Hindawi
Mobile Information Systems
Volume 2022, Article ID 7419920, 10 pages
https://doi.org/10.1155/2022/7419920

Research Article

A CSI 300 Index Prediction Model Based on PSO-SVR-GRNN Hybrid Method

Jialin Chen [1] and Hanyin Yang [2]

[1] *School of Finance, Southern University of Science and Technology, Shenzhen, Guangdong, China*
[2] *Zhongnan University of Economics and Law, Wuhan, China*

Correspondence should be addressed to Jialin Chen; 11912031@mail.sustech.edu.cn

Received 8 June 2022; Revised 27 June 2022; Accepted 5 July 2022; Published 4 August 2022

Academic Editor: Chia-Huei Wu

In this article, a PSO-SVR-GRNN nonparametric hybrid model is proposed for the CSI 300 stock index to forecast the problem. Particle Swarm Optimization (PSO) is utilized to optimize the parameters of the SVR model to enhance the prediction ability of the support vector machine's regression model for the original CSI 300 Index time series. The optimized residual sequence prediction results of the General Regression Neural Network (GRNN) are then used to optimize the time series prediction. The outcomes indicate that the PSO- SVR-GRNN model can greatly improve the prediction accuracy of the CSI 300 Index time series compared with individual models such as PSO-SVR, GRNN, GA-SVR, LSTM, PSO-LSTM, and SVR.

1. Introduction

Stock price indices are often compiled under a certain methodology based on the movement of stock prices during the corresponding period. As an effective measure of the capital market, the volatility of the stock index is usually observed as a barometer of changes in the national, social, economic, and political environment in which the market is based. After decades of development, the overall market capitalization of China's capital market has increased rapidly, and its volume has developed significantly. By contrast, the fact that the Chinese stock market is not well regulated and the supervision of listed companies is incomplete has led to the high volatility of the Chinese stock market. This has a considerable impact on the Chinese stock market, which retail investors dominate, so predicting stock market volatility is vital to warn investors of the risks. Furthermore, the stock index serves as a reflection of the stock market trend and as an underlying for financial innovation products such as index derivatives, achieving the establishment of the capital market in China. Accordingly, no matter for investors or regulators, the judgment of the stock index trend is of tremendous significance, so it has long been a concern by the majority of researchers. Nevertheless, the stock index is a reference for the movement of market stock prices and is influenced by diverse factors such as market sentiment. The time series created by the stock index has prominent nonstationary and nonlinear characteristics, which brings resistance to the advancement of research. This is why predicting stock index movements is still a popular research topic in the industry and academia. Currently, the research utilized to forecast stock index movements can be separated into four main categories: (1) general statistical models, (2) machine learning models, (3) combinatorial model algorithms, and (4) parameter optimization.

Researchers usually took statistical models to forecast stock price indices in the early days. Common statistical models comprise exponential smoothing models, ARIMA models (Ariyo et al. [1]), GARCH family of models, etc., (Wang et al. [2]). The autoregressive (AR) model is basically a model that operates past values and error terms of its series to forecast. Moreover, it is broadly applied for forecasting in economics for its robustness and better interpretability. However, stock markets are treacherous, and stock indices have powerful nonstationary and nonlinear features that are not easy to be caught by ARIMA. As a result, statistical methods have a limited position in stock index forecasting.

Later, with the development of machine learning and deep learning techniques, numerous researchers experimented with artificial intelligence techniques to predict stock index movements. Support Vector Machine (SVM) is a widespread algorithm in machine learning. It has an extensive number of kernel functions available, is adjustable in managing nonlinear regression problems with high characteristics and has proven its predictive effectiveness in the financial field (Cao and Tay [3], Huang et al. [4]). Artificial Neural Networks (ANNs) are famous models in deep learning. As an influential adaptive method, ANNs are able to continually decrease the error of the objective function to approximate the actual value effectively. In addition, it is able to seize the non-linear part of the time series, which is highly attractive to researchers (Khashei and Bijari [5]). Regardless, a single SVM model has the disadvantage of being challenging to execute large-scale sample training and handle multiclassification issues. Besides, ANNs models are sensitive to parameter settings so they may be slightly less effective in generalization in the financial domain. Consequently, while adopting the above techniques, developing the advantages and avoiding the disadvantages is required to enhance the CSI 300 Index prediction effectively.

To address the drawbacks of single model forecasting, researchers have started to experiment with hybrid models and have reached some outcomes in the field of financial time series forecasting [6–8]. Pai and Lin [9] took full advantage of the special advantages of ARIMA and SVM models in forecasting stock price problems to create a hybrid ARIMA-SVM model and acquired good consequences in predicting stock prices. Zhang [10] and Khashei et al. [11] formed a hybrid ARIMA-ANN model and proved that the neural network approach is well suited to capture non-linear features such as the prediction of residual sequences filled with noise factors. Some literature suggests that generalized regression neural network (GRNN) models have more substantial predictive power than other ANNs models. Yan [12] combined several GRNN models with various properties and obtained the best outcomes in prediction with 60 different models. Hence, this article tries to fuse Support Vector Regressor (SVR) and Generalized Regression Neural Network (GRNN) in order to be able to maximize the distinct advantages of both models to fulfill the goal of enhancing the prediction accuracy.

Based on this, the application of integrating hybrid models with parameter optimization on time series forecasting is also more common. Khandelwal et al. [13] introduced separate linear and nonlinear components of time series by DWT (Discrete Wavelet Transform) and identified and forecasted new models applying ARIMA and ANN models, separately. The consequences revealed that the performance of the combined model incorporating the DWT technique was enhanced compared to both the single model and the ARIMA-ANN model. Oliveira and Ludermir [14] made an ETS-ARIMA-AR-SVR hybrid evolutionary system for time series volatility so that the prediction outcomes were remarkably advanced compared to the traditional models. In recent years, scholars have paid more and more attention to the superior performance of the PSO algorithm in parameter optimization. Qin et al. [15] formed a PSO-ARIMA-DBN model for red tide biomass time series data, which can nicely seize the temporal correlation, spatial heterogeneity, and other nonlinear features between red tide biomass and environmental factors. In a study of traffic flow prediction, Hu et al. [16] observed that the SVR model optimized with PSO parameters could get more accurate predictions than other models. Hasanipanah et al. [17] applied the PSO-SVR model to the AOp prediction field and acquired the same conclusion. Therefore, this report will adopt the particle swarm algorithm (PSO) to optimize the model parameters to reach more accurate model prediction results.

