# RESEARCH

# **Open Access**

# The volatility mechanism and intelligent fusion forecast of new energy stock prices



Guo-Feng Fan<sup>1</sup>, Ruo-Tong Zhang<sup>1</sup>, Cen-Cen Cao<sup>1</sup>, Li-Ling Peng<sup>1</sup>, Yi-Hsuan Yeh<sup>2</sup> and Wei-Chiang Hong<sup>2,3\*</sup>

\*Correspondence: samuelsonhong@gmail.com

 <sup>1</sup> School of Mathematics and Statistics, Ping Ding Shan University, Ping Ding Shan 467000, Henan, China
 <sup>2</sup> Department of Information Management, Asia Eastern University of Science and Technology, No. 58, Sec. 2, Sichuan Rd., Panchiao District, New Taipei 22064, Taiwan
 <sup>3</sup> Department of Information Management, Yuan Ze University, Taoyuan 320315, Taiwan

# Abstract

The new energy industry is strongly supported by the state, and accurate forecasting of stock price can lead to better understanding of its development. However, factors such as cost and ease of use of new energy, as well as economic situation and policy environment, have led to continuous changes in its stock price and increased stock price volatility. By calculating the Lyapunov index and observing the Poincaré surface of the section, we find that the sample of the China Securities Index Green Power 50 Index has chaotic characteristics, and the data indicate strong volatility and uncertainty. This study proposes a new method of stock price index prediction, namely, EWT-S-ALOSVR. Empirical wavelet decomposition extracts features from multiple factors affecting stock prices to form multiple sub-columns with features, significantly reducing the complexity of the stock price series. Support vector regression is well suited for dealing with nonlinear stock price series, and the support vector machine model parameters are selected using random wandering and picking elites via Ant Lion Optimization, making stock price prediction more accurate.

**Keywords:** Empirical wavelet transform, Support vector machine, Ant Lion algorithm, Stock price index forecasting

# Introduction

Globalization has advanced recently, and although China has made continuous progress in the economic field, all these achievements depend on an extensive development model. The long-term development of this production has led to high pollution, emissions, and energy consumption. To date, these problems have not been resolved. Environmental problems and the consequent shortage of resources have become increasingly serious. Although residents' incomes have greatly increased, their happiness has also in fact decreased. Consequently, people have begun to attach significant importance to environmental protection and energy security. Thus, Guo (2023) pointed out that the past few years have witnessed a period of rapid development in China's environmental protection industry. The new energy industry is undergoing rapid development. The government attaches great importance to all kinds of new energy production equipment and industries and has issued a series of favorable policies to promote their development. The National Energy Administration issued a notice in June 2021 on submitting the pilot plan for rooftop distributed photovoltaic development in the entire county (including



© The Author(s) 2024. **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/.

the city name and its associated district) and announced the first batch of pilot counties (including the city name and its associated district) in September of the same year (Li and Li 2023). The financial sector is at the core of resource allocation; therefore, more capital should be used for the development of the new energy industry, which is an indispensable step in national development strategies. Stock price fluctuations in an industry may reflect its short-term development to some extent. By forecasting price changes in new energy stocks, investors can make appropriate decisions, improve the profitability of their investments, reduce potential losses, and reflect the development of new energy sectors to a certain extent, thus promoting the effective allocation of related resources (Jiang et al. 2023). New energy stock price forecasting plays an important role in financial decision making, investment management, and quantitative trading in related fields.

The new energy industry is strongly supported by the state, and accurate stock price predictions can provide a good understanding of the development of new energy sources. This study propose an accurate new-energy stock price prediction model that reflects the development of the new energy industry. China's energy consumption structure has long been dominated by coal, placing a huge environmental burden on the country (Sun et al. 2019). In September 2020, China pledged to reach its peak greenhouse gas emissions by 2030 and achieve carbon neutrality by 2060 (Nie et al. 2021). To achieve this strategic goal, governments, companies, and academics have placed great emphasis on the new energy industry. The rapid development of new energy sources has also attracted the attention of capital market investors and provided new investment opportunities (Elsayed et al. 2020; Wang et al. 2022). New energy stocks should have a place in any equity portfolio and investors should seek a more balanced, high-quality, and multi-strategy portfolio.

# Literature review

The new energy industry is undergoing rapid development. The Chinese government attaches great importance to all types of new energy production equipment and the new energy industry, and it has introduced a series of favorable policies to promote its development. The importance of eco-environmental protection in sustainable development strategies is self-evident, and eco-environmental protection with the new energy industry at its core has become a trend (Wang et al. 2010; Zhang et al. 2009). The financial sector is central to the allocation of resources and should allocate more capital to the development of the new energy industry, which is an essential step in China's development strategy. The new energy industry is among the industries that China strongly supports, and its healthy development is of great practical importance. Share price fluctuations in an industry may reflect its short-term development. By forecasting the price movements of new energy stocks, investors can make appropriate decisions to improve the profitability of their investments, reduce potential losses, and reflect the development of the new energy sector, thereby facilitating the efficient allocation of relevant resources. New energy stock price forecasting plays an important role in financial decision making, investment management, and quantitative trading in relevant sectors. However, the cost of new energy, ease of utilization, economic situation, need for secure energy supply, technical issues, and the impact of the policy environment lead to variable stock prices for new energy. Therefore, investors face significant challenges when making investment decisions (Bai et al. 2019; Wang and Luo 2021). For example, when economic growth slows down and market demand is low, product sales and prices in the new energy sector may be affected, resulting in lower company earnings and stock prices. Economic uncertainties, such as global trade friction, can also impact the industry's earnings and prospects. As clean energy continues to gain momentum, the traditional energy sector faces heavy challenges, and the operation of many traditional thermal power plants is under extreme pressure. The state has begun regulating prices in the traditional energy sector and has increased environmental inspections and penalties for traditional energy companies to promote the development of new energy sources and clean energy alternatives to traditional energy sources. These regulatory measures have created market pressure for many traditional energy companies and have likewise created development opportunities for new energy companies (Ba et al. 2022). This, in turn, affects stock prices. Accordingly, the energy sector is undergoing a "rapid and transformative green transformation," especially in the context of the energy crisis caused by geo-conflicts and the increased focus on environmental, social, and governance issues, and the potential of the new energy sector is clear. The risk-return behavior of the new energy stock market and its influencing factors have become a hot research topic.

Stock prediction is a popular research topic for scholars worldwide, and many stock prediction models have been studied, including support vector machines (SVMs), generalized regression neural networks (GRNNs), long short-term memory (LSTM), random forests, and recurrent neural networks (RNNs). However, with the continuous improvement and development of the statistical theory, SVMs have been increasingly receiving attention. Zhou et al. (2023) established PSO-BPNN and PSO-SVR models for Jacking prediction. The results show that both models perform well in prediction and that the PSO-SVR model outperforms the PSO-BPNN model in terms of prediction accuracy and generalization ability. SVMs can handle small sample learning problems. In the case of an insufficient sample size, SVM is a new small sample learning method with a solid theoretical foundation. It does not involve probability measures or laws of large numbers; therefore, it differs from existing statistical methods. Essentially, it avoids the traditional inductive reasoning process and achieves effective "transduction reasoning" from training samples to predicting samples, greatly simplifying common classification and regression problems. Neural networks have high classification accuracy, strong distributed storage, and learning capabilities. However, they require a large number of parameters, and the learning time is too long, which may result in failure to achieve the learning objectives. Moreover, when data are insufficient, the neural network cannot operate. The most serious problem is the inability to explain one's reasoning and processes. However, the theory and learning algorithms require further improvements. Therefore, this study selects support vector regression (SVR) as the basis for the prediction model. Vapnik (1995) was proposed in the mid-1990s and has since been rapidly applied in many fields, including financial forecasting, owing to its simple algorithm structure and strong generalization ability. Kim (2003) used the following three methods for stock price prediction: SVR, neural network, and case-based reasoning.In addition, he compared their results, confirming that SVR has better performance and prediction accuracy. Shen and Shafiq (2020) predicted short-term stock market price trends using a comprehensive

deep-learning system, and Lee (2009) applied a SVM with a hybrid feature selection method to predict stock trends. Ni et al. (2011) predicted stock trends by analyzing feature selection and SVM. Lei (2018) applied a wavelet neural network prediction method to predict stock price trends.

As research progressed, individual forecasting models no longer met the demand for accuracy, and various optimization algorithms were gradually used to refine individual forecasting models to better predict the future trends of stocks. For example, Ning et al. (2022) proposed a SVR model (GA-PSO-SVR) based on a genetic algorithm (GA), particle swarm optimization (PSO), and a hybrid GA-PSO algorithm to obtain a GA-PSO-SVR model and search for and obtain the best SVR parameters to improve the accuracy of the prediction model. Sun et al. (2022) combined SVM with PSO to make full use of the unique advantages of SVM in handling small-sample regression problems and PSO global search optimization to improve the convergence speed and achieve depth and breadth optimization. Experimental results showed that the method improves the efficiency of parameter selection for SVMs and provides more accurate predictions. Zhu et al. (2022) used the optimization performance of the cuckoo search algorithm (CS) and bat algorithm (BA) combined with the SVR principle to apply the above two algorithms to optimize the kernel function parameters in SVR. The prediction performance of the CS-SVR and BA-SVR models was also tested. In terms of the overall prediction rate, the two algorithms significantly outperformed the traditional SVR model. Zheng et al. (2021) introduced the BA algorithm to optimize three free parameters of the SVR machine-learning model and constructed a BA-SVR hybrid model. The experimental results showed that the BA-SVR model outperformed the polynomial and sigmoid kernel SVR models without optimizing the initial parameters. Gu et al. (2021) used the FOA to optimize the penalty factors and parameters of the RBF in the F-SVM. The initialization of the matrix, determination of the hyperplane, and solution of the transformation matrix were performed. The experimental results validated the performance of the FOA-F-SVM and showed that it could generate more suitable model parameters and significantly reduce the computational cost, resulting in a higher classification accuracy. Yang et al. (2022) used a SVM model to construct a proxy model of the performance function based on a small number of samples generated using the finite element method to achieve a small sample. An explicit expression of the implicit nonlinear performance function under the condition of small sample sizes was achieved. This method has significant advantages in terms of computational accuracy and efficiency, and it is suitable for solving structural reliability problems in complex engineering.

