensemble methodology. *Transport. Res. Part E Logist Transp. Rev.* **108**, 160–178 (2017)
 55. W. Chen, H. Zhang, M. K. M. and Lifan Jia, Mean-variance portfolio optimization using machine learning-based stock price prediction. *Appl. Soft Comput. J.* **100**, 106943, (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.