As seen previously, previous literature has proved the effectiveness of "artificial intelligence techniques," "hybrid models," and "parametric optimization" in the time series forecasting and has acknowledged the potential of PSO algorithms for parametric optimization and the excellent performance of SVR and GRNN models in seizing financial time series features. Therefore, this article forms a PSO-SVR-GRNN hybrid model to forecast the CSI 300 stock index. First, the support vector regression (SVR) parameters are optimized utilizing the particle swarm algorithm (PSO) to forecast the CSI 300 stock index time series using the PSO-SVR model. Then, a generalized regression neural network (GRNN) model is constructed based on the residual series. Ultimately, the prediction values of the PSO-SVR-GRNN hybrid model for the CSI 300 stock index are reached. The main contributions of this paper are as follows:

(1) Motivated by the potential of the PSO algorithm in parameter optimization and the incredible performance of SVR and GRNN models in seizing financial time series features, this article makes the first PSO-SVR-GRNN hybrid model and uses it in the field of stock index forecasting

(2) Experiments reveal that the PSO-SVR-GRNN model proposed in this report greatly optimizes the single model and the base model in forecasting the CSI 300 stock index

2. Establishment of the PSO-SVR-GRNN Model

The establishment of the combined PSO-SVR-GRNN model is separated into the following processes.

2.1. Description of PSO-SVR-GRNN Process

(1) Acquiring and inputting the calendar year daily HSS 300 data.

(2) Apply MSE (mean square error) as the fitness evaluation index and split the time series into a training set and test set, respectively, utilizing the five-fold cross-check method. Subsequently, the particle swarm algorithm is adopted to find the optimal RBF kernel function parameters, insensitive

loss parameters, and penalty parameters C of the time series corresponding to the SVR model.

(3) Predicting the test set by the PSO-SVR model under the optimal parameters calculated in the previous step.

(4) Outputting the residual sequences after the training of the PSO-SVR model and taking the GRNN model to correct the residual sequences.

(5) Set up a combined PSO-SVR-GRNN model, train, and output the CSI 300 Index prediction outcomes.

The flow of the PSO-SVR-GRNN model building is shown in Figure 1.

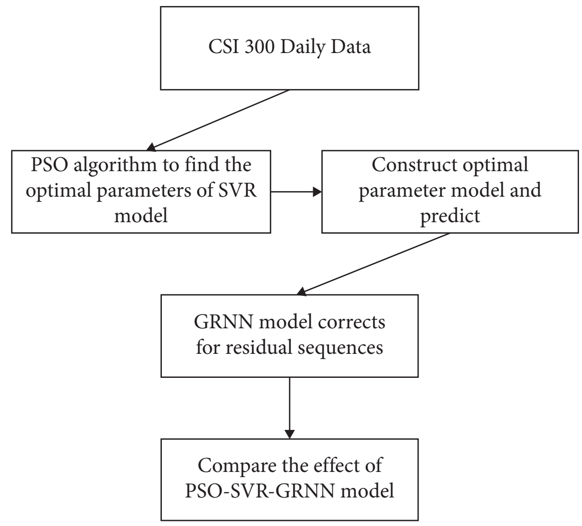

Figure 1: Flow chart of PSO-SVR-GRNN experiment.

2.2. Particle Swarm Algorithm. Eberhart and Kennedy first presented the particle swarm algorithm [18] in 1995. The concept of this algorithm is derived from the study of birds foraging behavior, by mimicking the foraging behavior of a flock of birds to discover the solution that makes the population optimal,

$$\begin{aligned} v_{\mathrm{id}}^{k+1} &= v_{\mathrm{id}}^{k} + c_1 \mathrm{rand}_1^k \left(p\mathrm{best}_{\mathrm{id}}^k - x_{\mathrm{id}}^k \right) \\ &\quad + c_2 \mathrm{rand}_2^k \left(g\mathrm{best}_{d}^k - x_{\mathrm{id}}^k \right) \\ x_{\mathrm{id}}^{k+1} &= x_{\mathrm{id}}^{k} + v_{\mathrm{id}}^{k+1}. \end{aligned} \quad (1)$$

In this algorithm, the parameters of each particle are its position $X_i\,(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{\mathrm{iD}})^T$ velocity $V_i = (v_{i1}, v_{i2}, v_{i3}, \ldots, v_{\mathrm{iD}})^T$, and fitness evaluation index. The original state is a flock of particles in a random position and with a random velocity (initial solution). Every particle updates the optimal value at each iteration by tracking the extreme points $p\mathrm{best}_{\mathrm{id}}^k$ of its trajectory and the extreme points $g\mathrm{best}_d^k$ of the particles in the swarm, where v_{id}^k is the velocity i of the particle at the kth latitude in the dth iteration, and x_{id}^k is the position of the particle i at the kth latitude in the dth iteration, Furthermore, C1 and C2 are the maximum step length of the particle i toward its historical optimal value point and the optimal value point in the population in the kth iteration, individually. If $v_{d\max}$ is too huge, it will cause the algorithm to fall into a local optimum, so it is essential to manage the maximum step sizes C1 and C2 in a reasonable range, At the end of each iteration, every particle's fitness is assessed, the flow of the particle swarm algorithm is shown in Figure 2.

2.3. SVR Model. In the 1990s, Vapnik [19] presented a support vector machine (SVM) model under statistical learning theory. After that, scholars took SVM to the field of regression prediction by proposing ε insensitive loss function to build a support vector regression (SVR) model; SVR is mainly adopted for time series prediction and optimization control, etc. The principle method is as follows.