The main optimization algorithms are the PSO algorithm (Ma et al. 2022; Ning et al. 2022; Sun et al. 2022), BA (Hao et al. 2018; Zheng et al. 2021; Zhu et al. 2022), FOA (Gu et al. 2021; Pan et al. 2021; Zhang and Fang 2015), and the grasshopper optimization algorithm (GOA) (Barman and Choudhury 2018; Yang et al. 2022). PSO is simple and easy to implement and does not require many parameter adjustments. However, the PSO algorithm is prone to falling into local optima and does not achieve satisfactory results. BA can be seen as a weakened PSO algorithm, which has shortcomings such as slow convergence speed and low convergence accuracy, and it is prone to falling into local minima, which seriously limits the application field of BA. The operation process of GOA is relatively complex and prone to premature phenomena. FOA

has drawbacks such as the inability to take negative values for algorithm candidate solutions, poor population diversity, and weak local search ability. Although the grey wolf algorithm has a strong convergence performance, is easy to implement, and can achieve a balance between local optimization and global search, it has problems such as premature convergence, low convergence accuracy when facing complex problems, and insufficient convergence speed. Saleem et al. (2021) compared their proposed Ant Lion Optimization (ALO) technology with other recent technologies such as ant colony, grasshopper, and moth flame optimizations. The comparison results demonstrated the effectiveness of ALO optimization technology in implementing energy-saving protocols for WBANs for remote monitoring applications. Ahamad et al. (2022) utilized a new search algorithm, ALO, for system approximation and compared it with a GA and PSO. Finally, the effectiveness of the ALO method was verified in terms of the convergence speed and CPU usage time.

# **Research motivation**

Previous stock index forecasting methods can easily fall into local optima, and the convergence speed is slow and highly dependent on the initial conditions. Highdimensional problems are not handled well, so the convergence accuracy needs to be increased and improved. Therefore, it is necessary to find a new forecasting method. To address these problems, this study proposes a forecasting method for the EWT-S-ALOSVR. ALO is easy to use, with fewer algorithm parameters and no gradient information in the objective function, and has efficient search and convergence efficiency (Mirjalili 2015). The main inspiration for this algorithm is based on the foraging behavior of ant lion larvae. We establish a mathematical model based on this behavior. Many assumptions of the Ant Lion algorithm can be found to be similar in stock prices. The model assumes that ants perform random walks in the search space, resembling changes in stock prices. Random walks are also influenced by ant lion traps, just like stock price changes are influenced by relevant policies, laws, and regulations. Therefore, using the ALO algorithm to study stock price fluctuations is rational. In response to the shortcomings and improvement needs of existing technologies, this study provides a construction method and application of a new energy index prediction model, aiming to provide an accurate and stable stock price prediction method to solve the problems caused by various macro- and micro-influencing factors in predicting the new energy index. It includes a hybrid EWT, ALO algorithm, and SVR model, extracts the characteristics of the stock index series through the EWT, and optimizes the parameter combination of the SVR model with ALO. A mixture model is proposed. This model uses the EWT to decompose vibration signals, thereby reducing the modal aliasing that currently exists in signal decomposition. The EWT decomposes stock index data into several intrinsic mode function components and finds that the decomposed intrinsic mode function components have a certain regularity. Subsequently, SVMs are constructed to predict these IMFs separately. To obtain the optimization parameters of SVMs, the ALO algorithm is used to automatically complete the parameter selection in SVM modeling. These predicted values are then reconstructed to obtain the final new energy index prediction results.

# **Research contribution**

The main contributions of this study are as follows:

- 1. We analyze the characteristics of the stock index sequence, calculate the Lyapunov index, and observe the Poincaré surface of the section, indicating that the sample has chaotic characteristics and the data have strong volatility and uncertainty.
- 2. Regarding the volatility of the stock data, the EWT is applied to decompose the original sequence and preprocess the data. Furthermore, a feature analysis is performed on the processed data.
- 3. Several conventional algorithms are proposed for this purpose. Considering the volatility and nonlinearity of stock prices, the SVR algorithm is applied to predict stock indices, and an ALO-optimized SVR model is used to propose a stock prediction model based on the ALO algorithm mixed SVR model.
- 4. The EWT method is combined with the ALO-based SVR model, and the ALOSVR model is applied to model each decomposed IMF. Conduct simulation experiments on the stock index to verify whether the optimized SVR has high stock prediction accuracy.

To better predict the stock price index trend and explore the fluctuation characteristics of different levels of the stock price index sequence. In this study, according to the concept of "decomposition and reconstruction–prediction–integration," an empirical wavelet decomposition model combining adaptive decomposition concept and wavelet transform theory is constructed. Simultaneously, ALO is performed based on the SVM model. We seek the optimal parameters for model prediction to realize accurate stock index prediction. Next, to verify and compare the effectiveness and practicability of the proposed EWT-S-ALOSVR model, this study set up nine other models as control models, applies them to the daily closing price of the CSI Green Power 50 Index, and evaluates the forecasting effect.

This paper is divided into four parts, which are organized as follows: The first section is an introduction. This section introduces the background and significance of this study. It then summarizes the research history and achievements of experts and scholars at home and abroad regarding the feasibility and methods of stock price prediction and expounds on the contributions and shortcomings of the research.

The second section introduces the theories related to the experimental and control models in this study, including the SVR model, ALO algorithm and EWT algorithm, which lays a theoretical foundation for subsequent modeling and prediction.

The third section introduces the selected stock price indices. This study introduces the fluctuation characteristics, stability evaluation, and significance of the research object of the stock price index. This is followed by an explanation of the stock price index and EWT decomposition of the daily closing price. The proposed method is compared and analyzed with various models such as the single SVR, PSO-SVR, LSTM, RNN, convolutional neural network (CNN), GRNN, and ARMA (1,2)-GARCH (1,1).

The fourth section presents the conclusion and prospects. The methods proposed in this study are summarized and prospected, and the development of the new energy industry is reviewed and evaluated.

# **Models and methods**

An important foundation of modern financial theory is the efficient market hypothesis proposed by Fama in 1965 (as quoted in Sandubete et al. 2023). According to this hypothesis, if a market is perfectly efficient, then the price of a stock will reflect all the information in the market, and updates in information will be quickly reflected in the price of the stock. This ideal situation does not correspond to the reality of the stock market. The efficient market hypothesis classifies markets according to the amount of information reflected in stock prices: (1) In weakly efficient markets, the current stock price already reflects all historical information, and no additional gains can be made by analyzing past prices; therefore, the only way to predict stock prices is to include public or private interest. (2) In semi-strong efficient markets, historical prices and public information do not generate additional returns because they are already fully reflected in the prices of financial assets, and stock prices need to be informed by insider information. (3) In strongly efficient markets,, the use of historical, public, and insider information can no longer generate excess returns. In these three markets, it is only in weakly efficient markets that it makes sense to predict stock prices because insider information is not available to the general investor population, whereas public news and media information are available. This study uses stock index closing price data to predict the rise and fall of the New Energy Concept Index, which is based on the Chinese stock market as a weakly efficient market. In addition, the study is based on the following four important assumptions: markets are not completely random, history repeats itself, markets follow the rational behavior of people, and markets are "perfect."

## **Empirical wavelet transform**

The EWT can adaptively detect different components of a signal without requiring prior assumptions or model fitting of the signal. This makes the EWT well suited for processing nonlinear and non-smooth signals. In addition, the EWT can handle the noise and interference present in the signal, thus improving its decomposition accuracy. In practice, the parameters must be selected and optimized according to the specific situation to obtain better decomposition results (Gilles 2013; Lou et al. 2022).

The formulae for calculating the detail factor  $W_f^{\sigma}(0, t)$  and the approximation factor  $W_f^{\sigma}(0, t)$  are shown in Eqs. (1) and (2):

$$W_f^{\sigma}(n,t) = f, \ \psi_n(t) = \int f(\tau) \overline{\psi_n(\tau-t)} d\tau = F^{-1}\left(\hat{f}(\omega)\overline{\psi_n(\omega)}\right),\tag{1}$$

$$W_f^{\sigma}(0,t) = f, \,\varphi_1 = \int f(\tau) \overline{\varphi_1(\tau-t)} d\tau = F^{-1} \Big( \hat{f}(\omega) \overline{\varphi_1(\omega)} \Big), \tag{2}$$

where  $\hat{\psi}_n(\omega)$ ,  $\hat{\varphi}_1(\omega)$  are the Fourier transform of the empirical wavelet function and the empirical scale function, respectively, as shown in Fig. 1. The detailed derivation of  $\hat{\psi}_n(\omega)$ ,  $\hat{\varphi}_1(\omega)$  is as follows:

Suppose that the Fourier support interval is divided into N consecutive intervals  $[0, \pi]$ .  $\Lambda_n = [\omega_{n-1}, \omega_n], \cup_{n=1}^N \Lambda_n = [0, \pi]$  is the boundary of each segmented section, and an excess region (of width  $2\lambda_n$ ) is defined, which is thought to be the centroid  $T_n$ .



Fig. 1 Segmentation of the Fourier axis

The empirical wavelets are band-pass filters on an interval. Following the construction of Littlewood–Paley and Meyer's wavelets, for any N>0, the sums are obtained using Eqs. (3) and (4):

$$\hat{\varphi}_{n}(\omega) = \begin{cases} 1, & (|\omega|) \leq (1-\rho)\omega_{n} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\rho\omega_{n}}(|\omega|) - (1-\rho)\omega_{n}\right)\right], & (1-\rho)\omega_{n} \leq |\omega| \leq (1+\rho)\omega_{n}, \\ 0, & others \end{cases}$$
(3)

$$\hat{\psi}_{n}(\omega) = \begin{cases} 1, & (1+\rho)\omega_{n} \leq (1-\rho)\omega_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\rho\omega_{n}}(|\omega|) - (1-\rho)\omega_{n+1}\right)\right] & (1-\rho)\omega_{n+1} \leq |\omega| \leq (1+\rho)\omega_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\rho\omega_{n}}(|\omega|) - (1-\rho)\omega_{n}\right)\right], & (1-\rho)\omega_{n} \leq |\omega| \leq (1+\rho)\omega_{n} \\ 0, & others \end{cases}$$
(4)

where

$$\beta(x) = x^4 \left(35 - 84x + 70x^2 - 20x^3\right), \forall x \in [0, 1], \gamma \in [0, 1], and \gamma < \min_n \frac{\omega_{n+1} - \omega_n}{\omega_{n+1} - \omega_n}, \lambda_n = \gamma \omega_n$$
  
The resulting original signal  $f(t)$  is shown in Eq. (5):

$$f(t) = W_f^s(0,t) * \phi_1(t) + \sum_{n-1}^N W_f^s(n,t) * \varphi_n(t) = \left(W_f^s(0,\omega)\phi_1(\omega) + \sum_{n-1}^N W_f^s(n,\omega) * \varphi_n(\omega), (5)\right)$$

where  $\hat{W}^{\sigma}_{f}(0,\omega)$  and  $\hat{W}^{\sigma}_{f}(n,\omega)$  are, respectively, the Fourier transform forms of  $\hat{W}^{\sigma}_{f}(0,t)$ and  $\hat{W}_{f}^{\sigma}(n, t)$ .

The decomposition of EWT is similar to that of EMD, and the result of the original signal decomposition is shown in Eq. (6):

$$f(t) = \sum_{i=0}^{N} f_i(t).$$
 (6)

Each  $f_i(t)$  is an AM-FM function that can be written as

$$f_i(t) = F_i(t)\cos(\phi_i(t)), F_i, \phi_i(t) > 0$$

According to Eq. (6), we obtain

$$f_0(t) = W_f^e(0, t) * \phi_1(t), \tag{7}$$

$$f_k(t) = W_f^e(k, t) * \psi_k(t).$$
(8)

The purpose of EWT is to decompose f(t) into N+1 single components of  $f_k(t)$ .