In formula (1), given a set of training samples $X = \{ (x_i, y_i) | x_i \in R^n, \quad y_n \in R, i = 1, 2, \ldots, N\}$ (where Xi is the input feature vector, Yi is the target value, n is the number of samples, and SVR maps the low-dimensional nonlinear data ϕ to the high-dimensional space by a nonlinear function, followed by linear regression, to acquire $f(x)$.

$$f(x) = \omega^T \phi(x) + bx \in R^n, \quad b \in R. \quad (2)$$

Here, w and b are the regression coefficients and intercept terms, respectively. The idea of SVR is to find the optimal estimate of these two parameters so that the deviation between the function value corresponding to the training data $f(x)$ and the actual value y_i is less than ε , and the corresponding optimization problem is to discover the optimal value of the function. The corresponding optimization problem is to find the minimum value of the following equation:

$$\begin{aligned} &\mathrm{Min:}\ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*), \\ &s.t. \begin{cases} y_i - f(x_i) \le \varepsilon + \xi_i, \\ y_i - f(x_i) \ge \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \ldots, N. \end{cases} \end{aligned} \quad (3)$$

Here, $C > 0$ is the penalty parameter, ξ_i , ξ_i^* is the positive slack variable, and ε represents the insensitive loss function. The Lagrangian transformation is performed for the pairwise problem of the above problem, and the following equation is reached:

$$\begin{aligned} L &= \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\lambda_i(\varepsilon + \xi_i - y_i + f(x_i)) \\ &\quad - \sum_{i=1}^{N}\lambda_i^*(\varepsilon + \xi_i^* + y_i - f(x_i)) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*). \end{aligned} \quad (4)$$

Here, $\lambda_i, \lambda_i^*, \eta_i, \eta_i^* \ge 0, i = 1, 2, \ldots, N$, the extreme value of the function is managed by the pairwise principle, so that its partial derivative concerning the variable $\omega, b, \xi_i, \xi_i^*$ is 0.

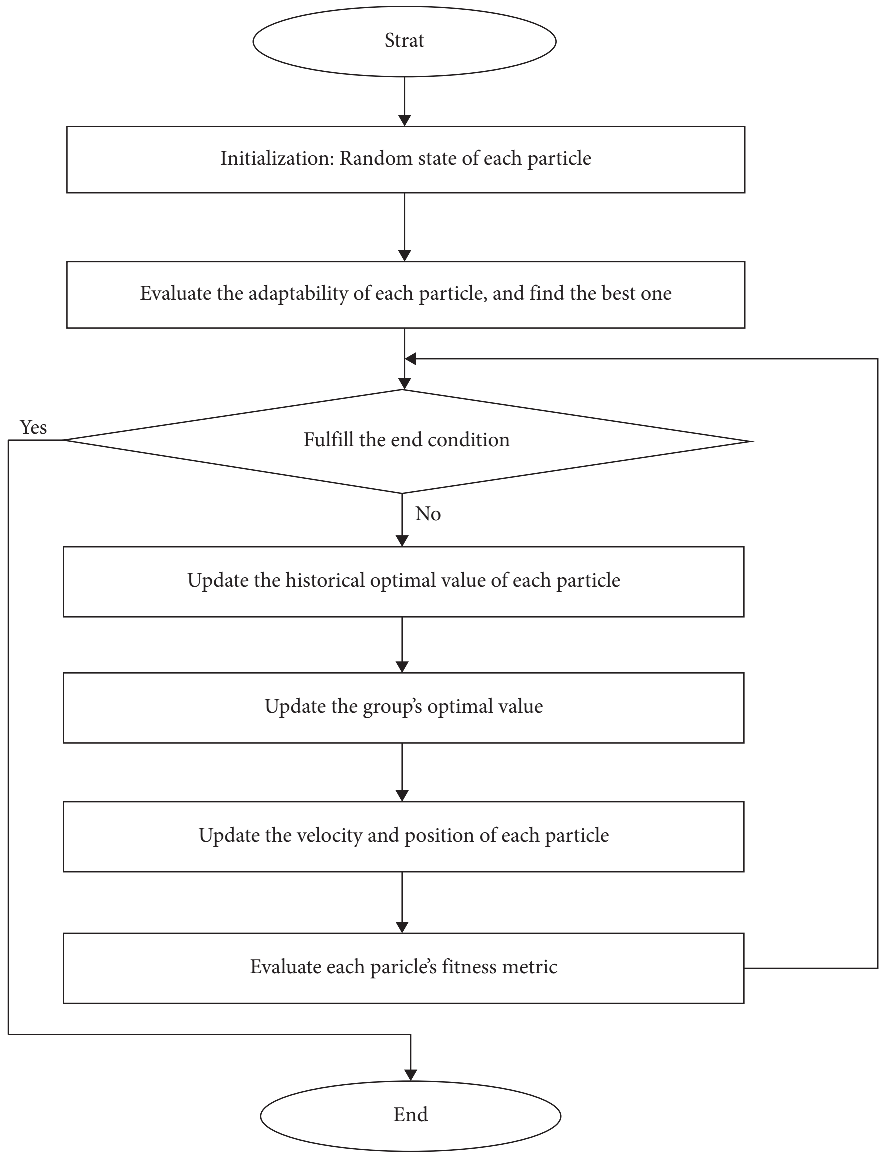

Figure 2: Iteration flow chart of particle swarm algorithm.

The kernel function $K(x, x_i) = \phi(x) \cdot \phi(x_i)$ is proposed to replace the inner product operation on the high-dimensional space, and the original regression function (1) is reduced to the following equation:

$$f(x) = \sum_{i=1}^{N} (\lambda_i - \lambda_i^*) K(x, x_i) + b. \quad (5)$$

Here, the usual kernel function selection contains linear kernel function, polynomial kernel function, and Gaussian radial basis kernel function (RBF), etc. From previous experience, the RBF kernel function is best applied in the absence of prior knowledge of the sample data, so in this paper, RBF is chosen as the kernel function, namely,

$$K(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{\sigma^2}\right), \quad \sigma > 0. \quad (6)$$

Here, σ denotes the bandwidth, and thus the original SVR model is explicitly an optimization problem solving for the bandwidth σ , insensitivity coefficient ε , and penalty parameters C.

2.4. GRNN Model

2.4.1. Introduction of the Model Algorithm. Generalized regression neural network (GRNN) is a kind of radial basis neural network presented by Sprecht. It has better nonlinear mapping ability and learning speed and is a generally utilized nonparametric regression model. A common nonparametric regression model can accomplish reasonable prediction results even with small sample data, GRNN largely includes a radial basis network layer and a linear network layer. Its MATLAB function-based network structure is shown below in Figure 3: (Abdollahi and Ebrahimi [20]):

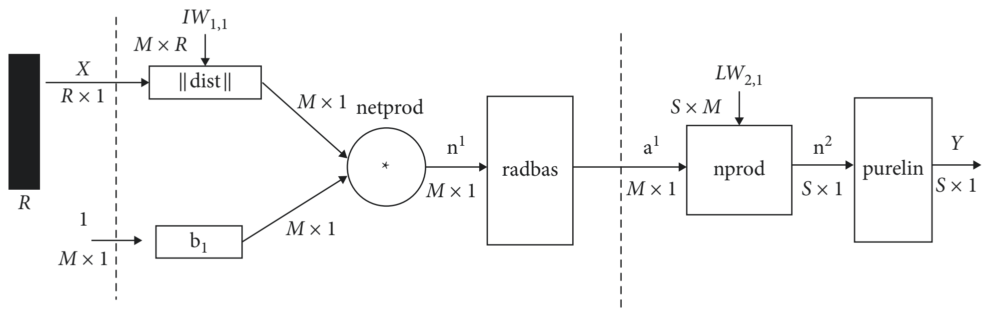

Figure 3: Schematic diagram of GRNN structure.

In formula (7), the input dimension is a vector X of $R \times 1$, M refers to the number of training samples, and the final network output results in Y. In the GRNN network structure, the number of neurons in the first radial basis network layer is the number of training samples M. The b_1 is the threshold of the hidden layer, and the weight function in this layer is the Euclidean distance function dist, which is taken to measure the distance between the network input and the first layer weight IW1,1. After that, the dist function output is multiplied with the threshold b_1 element, and the result n^1 is passed to the transfer function. Note that, the transfer function here is chosen as a Gaussian function normally operated in radial basis functions, namely,

$$a_i^1 = \exp\left(-\frac{\left(n_i^1\right)^2}{2\sigma_i^2}\right) = \exp\left(-\frac{\left(\|\text{dist}\| b_{1,i}\right)^2}{2\sigma_i^2}\right), \quad i = 1, 2, \ldots, M. \tag{7}$$

Among them, σ_i is named the smooth factor. The value of σ_i determines the shape of the basis function at the position of the ith hidden layer, and the larger the value, the smoother the basis function.

The linear output layer of the network utilizes the planned dot product function as the weight function. It separates the dot product obtained by dividing the result a^1 with the weight LW2,1 by the resultant vector element sum. The acquired consequence is passed into the purelin function, and the network output is measured.

2.4.2. Determination of Smoothfactor Parameters. The smooth factor σ impacts the prediction outcomes of the GRNN model. Since it decides the degree of smoothing of the basis function in the implicit layer of the GRNN model, it is essential to select a suitable smooth factor to improve the prediction accuracy of the network. This paper determines to perform the optimization search calculation for GRNN taking values in one dimension, and the comparison criterion is MSE. The exact optimization steps are as follows:

(1) Choose the initial smooth factor σ

(2) Pick one sample from the training set for comparison, and the rest of the samples for network construction

(3) Measure the absolute value of the detection sample error under the constructed network model

(4) Repeat steps (2) and (3) so that there is an error value for each sample in the sample set and measure the mean squared error for all sample points, and the minimum value is used as the test condition, where the mean squared error MSE formula is as follows:

$$MSE(\sigma) = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{y}_t - y_t\right]^2. \tag{8}$$

Here, y_t is the training sample value, and $\hat{y}_t$ is the GRNN network prediction value

The initial smooth factor is set to 0.1. The value of the smooth factor is chosen to increase by 0.1 each time for optimization; ultimately, the MSE of the sample prediction is picked as the criterion to discover the optimal smooth factor value with its minimum as the goal

3. Experiment

3.1. Preprocessing of the Original Data. The CSI 300 index covers the leading enterprises in 19 industries such as real estate, electronics, and pharmaceuticals. As an indicator reflecting the overall market situation of China's stock market, CSI 300 index not only reminds the market risks but also provides numerous investment products such as stock index futures. It is a key foundation for China's financial market development. In this study, the authors studied the real closing price per day of the CSI 300 index, taking 964 trading days (Feb. 2002–Dec.31, 2020) as the original time series (Figure 4).

The original time series of CSI 300 were normalized (Figure 5) and divided into a training set (the former 80% of the trading days) and a test set (the latter 20% of the trading days). Then, the data of the training set were input into the model. The sample was a matrix with a scale of 771 (964*0.8).

3.2. Parameter Optimization Based on Particle Swarm Optimization (PSO). In this study, the PSO method was used to optimize the penalty parameter C of the SVR model, the parameter g of the RBF kernel function, and the insensitive loss function p. In combination with 5-fold cross validation, MSE was selected as the measurement index of the fitness function; the number of evolution generations was 200; the

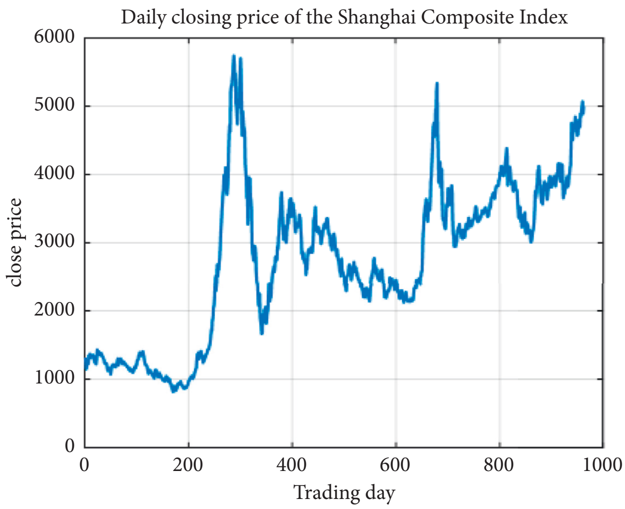

Figure 4: The original time series of CSI 300.