The purpose of Fourier spectral division is to separate the different parts of the spectrum corresponding to the modes. Two cases may arise:

If the case  $M \ge N$  appears, when the first N-1 maxima should be taken.

If M < N, that is, if there is a signal with a smaller number of modes than the ideal case, all detected maxima must be retained, and N must be reset to the appropriate value. *N* is the number of local maxima greater than the threshold corresponding to the limit for each segment. We add 0 and  $\pi$  to this limit set to obtain the N+1 limit.

## Support vector machines

The main details of the SVM model are as follows:

For a data set  $\{x_i\}_{i=1}^N$  with N observations, there is a feature mapping function  $\varphi(\cdot)$  that maps these observations to an N-dimensional feature space and also in real-number format,  $\varphi(\cdot) : \mathbb{R}^N \to \mathbb{R}^{N_f}$ . In the feature space, a linear function f is used to specify the linear relationship between the mapped feature points  $(x_i)$  and actual values  $(y_i)$ . The linear function is the SVR function, given in Eq. (9):

$$f(\mathbf{x}) = \mathbf{w}^{T} \varphi(\mathbf{x}) + b, \tag{9}$$

where  $f(\mathbf{x})$  are the forecasted values of the mapped feature points  $(x_i)$ . The weight  $\mathbf{w} \in \mathbb{R}^{N_f}$  and constant intercept b ( $b \in \mathbb{R}$ ) are determined by optimizing the empirical risk function using the SVR theory, yielding Eq. (10):

$$R_{emp}(f) = C \frac{1}{N} \sum_{i=1}^{N} K_{\varepsilon}(y_i, f(\mathbf{x})) + \frac{1}{2} \|\mathbf{w}\|^2,$$
(10)

$$K_{\varepsilon}(y_{i},f(\mathbf{x})) = \begin{cases} \left| f(\mathbf{x}) - y_{i} \right| - \varepsilon, & \text{if } \left| f(\mathbf{x}) - y_{i} \right| > \varepsilon \\ 0, & \text{otherwise} \end{cases},$$
(11)

where  $K_{\varepsilon}(y_{I}, f(x))$  represents the main empirical risk due to its theoretical definition, also known as the  $\varepsilon$  insensitive loss function; C and  $\varepsilon$  are parameters that need to be determined and play a key role in the SVR modeling process. According to Eq. (11), the empirical risk is zero only if the absolute value of the forecasting error  $(|f(x) - y_i|)$  is  $\leq \varepsilon$ . In the second term,  $\frac{1}{2}w^2$  is the gradient of the SVR function. C is used to balance the empirical risk and gradient.

To solve Eq. (10), the quadratic programming method is employed, and two slack variables,  $\Im$  and  $\Im^*$ , are used to measure the distances between the actual values and the boundary values of the  $\varepsilon$ -tube. Equation (10) is then converted into a normal programming form with the constraints given in Eq. (12):

$$\operatorname{Min} R_{emp}(\mathbf{w}, \mathfrak{I}, \mathfrak{I}^*) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} (\mathfrak{I}_i + \mathfrak{I}_i^*),$$
(12)

with the limitations,

$$y_i - f(\mathbf{x}_i) \le \varepsilon + \mathfrak{I}_i^*, - y_i + f(\mathbf{x}_i) \le \varepsilon + \mathfrak{I}_i, \mathfrak{I}_i^*, \mathfrak{I}_i \ge 0$$

Applying the Lagrange multiplier method yields weight w, as shown in Eq. (13):

$$\mathbf{w}^* = \sum_{i=1}^N \left( \alpha_i^* - \alpha_i \right) \varphi(\mathbf{x}_i), \tag{13}$$

where  $\alpha_I$  and  $\alpha_i^*$  are the Lagrangian multipliers and satisfy the condition  $\alpha_i^* * \alpha_i = 0$ . Finally, the SVR function is expressed by Eq. (14):

$$f(\mathbf{x}) = \sum_{i=1}^{N} \left( \alpha_i^* - \alpha_i \right) K\left( \mathbf{x}_i, \mathbf{x}_j \right) + b,$$
(14)

where  $K(x_i, x_j)$  is the kernel function, which is the inner product of the feature mapping functions,  $\varphi(\mathbf{x})$ , of two points,  $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ . The Gaussian function,  $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2upgamma^2}\right)$ , is the most used kernel function owing to its advanced mapping capability to deal with complicated nonlinear relationships. To reduce the computing loading, the Gaussian exponential kernel function (another classical Gaussian kernel function),  $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\gamma^2}\right)$ , is used here.

# Ant Lion Optimization algorithm

ALO is a nature-inspired metaheuristic algorithm introduced by Mirjalili (2015) in 2015. This optimizer demonstrates the hunting mechanism of ant lions. It is a species from the ant family that hunts during its larval stage. Normally, they live for 3–5 years, mostly in the larval stage and only a few weeks in the adult stage. At the larval stage, ant lions are toothed worms, and they walk back into the sand to form a funnel and hide to hunt for ants. During the random walk, when the ants reach the funnel, the ant lion immediately confuses them by throwing sand, making ants fall into the pit faster. After capturing the prayer, the ant lion eats it, leaving the pieces outside the pit and rebuilding the pit in the sand for the next ant hunt. The adaptability and quality of the ant lion depend on its number. Ants it pursues. The best ant lion can trap more ants and has the highest probability of capturing them. This process is implemented mathematically in the ALO by representing the position and fitness of ants and ant lions.

In the ALO problem, ants and ant lions are the solution vectors derived from the equation, and the optimal ant lion position in the search space for these vectors is the output of the control variables. Initialize the basic parameters of ALO, such as the number of ants and antlions, search space, and bounding constraints for each variable (Ansal 2020).

**Step 1** When an ant searches for food, it walks randomly in the search space. In each iteration, the random wandering of the ant is based on the new position of the control variable and is mathematically modeled in Eq. (15):

$$X(t) = [0, cumsum(2 \times r(t_1) - 1), cumsum(2r(t_2) - 1), ..., cumsum(2 \times r(t_T) - 1)],$$
(15)

where *cumsum* is the cumulative sum of the ant-walk positions; T is the maximum number of iterations; t is the current number of iterations; and r(t) is a random function. As shown in Eq. (16),

$$r(t) = \begin{cases} 1, \ rand > 0.5\\ 0, \ rand \le 0.5 \end{cases},$$
(16)

where rand is a random number in the [0,1] interval. As the ants move randomly in each step, their position is updated from the original position to a random position in the search space using Eq. (17). This equation helps deploy ordinary solutions to better solutions. The normalization process for  $X_i^t$  is

$$X_{i}^{t} = \frac{(X_{i}^{t} - a_{i}) \times (d_{i} - c_{i}^{t})}{(d_{i}^{t} - a_{i})} + c_{i},$$
(17)

where  $X_i^t$  is the normalized position of the *i*th dimensional variable at the *t*th iteration,  $a_i$  is the minimum random walk of the *i*th variable, and  $d_i^t$  is the maximum.  $c_i^t$  is the minimum value of the *i*th variable in the *t*th iteration, and  $d_i^t$  is the maximum value of the *i*th iteration.

**Step 2** As mentioned above, the ant lion trap affects the random wandering of ants. To model this hypothesis mathematically, we use Eqs. (18) and (19):

$$c^{t} = Antlion_{i}^{t} + c^{t}, \tag{18}$$

$$d^t = Antlion_i^t + d^t, \tag{19}$$

where  $c^t$  is the minimum of all variables in the *t*th iteration,  $d^t$  is the maximum of all variables in the *t*th iteration,  $c_i^t$  is the minimum of all variables in the *i*th ant,  $d_i^t$  is the maximum of all variables in the *i*th ant, and *Antlion*<sub>j</sub><sup>t</sup> is the position of the *j*th ant lion selected in the *t*th iteration. Equations (18) and (19) express the random walk of the selected ants in the hypersphere as defined by vectors *c* and *d*.

Step 3 The predatory process of the ant lion can be expressed as

$$I = 10^{\omega} \frac{t}{T} \begin{cases} w = 2, t > 0.1T \\ w = 3, t > 0.5T \\ w = 4, t > 0.75T , \\ w = 5, t > 0.9T \\ w = 6, t > 0.95T \end{cases}$$
(20)

where  $I = 10^{\omega} \frac{t}{T}$ , where *t* is the current iteration; *T* is the maximum number of iterations; and  $\omega$  is a constant defined according to the current iteration. The relationships between *T* and  $\omega$  are t > 0.1T,  $\omega = 2$ ; t > 0.5T,  $\omega = 3$ ; t > 0.75T,  $\omega = 4$ ; t > 0.9T,  $\omega = 5$ ; *andt* > 0.95T,  $\omega = 6$ .

An ant is captured by an ant lion when it is better adapted than the ant lion. At this point, the ant lion updates its position based on the position of the ant as shown in Eq. (21):

$$Antlion_{j}^{t} = Ant_{i}^{t} \text{ if } f\left(Ant_{i}^{t}\right) > f\left(Antlion_{j}^{t}\right), \qquad (21)$$

where  $Antlion_j^t$  represents the position of the *j*th ant lion in the *t*th iteration, and  $Ant_i^t$  represents the position of the *i*th ant in the tth iteration.

**Step 4** The elite phase of the ALO helps to obtain the best admixture solution in the full problem optimization process. In this algorithm, instead of walking around the ant lion selected by the roulette wheel, the ants walk around the elite ant lion. Finally, the new position of the ant is described according to the average of the two walks, as shown in Eq. (22):

$$Ant_i^t = \frac{R_A^t + R_E^t}{2},\tag{22}$$

where  $R_A^t$  is the random walk around the elite in the *t* iteration,  $R_E^t$  is the random walk around the elite in the *t*th iteration, and  $Ant_i^t$  represents the position of ant *i* in the *t* iteration.

In the ALO algorithm, mechanisms such as random wheel selection and ant random walks can be used by Ant Lion to ensure that the algorithm takes full advantage of its exploratory role in the research space and effectively optimizes its application model of the ALO algorithm. The use of a random roulette wheel selection by Ant Lion ensures the exploration of the research space. In addition, the random walk process and roulette-wheel strategy effectively prevent the ALO algorithm from falling into a local optimum dilemma. Thus, the ALO algorithm can improve the effectiveness of predictions to a certain extent. Mirjalili (2015) compared the ALO algorithm with other intelligent optimization algorithms in numerous experiments with different dimensions using single-peaked, multi-peaked, and hybrid functions with high convergence accuracy.

The idea of the ALO algorithm for optimizing SVR parameters is as follows: First, initialize the parameters in the SVR and then calculate the adaptive value for each ant based on the adaptive function. If the updated adaptive value is greater than that of the previous position, it is updated and replaced with the start time of the next iteration. By successive iterations, if the extreme value obtained is below a fixed threshold, or if the maximum number of iterations is reached, the best parameters C and  $\gamma$  are sent, and the prediction model ALOSVR is constructed using the best parameters.