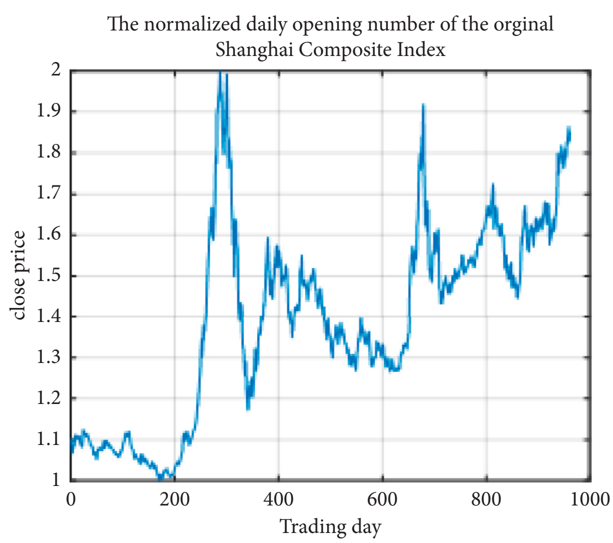

Figure 5: The normalized time series of CSI 300.

population quantity was 20. The fitness curve of the PSO-based parameter optimization is illustrated in Figure 6.

According to Table 1, the penalty parameter C of the SVR model, the parameter g of RBF kernel function, and the insensitive loss function p are 2.8141, 0.0100, and 0.0100, respectively, with MSE less than 0.01. It can be judged that the three parameters obtained in the PSO method have strong fitness.

3.3. PSO-SVR Model Prediction. Based on the above-given three parameters optimized by PSO, an SVR model was established to predict the time series of CSI 300, taking the latter 20% of the trading days as the test set. The final true value and predicted value are shown in Figure 6.

As observed in Figure 7, the PSO-SVR model trained with the former 80% of the trading days can accurately capture the linear and non-linear characteristics of the time series and fit the test set.

As revealed from the evaluation indices of the PSO-SVR model fitted the test set (Table 2), the MSE is only 12528, showing a good fitting effect.

3.4. GRNN Model-Based Error Correction. First, the residual series of the PSO-SVR model were obtained (Figure 8).

Then, the residual series were input into the GRNN model. The 5-fold cross validation method was used to divide the dataset into a training set and a test set, obtaining the optimum smoothing factor of 0.6. As observed in Figure 9, the GRNN model with the best parameters was used to fit the residual series.

Compared with the PSO-SVR model, the fused model subjecting to residual correction by GRNN was significantly improved in effect. The RMSE decreased from 111.93 to 111.24 and the MSE from 12528–12374 (Table 3).

4. Comparison

4.1. Contrast Test with the GRNN Model. The original series of CSI 300 indices were treated as the dataset and divided into a training set and test set using a 5-fold cross validation method. Followed by, the optimum smoothing factor was obtained, 0.1, and input into the GRNN model parameter to fit the original time series.

As discovered in Figure 10, the GRNN-predicted results of the original sample comply with the actual data, presenting a good fitting effect. However as displayed in Table 4, the prediction effect of GRNN is inferior to the combined model.

4.2. Contrast Test with the LSTM Model. The original series of CSI 300 indices were treated as the dataset and divided into a training set and test set based on the above-given contrast test design. Then, the LSTM model parameter was used to fit the original time series.

As can be seen from Figure 11, the LSTM-predicted results of the original sample conform to the actual data, demonstrating a good fitting effect. However as displayed in Table 4, the prediction effect of LSTM is inferior to the combined model significantly.

4.3. Contrast Test with the SVR Model. Also, the original series of CSI 300 indices were treated as the dataset and further divided into a training set and a test set based on the said contrast test design. Then, the SVR model parameter was used to fit the original time series. Figure 12 displays the optimization results of the SVR parameter and Figure 13 presents the SVR's fitting to the original series.

As exhibited in Figure 13, the SVR-predicted results of the original sample accord with the actual data, which presents a good fitting effect. However as implied in Table 4, the prediction effect of SVR is still inferior to the combined model.

4.4. Contrast Test with the PSO-LSTM Model. The original series of CSI 300 indices were divided into a training set and a test set, with the same design as the aforesaid test. Then, the

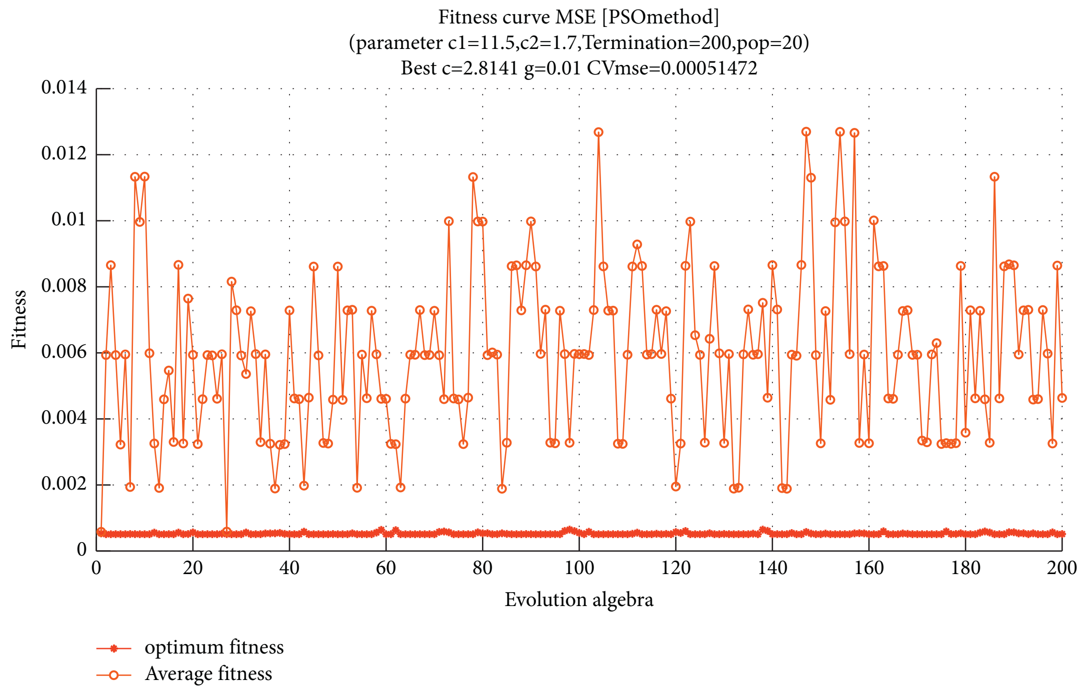

Figure 6: MSE changes in the fitness curve during the PSO-based parameter optimization.