## **Empirical analysis**

## Dataset selection and feature extraction

The CSI Green Power 50 Index (Teng 2023) selects 50 listed securities whose businesses involve hydropower, wind power, photovoltaic power generation, and other energy generation as samples to show the overall performance of listed securities with a green power theme. The index was officially released on May 9, 2022, with December 31, 2013, as the base date and 1,000 points as the basis points.

The CSI Green Power 50 Index is compiled by the China Securities Index Company Limited to reflect the performance of China's green power sector stock market. The index selects 50 stocks in the stock market with large market capitalization, good liquidity, and certain representativeness in the green power sector as sample stocks and reflects the overall performance of the stock market in this sector by calculating the weighted average price index of these sample stocks. The index's constituent stocks include companies in the fields of power generation, power grids, electrical equipment,



new energy, energy conservation, and environmental protection, reflecting the development of China's green power industry and investment opportunities, as shown in Fig. 2.

This index covers a wide range of industries with strong green attributes. In terms of industry distribution, the CSI Green Power Index has weightings of 34.11%, 26.12%, and 15.22% for thermal, hydro, and wind power generation, respectively, or over 70% in total, as shown in Fig. 3. Among them, the number of constituents involved in thermal power generation reached 17, whereas the number of constituents in hydropower generation, wind power generation, and photovoltaic power generation were 10, 8, and 7, respectively.



Fig. 4 Share of energy consumption structure

Looking at the data for the five years from 2017 to 2021, we can see that in China's energy consumption structure, the consumption of coal energy has been decreasing (as shown in Fig. 4), while the consumption of new energy is gradually increasing, indicating that China is paying increasingly more attention to the development of the new energy industry, and the CSI Green Power 50 Index selected for this study can show the development of the new energy industry to a large extent, so the study of this index is very meaningful.

Among the indicators observed, closing prices are a very important indicator. Dow's Theory states that the closing price is the most important of all prices and that the highest, lowest, and other prices represent short-term prices and should be more convincing in terms of their impact on the future. In many cases, the closing price is used as a proxy for the daily prices. The closing price not only reflects the day's trading but also serves as a reference for the opening price of the stock market the following day. It is a very important and observational indicator in the overall trading process, and its high and low prices are often used by market investors as a guide to which stocks to buy, especially in the eyes of short-term investors as a very important counting indicator. Therefore, we use the closing price of a stock to forecast its price. Figure 5 shows the closing prices from May 9 to October 24, 2022.

# Impact of policies on the stock market

The practice of stock market development worldwide has shown that the implementation of stock market policies can have a direct impact on stock market price fluctuations. However, different markets have different degrees of sensitivity and response to policies, which is the so-called policy effect of the stock market (Aktürk et al. 2022; Jang 2021). In the short history of China's stock market, the entire development process has been guided and promoted by the government, except for a certain degree of private spontaneity in the early stages of the stock market's development. The government has not



only used legal and economic means to regulate the operation of the stock market but has also frequently intervened in the stock market with policies, giving the stock market a strong policy market character. On the one hand, the authorities, with the mentality of "parental officials," hope to maintain the stable development of the market through various regulatory measures so that the stock market can play its expected role in economic reform and development; on the other hand, as an emerging stock market, the institutional basis and action structure of the Chinese stock market is still in a process of continuous improvement, and corresponding policy changes are therefore more frequent. In this environment, the stock market is often influenced by stock market policies to the extent that they become the dominant force in shaping the operation of the stock market to a large extent, generating large shocks to stock prices and triggering abnormal fluctuations in stock prices.

Many scholars have carried out policy analyses based on stock market trends as a sample and obtained a correlation between policy and trend (Abdelkafi 2018; Bekiros et al. 2016; Javaheri et al. 2022; Ko and Lee 2015; Sohangir et al. 2018). For example, Abdelkafi (2018), Javaheri et al. (2022), and Ko and Lee (2015) studied the Australian, US, and international stock markets, respectively. Trends in the stock market will lead to fluctuations in consumer sentiment, and the final result will exacerbate the uncertainty of the stock market (Sohangir et al. 2018), making the nonlinearity of stock prices more obvious and increasing the difficulty of prediction.

Disequilibrium is the natural state of an economy, which is always in a state of flux. This is not only because the economy is always exposed to external shocks or influences but also because disequilibria arise within the economy. Endogenous disequilibrium arises for two primary reasons. The first is fundamental uncertainty, but the authors believe that this is not the case. Disequilibrium is the natural state of an economy; therefore, it is always in a state of flux. This is not only because the economy is always exposed to external shocks or influences, but also because disequilibria arise within the economy. All questions about choice in the economy relate to future occurrences that may occur either immediately or later. Thus, the questions of choice in an economy must be related, to some degree, to the unknown.

During the transition from a planned to a market economy, almost all of the "institutional innovations" were led and nurtured by the government. When state-owned enterprises were not yet completely free from the infancy of the "separation of government and enterprises," the process of going public in the form of a shareholding system emerged, thus giving rise to several historical legacy problems unique to China. These "uncertainties" constitute a special type of exogenous market uncertainty. The lack of continuity and stability of stock market regulatory policies and securities regulators are constantly developing new approaches, policies, and measures. Market regulation is replaced by a large number of volatile policy measures, making it difficult for investors to form stable policy expectations that are extremely sensitive to policy reactions, thus increasing the impact of policy factors on the stock market and causing abnormal stock price fluctuations. Recently, China has increased its efforts in environmental protection and introduced a variety of policies to promote the development of green finance, which will also inevitably produce greater volatility in the share prices of the new energy industry, and as the policies to promote the new energy industry are also in the exploration stage, it will increase the volatility of the stock market in this industry and make it more difficult to forecast the new energy share price index.

## Sample complexity analysis

This study uses the Lyapunov index to represent the numerical characteristics of the average exponential dispersion rate of neighboring trajectories in phase space. This exponent is one of several numerical features used to identify chaotic motion.

The Lyapunov exponent is often used to determine the chaotic nature of a system, and an image can be used to visualize whether the system or mapping is chaotic. When lambda > 0, the system moves into a chaotic state, and the corresponding mapping is called a chaotic mapping. When lambda < 0, the motion of the system stabilizes and is insensitive to the initial state of the system, that is, the mapping is insensitive to the initial value at this point. Finally, when lambda = 0, the system is in a steady state.

By calculating the CSI Green Power 50 Index selected for this study, the lambda for these data is 1.5853.

This sample is characterized by chaos, which is by nature a nonlinear deterministic state of disorder and is characterized by a high degree of initial value sensitivity, so that a small perturbation causes the system to deviate completely from its original state after a sufficiently long period.

The Poincaré surface of the section was introduced by Poincaré at the end of the nineteenth century to analyze the motion of multivariate autonomous systems. Considering only the steady-state image of the Poincaré section, when there is only one immobile point and a few discrete points on the Poincaré section, the motion is considered periodic; when the Poincaré section is a closed curve, the motion is considered quasi-periodic; and when the Poincaré section is a patchwork of dense points with a hierarchical structure, the motion is considered to be in a chaotic state. Figure 6 shows that the sample data are in a chaotic state.

The calculation of the maximum amplitude is as indicated in Eq. (23):

$$AOV = \max\{aov_1, aov_2, aov_3 \dots\},\tag{23}$$

where *AOV* represents the maximum amplitude, and  $aov_i$  (i = 1, 2, 3, ...) represents the adjacent fluctuation value within the period.

The rising and falling volatilities are calculated using Eq. (24):

$$RV = \left(\frac{D_{i+1} - D_i}{DT_i}\right),\tag{24}$$

where RV represents rising volatility,  $D_i$  represents the *i*th bottom, and  $DT_i$  represents the distance between the two bottoms. Equation (25) is as follows:

$$DV = \left(\frac{U_{i+1} - U_i}{UT_i}\right),\tag{25}$$

where DV represents the falling volatility,  $U_i$  represents the *i*th top, and  $UT_i$  represents the time distance between the two tops.

The formula for calculating the average fluctuation interval is given by Eq. (26):

$$MFI = \frac{VP}{IL},\tag{26}$$



**Fig. 6** The Pomegranate interface

where *MFI* is the maximum fluctuation interval, *VP* is the number of fluctuation points in the forecast area, and *IL* is the interval length.

These four indicators are used to examine the complexity of the closing price movement of the CSI Green Power 50 Index, as shown in Table 1.

The volatility and amplitude of the raw closing price of the CSI Green Power 50 Index are so strong that direct analysis of the raw data would result in a large error in the forecast results.

The upper and lower limits of historical load stability can be expressed by Eqs. (27) and (28), respectively:

$$L_{upper} = \left[1 - \sqrt{\frac{\sum_{t=1}^{N} \left(\frac{|H(t)|}{P(t)}\right)^2}{N}}\right] \times 100\%,\tag{27}$$

$$L_{lower} = \left[1 - \sqrt{\frac{\sum_{t=1}^{N} \left(\frac{|L(t) + H(t)|}{P(t)}\right)^{2}}{N}}\right] \times 100\%,$$
(28)

where P(t) is the historical data sequence of residential electricity consumption, H(t) is the sequence of isolated high-frequency electricity consumption components, N is the number of electricity consumption data points in the analysis period,  $L_{upper}$  is the upper limit of the historical data stability, L(t) is a sequence of separated low-frequency electricity consumption components, and  $L_{lower}$  is the upper limit of the historical data stability.

The upper and lower limits of the closing price stability of the CSI Green Power 50 Index are 36.98% and 37.13%, respectively. The wide fluctuation range of the stability indicates that the original dataset is not stable, which reflects the strong uncertainty of the stock price index. Therefore, a decomposition analysis of the raw data is necessary.

After collation using EWT decomposition, data are divided into five groups, as shown in Fig. 7.

The stock price index is a weather vane for economic change, and movements of the stock price index are closely related to the state of the national economy. The main factors influencing the stock price index are the gross domestic product, deposit rates, exchange rates, and disposable income of the population. The main macro factors are exchange rates, corporate commodity price indicators, interest rates, macroeconomic indices, and consumer information indices. Empirical wavelet decomposition simplifies forecasting by decomposing the closing price of a stock index, which is disturbed by multiple factors, into a representation in the form of a combination of more characteristic columns of data through the volatility characteristics of the data only.

	Max amplitude	<b>Rising volatility</b>	Falling volatility	Max fluctuation interva
CSI Green Power 50	87.2	20.84	23.6	1.93

# Table 1 Data characterization



# **Forecast preparation**

As China continues to promote the development of green finance, the new energy industry has also increased its services to the real economy; however, the development of the new energy industry has also faced many tests, especially in the performance of the new energy stock price index is particularly prominent. First, in this study, because the background of energy stocks is highly volatile, the EWT algorithm is selected to decompose the original sequence to simplify the difficulty of volatility in prediction. Second, the SVR model can target the sequence of nonlinear factors modeled for forecasting, while the ALO optimization algorithm is used to enhance the forecasting effect and generalization ability. The overall modeling and forecasting processes of the model proposed in this study are shown in Fig. 8.

# Parameter setting of the model

The parameters of the model can have a serious impact on the effectiveness of the model prediction, and all the prediction model parameters employed in this study are presented in Table 2.

# Setting the parameters of the Ant Lion Optimization algorithm

# (1) Data processing

The two data subsets of the stock price index are the training and test sets. Data from the first 69 samples are selected as the training set, and data from the last 45 samples are selected as the test set.



Fig. 8 Structure of stock price forecasts

Table 2	Parameter	settings	of the	compared	models
---------	-----------	----------	--------	----------	--------

Models	Parameter settings
GRNN	MSE=0.001 to start the evolution, and the maximum hidden layer is 10
LSTM	250 epochs of training, Gradient t, Threshold = 1, Learning Rate = 0.005, hidden units = 200
RNN	Lag = 1:1; horizon = 50; MiniBatchSize = 64; MaxEpochs = 600; learningrate = 0.00611
SVRPSO	eps = 1e-3; v = 5; C1 = 1.5; C2 = 1.5;
SVRALO	Agents_no = 10; dim = 10; Max_iteration = 100; lb = $0.01$ , ub = 100
CNN	Lag = 1:1; horizon = 50; MiniBatchSize = 64; MaxEpochs = 600; learningrate = 0.00611

# (2) Control ALOSVR parameter settings

Set the model parameters: In the method proposed in this study, where the number of ants and ant lions is set to 10, the number of variables is 10, the maximum number of iterations is 100, the lower bound is 0.01, and the upper bound is 100:

# (3) Set the fitness function

The mean squared error (MSE) between the actual and predicted capacity values is used to establish the fitness function, as shown in Eq. 29:

$$MSE = \frac{1}{n} \sum_{i}^{n} (\hat{y} - y_i)^2$$
(29)

(4) ALO algorithm to optimize SVR parameters

After initializing the parameters and selecting the fitness function, the SVR parameters can be optimized using the ALO algorithm.

(5) Predicting the stock index

Test data are used to validate the existing model and predict the closing price of the stock index. The SVRALO method is adopted to predict each component of decomposition separately, and the predicted results are shown in Fig. 9

Simultaneously, from the model proposed in this study to predict the closing price logarithmic yield of the CSI Green Power 50 Index, as shown in Fig. 10, the prediction accuracy is also very high, and the scope of application of the model is relatively wide. However, because the logarithmic yield is calculated based on the closing price, this study focuses on the prediction and analysis of the closing price of the green energy stock index, supplemented by the logarithmic return on the stock index to carry out a complementary analysis to strengthen the reliability of the research in this paper and the applicability of the proposed model is broad:

$$R = Ln\left(\frac{P_t}{P_{t-1}}\right),\tag{30}$$

When the Ant Lion is optimized, the ants wander randomly through the search space with a high degree of uncertainty and randomness, and the stock price index is also



ple ones.



highly uncertain and random because of a variety of factors. Random wandering is influenced by the ant lion's traps. By observing the ant lion's habits, it is interesting to note that the pits that the ant lion digs when hunting are related to its hunger level and the Moon phase. For example, when the ant lion is more hungry, or when the Moon is fuller, it will usually choose to dig a larger pit. They have evolved and adapted this approach to improve their chances of survival. Whereas the goodness of fit of the SVR model is largely determined by the parameters, and the range of parameters selected is so large that it is difficult to find the optimal parameters, ALO determines the optimal parameter values by elastically selecting and constantly replacing the optimal Ant Lion from multi-

The ALO algorithm for the SVR algorithm focuses on optimizing the values of the penalty factors C and  $\gamma$  in the SVR algorithm. C is an important parameter of the SVM which controls the strength of the penalty of the model. By adjusting the value of C, we balance the tolerance of the model to classification errors and its control over overfitting. When the value of C is small, the SVR model is more forgiving of misclassified points and may lead to overfitting; when C is large, the SVR model is less forgiving of misclassified points and may lead to underfitting. Therefore, in practical applications, the optimal value of C is determined using methods such as cross-validation, which can improve the generalization ability and prediction effect of the model.

The value of  $\gamma$  determines the distribution of the data when mapped to the new feature space, with smaller support vectors for high values of  $\gamma$  and vice versa. The number of support vectors affects the speed of model training and prediction (Aktürk et al. 2022). In general, SVR can accommodate all data with nonlinear characteristics, but it tends to

suffer from over-tuning results, and the possibility of low detection accuracy is called overtraining; if the  $\gamma$  value is too small, it may have a significant smoothing effect on the selected model, leading to the particularly high accuracy of the driver but affecting the accuracy of the test device.

Choosing the right parameters is key to optimizing SVR, but the wide range of values for the two SVR parameters makes it difficult to construct a suitable model; even if a suitable model is constructed, it does not indicate an optimal model. The ALO algorithm uses a roulette wheel to randomly select ants to ensure that the exploration of space can be performed completely. The random walk of ants around the colony also emphasizes exploration of the search space around the colony. Calculating the random walks for each ant and each dimension results in different movement behaviors of the ants within the ant lion trap, allowing for a diversity of locations around the ant lion to be maintained. The Ant Lion provides a large space that does not limit the range, and the ALO algorithm uses a large number of search agents to approximate the global optimum; thus, avoiding local optima is very high. In addition, during the iterations, the intensity of the ant movement is reduced adaptively, ensuring convergence of the ALO algorithm. The parameters for predicting the closing price of the CSI Green Power 50 Index obtained using ALO are listed in Table 3.

	IMF1	IMF2	IMF3	IMF4	IMF5
С	53.96	68.99	37.45	75.66	38.43
γ	0.26	0.08	0.17	31.15	1.86

Table 3 Optimization parameters of EWT-S-ALOSVR model for each EWT and residual



In addition to the EWT-S-ALOSVR model proposed in this study, after decomposing the original data in the EWT way, PSO-optimized SVR, LSTM, RNN, CNN, GRNN, RNN, CNNRNN, and the financial time series model ARMA-GARCH are selected for comparison, and the prediction results are shown in Fig. 11.

# Forecast accuracy indicators

The prediction accuracy of the model is compared with those of other models using four known indicators of prediction accuracy.

The first is the mean squared error (MSE) given by Eq. (29) and the second is the mean absolute percentage error (MAPE) (Safari et al. 2023) given by Eq. (30); the third is the mean absolute error (MAE) given by Eq. (31), and the last is the root mean squared error (RMSE), given by Eq. (32). The associated values of the accuracy indicators for each model are listed in Table 4.

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - f_i}{y_i} \right| \times 100\%,$$
(31)

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - f_i|,$$
(32)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - f_i)^2},$$
(33)

where *N* is the total number of forecast results,  $y_i$  is the actual load at point *i*, and  $f_i$  is the forecast load at point *i*.

The analysis index we selected, r-square value, is a statistical index used to measure the strength of the relationship between two variables, indicating the degree of linear correlation between the variables, with a value range of 01. The closer the r-squared value is to 1, the stronger the linear relationship between the variables. The closer the R-squared

Table 4	Model's	stock index	closing	price	prediction	accuracy	<sup>,</sup> metrics

Forecasting models	R <sup>2</sup>	MAE	MSE	RMSE	MAPE	TIME
EWT-S-ALOSVR	0.9999	0.9938	1.4523	1.2051	0.05%	39.0849
EWT-GRNN	0.8032	38.3487	2206.0237	46.9683	1.86%	43.5990
EWT-SVR	0.8313	36.6347	1891.1719	43.4876	1.75%	31.9757
ARMA(1,2)-GARCH(1,1)	0.9234	24.3500	858.9023	29.3070	1.15%	23.4357
EWT-RNN	0.9031	27.3415	1085.9663	32.9540	1.30%	166.5626
EWT-CNNRNN	0.9038	27.0270	1078.4001	32.8390	1.29%	202.4555
EWT-LSTM	0.8529	35.0267	1648.8686	40.6063	1.69%	223.8612
EWT-CNN	0.9037	27.4515	1079.1984	32.8512	1.32%	155.4874
EWT-PSOSVR	0.8152	38.6043	2071.8264	45.5173	1.86%	380.8867

The bold words imply that the model is the best one among other compared models

value is to 0, the weaker or nonexistent the linear relationship between the variables. In addition, the MSE in mathematical statistics refers to the expected value of the square difference between the parameter estimate and parameter true value. MSE is a convenient method to measure the "mean error," MSE can evaluate the degree of change of the data, the smaller the value of MSE, the better the accuracy of the prediction model to describe the experimental data. In machine learning, to examine the accuracy of a model, the mathematical average is used to calculate the root mean square of the error that occurs between the tested and predicted values. The RMSE adds a square root to the MSE. The average absolute error is the average of the absolute values of deviations from the arithmetic mean of all individual observations. The average absolute error can avoid the problem in which errors cancel each other; thus, it can accurately reflect the size of the actual prediction error.

The closing price logarithmic yield prediction results for the CSI Green Power 50 Index are presented in Table 5. Given that the yield data processing does not meet the requirements of ARMA-GARCH prediction, only the remaining eight groups of data prediction indices are analyzed for comparison. It can be seen that the prediction effect of EWT-S-ALOSVR, the method proposed in this study, is more precise.

Common optimization methods have many limitations; for example, the grey wolf optimization algorithm is prone to premature convergence, its convergence accuracy is not high when facing complex problems, and its convergence speed is not sufficient. Moreover, in the process of evolution, super bats in a population may attract other individuals to gather around them rapidly, which dramatically decreases the diversity of the population. Simultaneously, as bats get closer to the optimal individuals of the population, the rate of convergence is greatly reduced, or there is evolutionary stagnation, and the population loses the ability to evolve further. In many cases, especially for an optimization space with high-dimensional, multi-peak, and complex terrain characteristics, the algorithm does not converge to global extremes. Thus, it is difficult to find the global optimal point distributed in the local optimal neighborhood. Therefore, the basic bat optimization algorithm should be improved to increase the diversity of the population so that the population can maintain the ability of continuous optimization in the iterative process. The BA suffers from the shortcomings of slow convergence speed, low convergence accuracy, and ease of falling into local minima, which seriously limits its application. The most common PSO algorithm is the focus of this study.