Table 1: The results of parameters optimized by PSO.

Sequence of information granulation parameter	Bestc	Best_g	Best_p	MSE
All	2.8141	0.0100	0.0100	0.0005

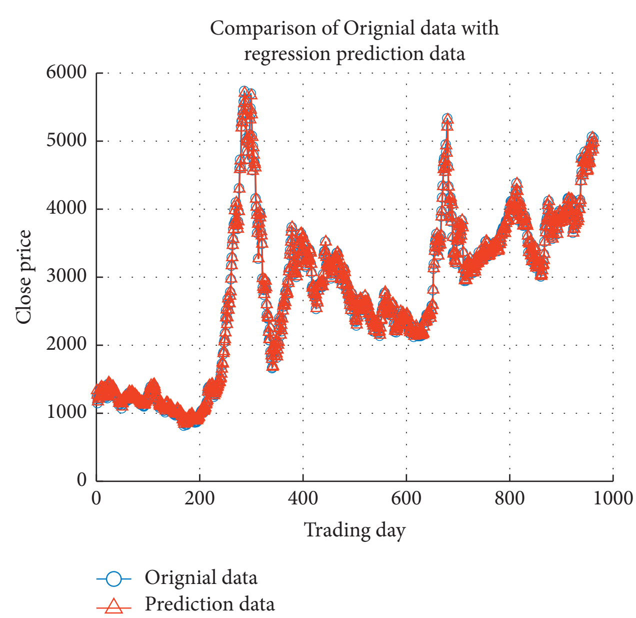

Figure 7: The PSO-SVR's fitting results.

Table 2: PSO-SVR model effect.

Evaluation index	RMSE	MSE	MAPE	MAE
Index value	111.93	12528	2.7193	76.0757

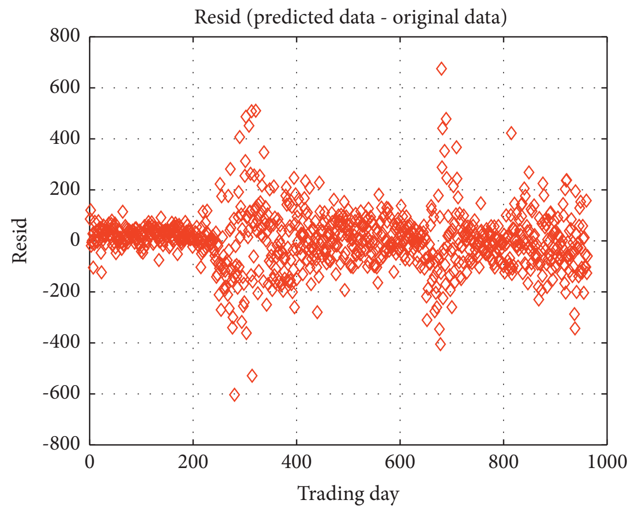

Figure 8: Residual scatter diagram of the PSO-SVR model.

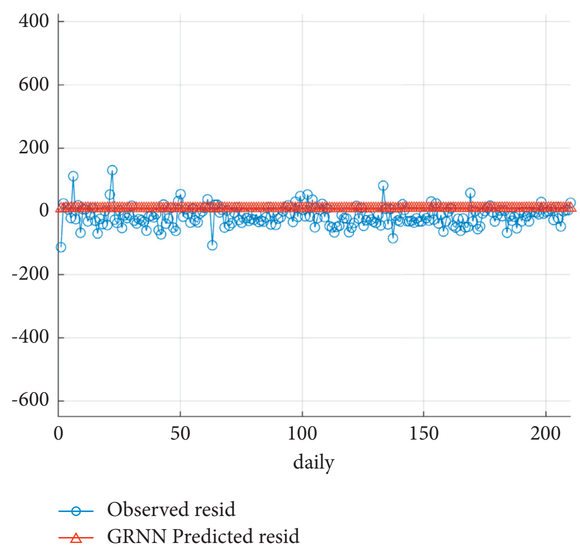

Figure 9: Residual series fitting results of the GRNN.

Table 3: Comparision between the effects of the PSO-SVR model and the PSO-SVR-GRNN model.

Evaluation index	RMSE	MSE	MAPE	MAE
PSO-SVR	111.93	12528.00	2.7193	76.0757
PSO-SVR-GRNN	111.24	12374.00	2.4387	75.8048

PSO-LSTM model parameter was used to fit the original time series. The PSO-LSTM's fitting to the original series is shown in Figure 14.

According to Figure 14, the PSO-LSTM model-predicted results of the original sample tally with the actual data, displaying a preferable fitting effect. However, the prediction effect of the PSO-LSTM model is still inferior to the combined model as clearly shown in Table 4.

4.5. Contrast Test with the GA-SVR Model. Taking the original series of CSI 300 indices as the dataset, the training

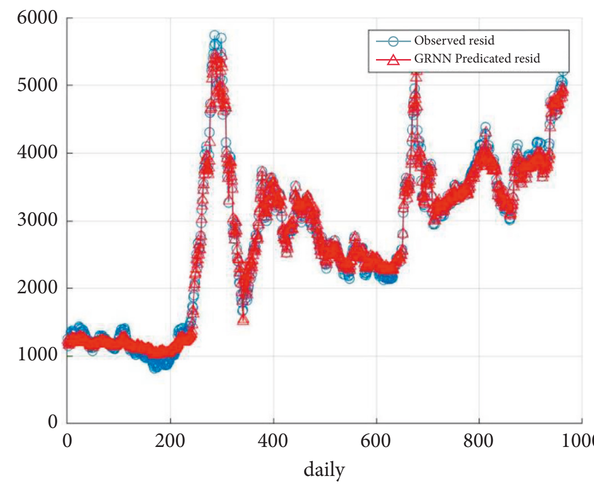

Figure 10: The GRNN's fitting to the original series.