	R <sup>2</sup>	MAE	MSE	RMSE	MAPE
EWT-SVR	0.676475651	40.7906966	2254.522143	47.48180855	1.9850%
EWT-PSOSVR	0.580166447	50.02146896	2925.665553	54.08942182	2.4404%
EWT-S-ALOSVR	0.999876065	0.753951482	0.863660746	0.929333496	0.0365%
EWT-GRNN	0.562345593	45.8269164	3049.85253	55.22546994	2.2345%
EWT-CNN	0.766315883	34.24874618	1628.45863	40.35416497	1.6741%
EWT-LSTM	0.751358377	34.20940481	1732.691985	41.62561693	1.6556%
EWT-RNN	0.711200442	39.1601227	2012.537853	44.86131801	1.9055%
EWT-RNNCNN	0.903051284	21.09031406	675.599932	25.99230525	1.0221%

 Table 5
 Model's predictive accuracy metrics for log stock index closing returns

The bold words imply that the model is the best one among other compared models

Although PSO optimization algorithms have stochasticity and parallelism, such as gradient descent, PSO optimization algorithms have many advantages, including relatively simple algorithms, relatively fast convergence, applicability to nonlinear, nonconvex, and multipeaked functions, and the ability to handle high-dimensional problems. However, the PSO algorithm is prone to falling into local optimum solutions because each particle in the algorithm focuses only on local optimum solutions and is unable to discover solutions for the entire search space. This also leads to the possibility that the algorithm may miss the globally optimal solution. In addition, the implementation of the algorithm requires tuning of the parameters, which must be chosen appropriately to balance the exploration and exploitation of the particles. The two main reasons for this phenomenon are: first, the nature of the function to be optimized, and second, the rapid disappearance of the diversity of the particles in the algorithm, which leads to premature convergence. These two causes are intricately linked and difficult to analyze individually. Second, the PSO algorithm does not have a sophisticated search method for matching; therefore, the results obtained by the algorithm are often not the most accurate. Finally, the PSO algorithm is not based on rigorous theory but is simply a simplified simulation of a group search phenomenon that cannot be justified in terms of the mechanism of its occurrence, and the scope of its use is not clearly stated.

The improved approach of the ALO algorithm can be considered as a search for a global optimum that is more accurate than the initial setup. The strengths of the ALO algorithm lie primarily in the independence of the process and problem from each other, and in the reception of the input and output data. Thus, the problem only represents the operation of the focal step of its optimization algorithm, whereas the essence does not have a significant impact on the algorithm. Therefore, evolutionary algorithms do not need to derive global optima. Given that the ALO algorithms are highly stochastic, they can effectively address the interference of local optima on the resulting output errors. Another advantage of the ALO algorithm is its simplicity. Most bionic optimization algorithms are inspired by the evolution of nature or group behaviors, and are relatively simple. In addition, optimization algorithms have a general framework based on a set of randomly generated solutions reinforced by iterative updates.

The method proposed in this study shows a higher accuracy for peaks or inflection points. The reason for this is primarily that neural network algorithms have the disadvantages of neural network models when predicting, which include: (1) the most serious problem is the inability to explain their reasoning process and the basis of their reasoning; (2) the inability to ask the user the necessary queries, and when there is insufficient data, the neural network cannot work; (3) turning all problem features into numbers and all reasoning into numerical calculations inevitably results in lost information; and (4) the theory and learning algorithms need to be further refined and improved. RNN, LSTM, and their derivative machine-learning algorithms are based on their natural sequential processing over time. In addition, long periods of information traverse all the units sequentially as they enter the current processing unit. At this point, the gradient disappears. Although the RNN algorithm achieves a certain degree of resolution for the gradient problem compared to the LSTM algorithm and its variant hybrid algorithms,



Fig. 12 Comparison of forecasting at inflection points



Fig. 13 Comparison of fitting effects

the strength of the solution is inadequate. This problem has a serious impact on the accuracy and validity of the results when faced with long time series. In addition, the GRNN limitations are primarily due to the high computational complexity of the prediction process, in which all test samples must be computed against all training samples. Therefore, it is often necessary to retain every training sample. To better observe the fit of the model predictions, the above graph selects models with R<sup>2</sup> above 90% and plots the effect of zooming in to compare the inflection point predictions, as shown in Fig. 12.

To compare the indicators of the nine forecasting methods used in this study, two groups of forecasting methods with the smallest MAPE and the original data are selected to graph the results, as shown in Fig. 13.

$$y_t = \vartheta_1 y_{t-1} + \dots + \vartheta_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$
(34)

$$\sigma_t^2 = a + \sum_{i=1}^p b_i y_{t-i}^2 + \sum_{j=1}^q c_j \sigma_{t-j}^2$$
(35)

All time series can be split into a sum of two parts: one composed of the mean equation and the other part is made up of the variance equation. When using the ARMA model alone for time series analysis and forecasting, the variance equation contained in the series is often missed, simply because the residuals are white noise series, and there is no longer any information that can be mined. By contrast, when using the GARCH model alone for time-series analysis, the mean equation is usually treated as a fixed constant, that is, in the formula, and therefore has an ARCH effect. The commonly mentioned ARMA-GARCH model applies the ARMA model to the mean for forecasting, while building a GARCH model for the variance. Thus, the mean is consistent with the ARMA prediction process, whereas the residuals satisfy the stochastic GARCH process.

The error bands for both EWT-S-ALOSVR and ARMA(1,2)-GARCH(1,1) are drawn as indicated in Fig. 14.

The error plot Fig. 14 shows that the error band of EWT-S-ALOSVR is narrower than the error band interval of ARMA(1,2)-GARCH(1,1), so it can be concluded that the proposed EWT-S-ALOSVR model is superior to other comparative models.

The error plot Fig. 15 shows that the error of EWT-S-ALOSVR has a narrower error band interval than that of ARMA(1,2)-GARCH(1,1), and therefore the proposed EWT-S-ALOSVR model is better.

The model has been used extensively in the financial sector with good results (Challa et al. 2020; Liu et al. 2021; Quaicoe et al. 2015). However, the ARMA-GARCH model is not perfect, and its limitations primarily lie in the fact that, first, the ARMA model requires a series to be stable when predicting, or the series must be stable after processing by differencing; otherwise, it cannot build a fixed-order model. Second, ARMA





Fig. 15 EWT-S-ALOSVR and ARMA(1,2)-GARCH(1,1) error bar chart

 Table 6
 Parameter settings for ALO forecasting of stock indices

	Maximum of iterations	Population size	Lower bound	Number of ant lions	Upper bound
ALO optimization parameters	20	20	- 100	30	100

models can only deal with linear data and cannot capture nonlinear relationships. Finally, a disadvantage of the GARCH model is that if the distribution of errors is not judged properly, it will reduce the accuracy of the results and make the established model unstable.

Stock data are chaotic, complex, formed by a variety of factors. ALO optimizes the SVR model with two parameters including the maximum number of iterations and the number of populations and so on. As shown in Table 6, the maximum number of iterations is to avoid the parameter optimization from falling into a cycle, and the number of populations can be designed to avoid local optimums in the process of stock prediction. The design of traps for ant lion foraging is also confirmed by the characteristics of each ant's action, and this process is similar to the impact of different factors on the stock price index to determine the optimal parameters in the prediction model; therefore, the use of the ALO model for the prediction of the stock index is reasonable and necessary.

# Robustness analysis of model effects

Robustness testing is a common method of statistical analysis, the main purpose of which is to assess the reliability and stability of research results. Testing the robustness of a research result, that is, whether varying a parameter or making a certain change to varying degrees, will have a substantial impact on the results of the research. If the results of a study remain stable with different parameters or changes, they can be considered robust.

To verify the robustness of the stock index prediction effect of the EWT-S-ALOSVR proposed in this study, we again divide the data test set and training set, and again predict the results of the model through the indicators, the prediction analysis comparison

	R <sup>2</sup>	MAE	MSE	RMSE	MAPE
EWT-SVR	0.322624675	0.009223196	0.000131018	0.011446298	45.2522%
EWT-PSOSVR	0.779402482	0.005332022	4.26679E-05	0.006532068	16.3844%
EWT-S-ALOSVR	0.999530122	0.000244515	9.08836E-08	0.000301469	0.1771%
EWT-GRNN	0.70494472	0.005991242	5.70695E-05	0.007554436	- 22.3424%
EWT-CNN	0.878829203	0.003869251	2.34368E-05	0.004841159	21.9221%
EWT-LSTM	0.840191417	0.004534262	3.09101E-05	0.005559688	11.7250%
EWT-RNN	- 0.029625582	0.011786614	0.00019915	0.014112048	- 44.3026%
EWT-RNNCNN	0.669243668	0.006815047	6.39748E-05	0.007998425	- 162.3404%

# Table 7 Robustness testing

The bold words imply that the model is the best one among other compared models

Tal	ble	8	Predictive	analyses	for a sam	ole o'	f one y	year
-----	-----	---	------------	----------	-----------	--------	---------	------

	R2	MAE	MSE	RMSE	MAPE
EWT-SVR	0.720473103	15.6034366	377.5051329	19.42949132	0.7882%
EWT-PSOSVR	0.79298626	12.58057204	279.5750622	16.72049826	0.6313%
EWT-S-ALOSVR	0.998958106	0.932248487	1.407093617	1.18620977	0.0470%
EWT-GRNN	0.650209266	16.46463634	472.3974652	21.73470647	0.8255%
EWT-CNN	0.907873091	9.173742964	124.4187278	11.15431431	0.4611%
EWT-LSTM	0.803847794	13.06862743	264.9064019	16.2759455	0.6578%
EWT-RNN	0.913535694	9.249668318	116.771301	10.80607704	0.4655%
EWT-RNNCNN	0.809604256	12.3327014	257.132216	16.03534272	0.6215%

The bold words imply that the model is the best one among other compared models

in the above paper is to divide 60% of the data into the test set and 40% into the training set; therefore, the data is again re-divided: 70% of the data is divided into the test set, and 30% is divided into the training set, and the results of the prediction indicator comparison are as shown in Table 7.

By comparing the EWT-S-ALOSVR model proposed in this study, the prediction effect is still better than the other models in the case of data division change, so it can be seen that the model proposed in this study is robust and the conclusions of the study are accurate and reliable.

The CSI Green Power 50 Index data are selected as a supplementary analysis to compare the accuracy of the model proposed in this study, as there are more than 1 year of sample data, and the results are presented in Table 8.

## Significance tests for predictive performance

To validate the improvement in predictive performance provided by the proposed EWT-S-ALOSVR prediction model, a comparison with other models must be made, and the statistical significance of any differences must be determined. The comparison is based on a one-to-one rule (two-by-two comparison) so that only one model at a time is compared with the proposed ALOSVR model. As each model is independent of the other, the comparison can simply be made as described above. Furthermore, multiple comparisons are required to ensure that the improvements provided by the proposed model are significant compared with other models. Recently, Derrac et al. (2011) came to the structural conclusion that the Wilcoxon signed-rank test can be used to make simple two-by-two comparisons, while the Friedman test can be used to make multiple comparisons.

The Wilcoxon signed-rank test is used to determine the significance of the prediction error when two forecasting models with the same amount of data forecast the central tendency. The definition  $e_i$  is the absolute prediction error of the *i*th forecast outcome produced by the two forecasting models. If  $e_i > 0$ ,  $r^+$  is set as the rank sum; if  $e_i < 0$ ,  $r^-$  is set as the rank sum, and if  $e_i = 0$ , this comparison is eliminated, and the sample size is removed. The statistic W is defined in Eq. (35):

$$W = \min\{r^+, r^-\} \tag{36}$$

The Friedman test is a nonparametric statistical procedure used to detect significant prediction errors between two or more prediction models. The null hypothesis of the Friedman test is that the means of the prediction errors associated with all models of interest are equal. The statistic F in the Friedman test is given by Eq. (36):

$$F = \frac{12N}{m(m+1)} \left[ \sum_{j=1}^{m} R_j^2 - \frac{m(m+1)^2}{4} \right],$$
(37)

where N is the total number of prediction errors; m is the number of models compared; and Rj is the average rank sum obtained for each prediction error of each prediction model, as defined in Eq. (37):

$$R_{j} = \frac{1}{N} \sum_{i=1}^{N} s_{i}^{j},$$
(38)

where is the *i*th prediction error of the *j*th comparison model.

The Wilkerson signed-rank test and the Friedman test are used to determine the significance of the proposed EWT-S-ALOSVR model. The results of these two statistical tests are given in Table 5, with a one-tailed test used at the  $\alpha = 0.05$  level of significance. This indicates that the model performs significantly better than the other models.

Compared models	Wilcoxon α = 0.05 signed-rank test; <i>p</i> value	Friedman test		
EWT-S-ALOSVR versus EWT-GRNN	< 0.01	H0: $e1 = e2 = e3 = e4 = e5 = e6 = e7 = e8$		
EWT-S-ALOSVR versus EWT-LSTM	< 0.01	= e9 F = 284.997 p < 0.01 (Reject H0)		
EWT-S-ALOSVR versus EWT-RF	< 0.01			
EWT-S-ALOSVR versus EWT-PSORF	< 0.01			
EWT-S-ALOSVR versus EWT-PSOSVR	< 0.01			
EWT-S-ALOSVR versus EWT-SVR	< 0.01			
EWT-S-ALOSVR versus EWT-CNNRNN	< 0.01			
EWT-S-ALOSVR versus EWT-RNN	< 0.01			
EWT-S-ALOSVR versus ARMA(1,2)- GARCH(1,1)	< 0.01			

Table 9	Signed	rank te	est and	Friedm	nan's	test
---------	--------	---------	---------	--------	-------	------

Compared models	One-sided KSPA test	Two-sided KSPA test	
	< 0.01	< 0.01	
EWT-S-ALOSVR versus EWT-LSTM	< 0.01	< 0.01	
EWT-S-ALOSVR versus EWT-RF	< 0.01	< 0.01	
EWT-S-ALOSVR versus EWT-PSORF	< 0.01	< 0.01	
EWT-S-ALOSVR versus EWT-PSOSVR	< 0.01	< 0.01	
EWT-S-ALOSVR versus EWT-SVR	< 0.01	< 0.01	
EWT-S-ALOSVR versus EWT-CNNRNN	< 0.01	< 0.01	
EWT-S-ALOSVR versus EWT-RNN	< 0.01	< 0.01	
EWT-S-ALOSVR versus ARMA(1,2)-GARCH(1,1)	< 0.01	< 0.01	

## Table 10 Forecasting results of the KSPA test



The Wilcoxon signed-rank test and Friedman test are used to compare the results of the proposed EWT-S-ALOSVR model, as listed in Table 9.

The KSPA test is a complementary statistical test used to determine the accuracy of the two sets of predictions. It is a nonparametric test based on the Kolmogorov——Smirnov test. The advantage of the KSPA test is that it determines not only the predictive distribution of the two models but also whether the model has a minimal random error. The test is not affected by autocorrelation in the prediction errors.

A two-sample bilateral KSPA test (hereafter referred to as the bilateral KSPA test) is used to determine whether there is a statistically significant difference between the two prediction error distributions. The null hypothesis is that there is no statistically significant difference between the two statistical predictions; when the bilateral KSPA test produces a test statistic below the significance level (typically 1%, 5%, or 10%), the null hypothesis is rejected, and the alternative hypothesis of the distribution of prediction errors per unit area is accepted. In this case, there is a statistically significant difference, there is a statistically significant difference between the two predictions tested in the bilateral KSPA test. The purpose of the two-sample one-sided KSPA test (hereafter, one-sided KSPA test) is to determine whether a model based on a loss function that minimizes



the reported error provides a smaller random error than the predicted model. The results of the model KSPA test are presented in Table 10 and Figs. 16 and 17.

First, statistically significant differences between the proposed prediction models and the eight comparison models are confirmed. A one-sided KSPA test is used to identify the proposed models and to compare the low random errors reported for the predictions. The results indicate that the EWT-S-ALOSVR model has the greatest predictive performance. The forecasts obtained using the EWT-S-ALOSVR model outperform those of the comparative model based on the computed error statistics. Therefore, there is a significant difference between the proposed and comparative models. The KSPA error distribution and empirical cumulative distribution function are shown in the figure. The proposed EWT-SVRALO model better describes the random deviation, resulting in smaller errors and high prediction accuracy.

In short, it is clear from the tests that the proposed model has better prediction accuracy than the comparison model. It captures random deviations with fewer errors.

# **Conclusion and prospects**

## Conclusion

Recently, techniques and theories on deep learning have advanced, particularly the powerful learning ability of machine learning and other advantages that are widely used in stock price prediction, which is often more accurate and effective than traditional stock price prediction. With the country's promotion of sustainable development, support for the new energy industry has increased; thus, the healthy, stable, and sustainable development of the new energy industry has far-reaching consequences for the economy and society, and the stock price of the new energy industry can, to a certain extent, reflect its recent development of the new industry. Therefore, this study considers the new energy sector stock index as the research object. However, the price of new energy stocks is influenced not only by the development of the issuing company itself, but also by the general economic environment, domestic and international conditions, and socio-political factors. Consequently, there is usually no clear pattern for the movement of new energy stock prices; therefore, to accurately predict their fluctuations, a large amount of historical data are needed as a basis for training and building models. Some traditional methods, such as ARIMA and GARCH, need to first smooth the data, and then fix the order, optimize it, and perform several stability and cointegration tests; however, they are often more accurate in the expected short term, but less effective in the long term. Although some machine learning models are able to process long-term large data, often due to the amount of calculation, resulting in the calculation being too slow, which leads to its results also being less than ideal. In the case of small datasets, the model often chooses a simpler model.

A new method for forecasting stock price indices, EWT-S-ALOSVR, is proposed in this study to provide a new approach to stock price index forecasting, offering new models for use in other time-series forecasting problems. The findings of this study are as follows:

- 1. The new energy stock index exhibits nonlinearity and is severely affected by policy changes. By calculating the Lyapunov index and observing the Poincaré surface of the section, it can also be shown that the CSI Green Power 50 Index sample has chaotic characteristics and that the data have strong volatility and uncertainty. Therefore, accurately predicting new energy industry stock prices is difficult.
- 2. The EWT-S-ALOSVR model is more effective than the other models in predicting the closing price of the new energy stock price index. The results show that the MAE, MSE, RMSE, and MAPE predicted by the EWT-S-ALOSVR model are 0.9938%, 1.4523%, 1.2051%, and 0.05%, respectively. To test the accuracy of the model, the RNN, LSTM, GRNN, PSOSVR, and other models are also adopted for comparison, which shows that the R2 of the proposed model reaches the maximum.
- 3. The prediction results are tested using the Wilcoxon signed-rank, Friedman, and KSPA tests. Observing the error distribution of the models, it can be seen that the EWT-S-ALOSVR model in this study has smaller errors than the other models.

# Prospects

Although the ability of stock prices to be predicted is controversial, the feasibility of stock price forecasting has been recognized by many scholars, and the search for an excellent forecasting method is ongoing. Stock price series or stock price index series are essentially financial time series, which have many commonalities with other kinds of time series and also have their own unique and difficult characteristics. Therefore, this study has some shortcomings in modeling stock price series for forecasting.

1. Owing to hardware constraints and the relatively small amount of research on the model, the model can only be semi-manual or can rely on the experience of similar studies by previous scholars. For investors keen on short-term operations, short-term fluctuations significantly affect their operations. The model proposed in this study has better prediction results for the low and medium frequency and trend terms after several rounds of moderation; however, the prediction results for the high-frequency components are still not promising, even though the final prediction results appear to be good after integration. Therefore, in subsequent research, the

automatic model-tuning method can be extended to select the best parameters for improving the model effect.

2. The daily closing price of the stock index is predicted using only one feature. By contrast, the trading data indicators of the stock price index include the opening price, high price, low price, volume, and turnover. Daily trading data are empirically modeled and decomposed to quantify several influencing factors; however, these influencing factors are not specific to individual cases. Macroeconomic trends and the overall financial market conditions are closely related to the rise and fall of the stock market, and the algorithm can only derive this information from the data.

## Abbreviations

CSI	China Securities Index
SVR	Support vector machine
EWT	Empirical wavelet transform
ALO	Antlion algorithm
LSTM	Long-short term memory
CNN	Convolutional neural network
RNN	Recurrent neural network
GRNN	Generalized regression neural network
PSO	Particle swarm optimization
BA	Bat algorithm
CS	Cuckoo search algorithm
FOA	Fruit fly optimization algorithm
GOA	Grasshopper optimization algorithm
MFL	Max fluctuation interval
DV	Falling volatility
RV	Rising volatility
AOV	Max amplitude
MAE	Mean absolute error
MSE	Mean squared error
RMSE	The root mean squared error
MAPE	Mean absolute percentage error
KS	Kolmogorov-Smirnov

#### Acknowledgements

Guo-Feng Fan thanks the support from the project grants: Key Research Project in Universities of Henan Province (No. 24B480012), Science and Technology of Henan Province of China (No. 182400410419), and the Foundation for Fostering the National Foundation of Pingdingshan University (No. PXY-PYJJ-2016006), and Wei-Chiang Hong thanks National Science and Technology Council, Taiwan (MOST 111-2410-H-161-001).

## Author contributions

G-FF: Investigation, Methodology, Validation. Funding acquisition, Supervision, Writing-Original draft preparation. R-TZ: Conceptualization, Investigation, Methodology, Software, Data curation, Formal analysis, Validation. C-CC: Conceptualization, Investigation, Software, Data curation, Formal analysis. L-LP: Investigation, Methodology, Software, Data curation, Funding management. Y-HY: Software, Data curation, Formal analysis, Validation. W-CH: Investigation, Methodology, Funding acquisition, Supervision, Writing-Reviewing, and Editing. All authors read and approved the final manuscript.

## Funding

Key Research Project in Universities of Henan Province (No. 24B480012), Science and Technology of Henan Province of China (No. 182400410419), the Foundation for Fostering the National Foundation of Pingdingshan University (No. PXY-PYJJ-2016006), and National Science and Technology Council, Taiwan (MOST 111-2410-H-161-001).

## Availability of data and materials

Data is available if reasonable request.

## Declarations

#### **Competing interests**

The authors declare that they have no competing interests.