Table 4: Comparison of the models' effects.

Evaluation index	RMSE	MSE	MAPE	MAE
PSO-SVR	111.93	12528.00	2.7193	76.0757
PSO-SVR-GRNN	111.24	12374.00	2.4387	75.8048
GRNN	128.33	16470.79	4.2362	96.9851
LSTM	123.14	15165.78	3.0256	83.1749
SVR	110.75	12266.14	2.6367	73.9106
GA-SVR	110.43	12197.00	2.6556	74.4111
PSO-LSTM	147.46	21744.65	3.2696	109.6961

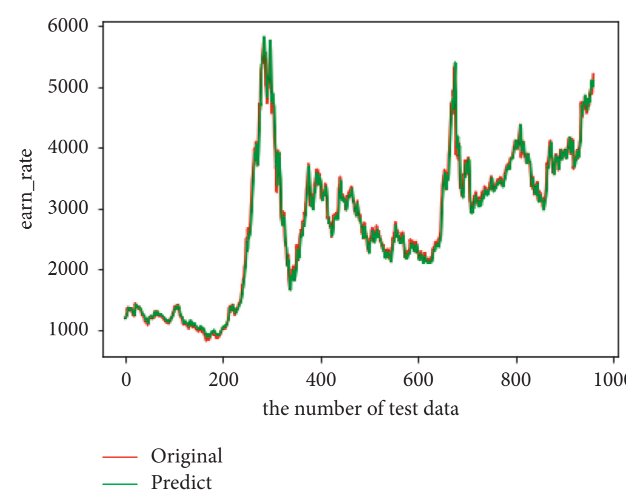

Figure 11: The LSTM's fitting to the original series.

set and the test set were also divided into with the same design as the aforesaid test. Furthermore, the GA-SVR model parameter was used to fit the original time series. The optimization results of the GA-SVR parameter are displayed in Figure 15 and the GA-SVR's fitting to the original series is demonstrated in Figure 16.

As revealed in Figure 16, the GA-SVR model-predicted results of the original sample correspond to the actual data, demonstrating a sound fitting effect. However as observed in Table 4, the prediction effect of this model is also weaker than the combined model.

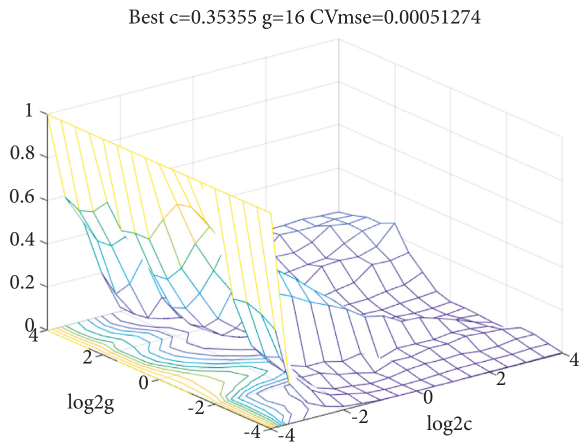

Figure 12: The optimization results of the SVR parameter.

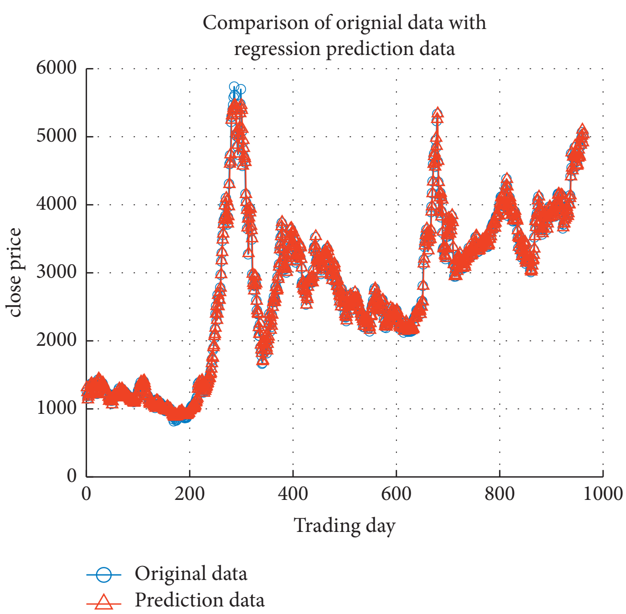

Figure 13: The SVR's fitting to the original series.

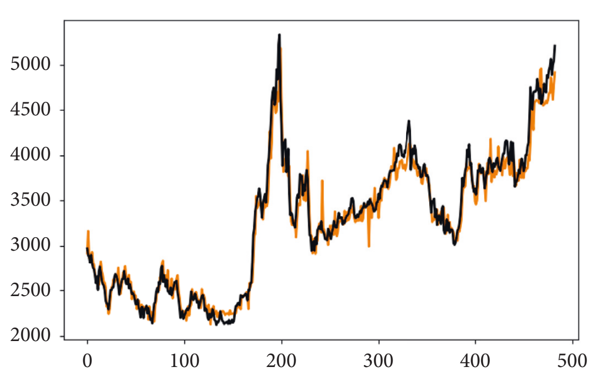

Figure 14: The PSO-LSTM's fitting to the original series.

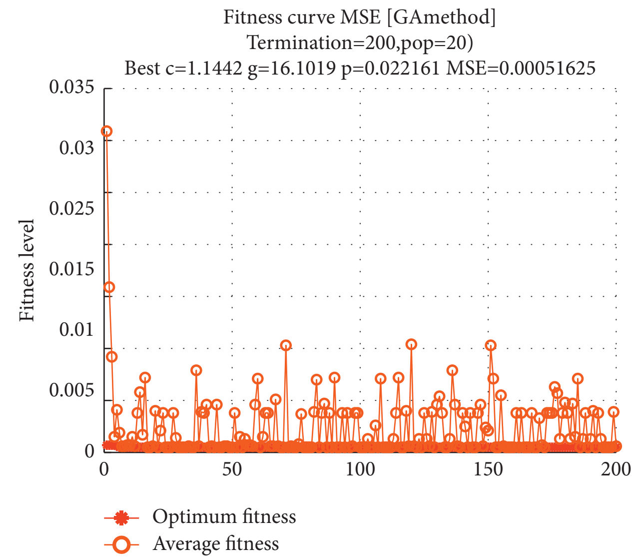

Figure 15: The optimization results of the GA-SVR parameter.