Received: 5 June 2023 Accepted: 3 January 2024 Published online: 22 February 2024

## References

- Abdelkafi I (2018) The relationship between public debt, economic growth, and monetary policy: empirical evidence from Tunisia. J Knowl Econ 9:1154–1167. https://doi.org/10.1007/s13132-016-0404-6
- Ahamad N, Sikander A, Singh G (2022) A novel reduction approach for linear system approximation. Circuits Syst Signal Process 41:700–724. https://doi.org/10.1007/s00034-021-01816-4
- Aktürk E, Karan MB, Pirgaip B (2022) Is the effect of dividend policy on the volatility of stock prices stable? An empirical study on European countries. Span J Finance Acc 51(4):484–504. https://doi.org/10.1080/02102412.2022. 2027647
- Ansal V (2020) ALO-optimized artificial neural network-controlled dynamic voltage restorer for compensation of voltage issues in distribution system. Soft Comput 24:1171–1184. https://doi.org/10.1007/s00500-019-03952-1
- Ba Z, Zhao Y, Liu X, Gang Li G (2022) Spatio-temporal dynamics and determinants of new energy policy diffusion in China: a policy citation approach. J Clean Prod 376:134270. https://doi.org/10.1016/j.jclepro.2022.134270
- Bai L, Liu Y, Wang Q, Chen C (2019) Improving portfolio performance of renewable energy stocks using robust portfolio approach: evidence from China. Physica A 533:122059. https://doi.org/10.1016/j.physa.2019.122059
- Barman M, Choudhury NBD (2018) Hybrid GOA-SVR technique for short term load forecasting during periods with substantial weather changes in North-East India. Procedia Comput Sci 143:124–132. https://doi.org/10.1016/j.procs. 2018.10.360
- Bekiros S, Gupta R, Kyei C (2016) On economic uncertainty, stock market predictability and nonlinear spillover effects. N Am J Econ Finance 36:184–191. https://doi.org/10.1016/j.najef.2016.01.003
- Challa ML, Malepati V, Kolusu SNR (2020) S&P BSE Sensex and S&P BSE IT return forecasting using ARIMA. Financ Innov 6:47. https://doi.org/10.1186/s40854-020-00201-5
- Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol Comput 1:3–18. https://doi.org/ 10.1016/j.swevo.2011.02.002
- Elsayed AH, Nasreen S, Tiwari AK (2020) Time-varying co-movements between energy market and global financial markets: Implication for portfolio diversification and hedging strategies. Energy Econ 90:104847. https://doi.org/10. 1016/j.eneco.2020.104847
- Gilles J (2013) Empirical wavelet transform. IEEE Trans Signal Process 61(16):3999–4010. https://doi.org/10.1109/TSP.2013. 2265222
- Gu Q, Chang Y, Li X, Chang Z, Feng Z (2021) A novel F-SVM based on FOA for improving SVM performance. Expert Syst Appl 165:113713. https://doi.org/10.1016/j.eswa.2020.113713
- Guo CZ (2023) Fully implement the spirit of the 20th National Congress of the Communist Party of China and jointly promote the new development of the environmental protection industry. China's Environ Protect Ind 1:7–9 (**in Chinese**)
- Hao Y, Wang Q, Li Y (2018) An intelligent algorithm for fault location on VSC-HVDC system. Int J Electr Power Energy Syst 94:116–123. https://doi.org/10.1016/j.ijepes.2017.06.030
- Jang WW (2021) Monetary policy effects on equity returns: application of SVAR identified with high-frequency external instrument. J Deriv Quant Stud 29(4):319–331. https://doi.org/10.1108/JDQS-08-2021-0021
- Javaheri B, Habibi F, Amani R (2022) Economic policy uncertainty and the US stock market trading: non-ARDL evidence. Future Bus J 8:36. https://doi.org/10.1186/s43093-022-00150-8
- Jiang M, Chen W, Huilin Xu, Liu Y (2023) A novel interval dual convolutional neural network method for interval-valued stock price prediction. Pattern Recogn 145:109920. https://doi.org/10.1016/j.patcog.2023.109920
- Kim K (2003) Financial time series forecasting using support vector machines. Neurocomputing 55:307–319. https://doi. org/10.1016/S0925-2312(03)00372-2
- Ko JH, Lee CM (2015) International economic policy uncertainty and stock prices: wavelet approach. Econ Lett 134:118– 122. https://doi.org/10.1016/j.econlet.2015.07.012
- Lee MC (2009) Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Syst Appl 36(8):10896–10904. https://doi.org/10.1016/j.eswa.2009.02.038
- Lei L (2018) Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. Appl Soft Comput 62:923–932. https://doi.org/10.1016/j.asoc.2017.09.029
- Li MD, Li JW (2023) The development status and prospects of distributed photovoltaic power generation in China under the "Dual Carbon" goal. Sol Energy 5:5–10 (in Chinese)
- Liu J, Zhang Z, Yan L, Wen F (2021) Forecasting the volatility of EUA futures with economic policy uncertainty using the GARCH-MIDAS model. Financ Innov 7:76. https://doi.org/10.1186/s40854-021-00292-8
- Lou Q, Wan X, Jia B, Song D, Qiu L, Yin S (2022) Application study of empirical wavelet transform in time–frequency analysis of electromagnetic radiation induced by rock fracture. Minerals 12(10):1307–1330. https://doi.org/10.3390/min12101307
- Ma L, Wang G, Zhang P, Huo Y (2022) Fault diagnosis method of circuit breaker based on CEEMDAN and PSO-GSA-SVM. IEEJ Trans Electr Electron Eng 17(11):1598–1605. https://doi.org/10.1002/TEE.23666
- Mirjalili S (2015) The ant lion optimizer. Adv Eng Softw 83:80–98. https://doi.org/10.1016/j.advengsoft.2015.01.010
- Ni LP, Ni ZW, Gao YZ (2011) Stock trend prediction based on fractal feature selection and support vector machine. Expert Syst Appl 38(5):5569–5576. https://doi.org/10.1016/j.eswa.2010.10.079
- Nie D, Li Y, Li X (2021) Dynamic spillovers and asymmetric spillover effect between the carbon emission trading market, fossil energy market, and new energy stock market in China. Energies 14:6438–6460. https://doi.org/10.3390/en141 96438
- Ning SC, Cao CJ, Wang LL, Xiao J, Zhao QF (2022) The prediction model for transverse thickness difference of electric steel in 6-high cold rolling mills based on GA-PSO-SVR approach. Steel Res Int 93(11):2200302. https://doi.org/10.1002/ SRIN.202200302
- Pan WT, Liu Y, Jiang H, Chen YT, Liu T, Qing Y, Huang GH, Li R (2021) Model construction of enterprise financial early warning based on quantum FOA-SVR. Sci Program 2021:5018917. https://doi.org/10.1155/2021/5018917

- Quaicoe MT, Twenefour FBK, Baah EM, Nortey ENN (2015) Modeling variations in the cedi/dollar exchange rate in Ghana: an autoregressive conditional heteroscedastic (ARCH) models. Springerplus 4:329. https://doi.org/10.1186/ s40064-015-1118-0
- Safari M, Rabiee AH, Joudaki J (2023) Developing a support vector regression (SVR) model for prediction of main and lateral bending angles in laser tube bending process. Materials 16(8):3251–3265. https://doi.org/10.3390/ma16083251
- Saleem F, Majeed MN, Iqbal J, Waheed A, Rauf A, Zareei M, Mohamed EM (2021) Ant lion optimizer based clustering algorithm for wireless body area networks in livestock industry. IEEE Access 9:114495–114513. https://doi.org/10. 1109/ACCESS.2021.3104643
- Sandubete JE, Beleña L, García-Villalobos JC (2023) Testing the efficient market hypothesis and the model-data paradox of chaos on top currencies from the foreign exchange market (FOREX). Mathematics 11(2):286–315. https://doi.org/ 10.3390/math11020286
- Shen J, Shafiq MO (2020) Short-term stock market price trend prediction using a comprehensive deep learning system. J Big Data 7:66. https://doi.org/10.1186/s40537-020-00333-6
- Sohangir S, Wang D, Pomeranets A, Khoshgoftaar TM (2018) Big Data: Deep Learning for financial sentiment analysis. J Big Data 5:3. https://doi.org/10.1186/s40537-017-0111-6
- Sun C, Ding D, Fang X, Zhang H, Li J (2019) How do fossil energy prices affect the stock prices of new energy companies? Evidence from Divisia energy price index in China's market. Energy 169:637–645. https://doi.org/10.1016/j.energy. 2018.12.032
- Sun Y, He D, Li J (2022) The PSO optimisation SVM prediction model for the asphalt pavement environment and service fatigue life. Int J Inf Commun Technol 20(4):355–366. https://doi.org/10.1504/JJICT.2022.123173
- Teng JL (2023) Rich Country's quantitative dream team launches a new "Base" Green Power 50ETF to keep up with the trend of the times. Invest Right Way 3:94–94 (**in Chinese**)
- Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
- Wang Y, Luo C (2021) An intelligent quantitative trading system based on intuitionistic-GRU fuzzy neural networks. Appl Soft Comput 108:107471. https://doi.org/10.1016/j.asoc.2021.107471
- Wang F, Yin H, Li S (2010) China's renewable energy policy: Commitments and challenges. Energy Policy 38(4):1872–1878. https://doi.org/10.1016/j.enpol.2009.11.065
- Wang CD, Chen Z, Lian Y, Min Chen M (2022) Asset selection based on high frequency Sharpe ratio. J Econometr 227(1):168–188. https://doi.org/10.1016/j.jeconom.2020.05.007
- Yang Y, Sun W, Su G (2022) A novel support-vector-machine-based grasshopper optimization algorithm for structural reliability analysis. Buildings 12(6):855–871. https://doi.org/10.3390/BUILDINGS12060855
- Zhang Q, Fang L (2015) Parameters optimization of SVM based on improved FOA and its application in fault diagnosis. J Softw 10(11):1301–1309. https://doi.org/10.17706/jsw.10.11.1301-1309
- Zhang P, Yang Y, Shi J, Zheng Y, Wang L, Li X (2009) Opportunities and challenges for renewable energy policy in China. Renew Sustain Energy Rev 13(2):439–449. https://doi.org/10.1016/j.rser.2007.11.005
- Zheng J, Wang Y, Li S, Chen H (2021) The stock index prediction based on SVR model with bat optimization algorithm. Algorithms 14(10):299–330. https://doi.org/10.3390/A14100299
- Zhou H, Huang S, Zhang P (2023) Prediction of jacking force using PSO-BPNN and PSO-SVR algorithm in curved pipe roof. Tunn Undergr Space Technol 138:105159. https://doi.org/10.1016/j.tust.2023.105159
- Zhu Y, Huang C, Wang Y, Wang J (2022) Application of bionic algorithm based on CS-SVR and BA-SVR in short-term traffic state prediction modeling of urban road. Int J Automot Technol 23(4):1141–1151. https://doi.org/10.1007/S12239-022-0100-4

# **Publisher's Note**

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