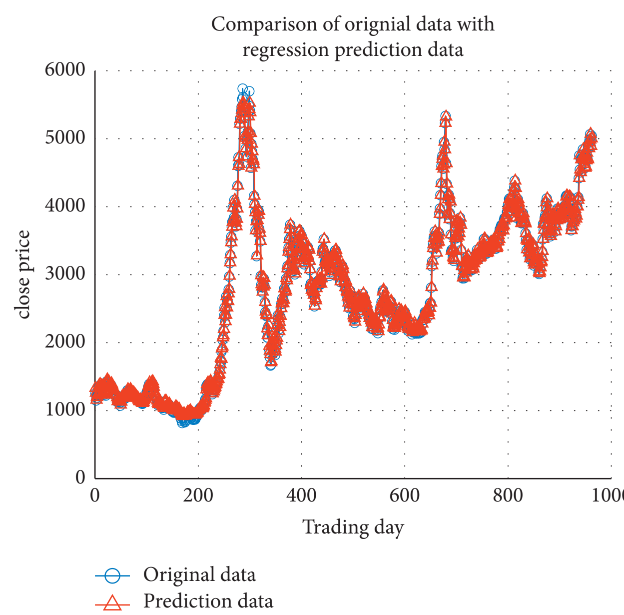

Figure 16: The GA-SVR's fitting to the original series.

5. Conclusion

Considering the inaccuracy of the CSI 300 index, relevant studies are analyzed and a combined CSI 300 index prediction model is proposed based on the PSO-SVR-GRNN.

First, this model has a short predictive period as it is established as per the historical data of the CSI 300 index. For a long-term prediction of the CSI 300 index, it is needed to further study how to improve the model by introducing economic change and other external factors. The complementarity of different prediction methods, the direct fusion of different methods, and how to improve the accuracy and reliability of the CSI 300 index prediction model are also the development trends in future studies on the prediction.

Second, accurate prediction of the CSI 300 index is a difficulty in many studies and is significant for the government to grasp the future macro-economic situation and trend. As demonstrated by contrast tests, the MSE, RMSE, MAPE, and MAE of the proposed model's prediction results are smaller than other models. Hence, the proposed model's prediction results are more accurate. This model combines the advantages of the PSO-SVR model and the GRNN model and was also used for actual prediction. The obtained prediction results varied in identical scope and trend to the actual data, which verifies the effectiveness and practicability of this model.

Third, the combined nonparametric model has more accurate and effective prediction results than the single model, as discovered over a comparison between the PSO-SVR model and the PSO-SVR-GRNN model. This result also indicates that the CSI 300 index series practically has complicated nonlinear structural features.

In further studies, more input data such as macro data, investor sentiment, and fundamental information will be used to increase the prediction accuracy. Moreover, PSO or GA algorithms will be used to find the best parameters of GRNN to optimize the existing model. Finally, the model will be extensively evaluated using various stock test data.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the ARIMA model," in *Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp. 106–112, IEEE, Cambridge, UK, March, 2014.

[2] J. J. Wang, J. Z. Wang, Z. G. Zhang, and S. P. Guo, "Stock index forecasting based on a hybrid model," *Omega*, vol. 40, no. 6, pp. 758–766, 2012.

[3] L. Cao and F. E. Tay, "Financial forecasting using support vector machines," *Neural Computing & Applications*, vol. 10, no. 2, pp. 184–192, 2001.

[4] W. Huang, Y. Nakamori, and S. Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers & Operations Research*, vol. 32, no. 10, pp. 2513–2522, 2005.

[5] M. Khashei and M. Bijari, "An artificial neural network (p, d,q) model for timeseries forecasting," *Expert Systems with Applications*, vol. 37, no. 1, pp. 479–489, 2010.

[6] G. P. Zhang and V. L. Berardi, "Time series forecasting with neural network ensembles: an application for exchange rate prediction," *Journal of the Operational Research Society*, vol. 52, no. 6, pp. 652–664, 2001.

[7] Y. Wang and Y. Guo, "Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost," *China Communications*, vol. 17, no. 3, pp. 205–221, 2020.

[8] Z. Luo, W. Guo, Q. Liu, and Z. Zhang, "A hybrid model for financial time-series forecasting based on mixed methodologies," *Expert Systems*, vol. 38, no. 2, Article ID 12633, 2021.

[9] P. F. Pai and C. S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[10] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[11] M. Khashei, M. Bijari, and G. A. Raissi Ardali, "Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs)," *Neurocomputing*, vol. 72, no. 4–6, pp. 956–967, 2009.

[12] W. Yan, "Toward automatic time-series forecasting using neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1028–1039, 2012.

[13] I. Khandelwal, R. Adhikari, and G. Verma, "Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition," *Procedia Computer Science*, vol. 48, pp. 173–179, 2015.

[14] J. F. de Oliveira and T. B. Ludermir, "A hybrid evolutionary decomposition system for time series forecasting," *Neurocomputing*, vol. 180, pp. 27–34, 2016.

[15] M. Qin, Z. Li, and Z. Du, "Red tide time series forecasting by combining ARIMA and deep belief network," *Knowledge-Based Systems*, vol. 125, pp. 39–52, 2017.

[16] W. Hu, L. Yan, K. Liu, and H. Wang, "A short-term traffic flow forecasting method based on the hybrid PSO-SVR," *Neural Processing Letters*, vol. 43, no. 1, pp. 155–172, 2016.

[17] M. Hasanipanah, A. Shahnazar, H. Bakhshandeh Amnieh, and D. Jahed Armaghani, "Prediction of air-overpressure caused by mine blasting using a new hybrid PSO–SVR model," *Engineering with Computers*, vol. 33, no. 1, pp. 23–31, 2017.

[18] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE*, vol. 4, pp. 1942–1948, 2017.

[19] N. Vladimir, *Vapnik. The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 2000.

[20] H. Abdollahi and S. B. Ebrahimi, "A new hybrid model for forecasting Brent crude oil price," *Energy*, vol. 200, Article ID 117520, 2020.