Fuzzy Systems and Data Mining IX
A.J. Tallón-Ballesteros and R. Beltrán-Barba (Eds.)

doi:10.3233/FAIA231018

A Novel Nonferrous Metals Price Prediction Model Based on BiLSTM-ResNet with Grey Wolf Optimization and Wavelet Transform

ZHANGLONG LI [a], YUNLEI YANG [a,1], JIACHUN ZHENG [a], YANG WU [a] and YINGHAO CHEN [b]

[a] *School of Mathematics and Statistics, Guizhou University, Guiyang, Guizhou 550025, China*

[b] *School of Mathematics and Statistics, Central South University, Changsha, 410083, China; Eastern Institute for Advanced Study, Yongriver Institute of Technology, Ningbo, Zhejiang 315201, China*

Abstract. Nonferrous metals are important commodities, and it is of great significance for policy makers and investors to accurately predict their price changes. Nevertheless, because the price of nonferrous metals present drastic fluctuations, developing a robust price prediction method is a tricky task. In this research, a hybrid model based on discrete wavelet transform (DWT), bidirectional long short-term memory (BiLSTM) and residual network (ResNet) is constructed for nonferrous metals price prediction. The hyper-parameters of the hybrid neural network are searched by grey wolf optimization (GWO) algorithm. Configuring reasonable parameters, which enhances the final prediction effect. Additionally, behind the second hidden layer, the low and high dimensional features are fused to prevent the degradation of the model. The original sequence is processed by DWT technology, then the sequence is reconstructed, which is beneficial to capture the essential trend. The experimental results show that the proposed BiLSTM-ResNet-GWO-DWT model is more accurate compared with the other benchmark models, which provides an effective reference significance.

Keywords. Nonferrous metals price prediction, BiLSTM, ResNet, Discrete wavelet transform, Grey wolf optimization

1. Introduction

Nonferrous metals, as the bulk commodity, have a vital role in the global economy. For instance, aluminum, copper and zinc are closely related to aviation, construction and other industries. Unique material characteristics make them more widely use. Meanwhile, the fluctuation of these nonferrous metals price is a barometer of economic action,

[1]Corresponding Author: Yunlei Yang, School of Mathematics and Statistics in Guizhou University, ylyang5@gzu.edu.cn

which offers a risk reference for the financial participants [1,2,3]. There is no denying that it is a meaningful work to carry out the relevant price prediction research.

In the past, some traditional statistical models had been proposed for the metals price prediction. For example, Thomas Kriechbaumer et al. proposed the improved wavelet-ARIMA approach, and demonstrated its usefulness in predicting monthly prices for aluminum, copper, lead and zinc [4]. Nevertheless, the ARIMA-type forecast models require that temporal data is stable and essentially only capture linear relationships. Lasheras et al. tested that the prediction performance of both neural network models (multilayer perceptron and Elman), using copper spot price data released on the New York Mercantile Exchange, was better than ARIMA [5]. What's more, the ARIMA model has an error assumption, namely, the constant variance. But generally for highly fluctuating data, the variance of the error is nonconstant. Consequently, the ARIMA model is basically not applicable to gold prices. On account of the above considerations, Yaziz et al. combined ARIMA and Symmetric GARCH-type Models to forecast Malaysian gold price [6]. However, these equations are based on the hypothesis of independence of local laws, that is why they are only suitable for short-term forecast.

In the advance of artificial intelligence (AI), many excellent data prediction models have been published based on extreme learning machine (ELM), support vector machine (SVM), decision tree and so on [7,8,9]. For instance, Kumar Chandar Sivalingam et al. taken full advantage of ELM method to focus on predicting the future gold prices of four commodities to predict business trend[10]. Results proved that ELM model was better than other feed-forward networks. Moreover, Shao Bilin et al. used the reformative particle swarm optimization (PSO) to optimize LSTM network and then predicted the nickel price [11]. Contrasted with the traditional LSTM and ARIMA model, the prediction error of this model was reduced by 9% and 13%, respectively, which had high reliability. Notwithstanding, the generalization ability of the single model is limited. Hou Muzhou et al. proposed the hybrid constructive neural network method (HCNNM) to fix the impact values in the raw data [12,13]. Through a series of numerical experiments, compared with the machine learning algorithms (including ELM, SVM and deep learning), they fully verify the correctness and effectiveness of the proposed theory. Chengshi Tian and Yan Hao proposed a newly designed hybrid optimization algorithm and applied to train adaptive neuro-fuzzy inference system, which provides more reliable information for decisions according to the point and interval prediction of carbon price sequence [14]. Dania Batool et al. Designed a tea yield prediction model based on combining different data sources through a crop simulation model and machine learning algorithms, which offers guidance for improving management practice [15].

At present, it is a consensus that fusion model is more generalized than single model. The hybrid stochastic-grey model with the singular spectrum analysis is proposed by Zoran Gligoric et al. to predict the behavior of metal price, as a result, it has better prediction accuracy than the traditional grey model [16]. Moreover, Ni Jian et al. combined LSTM-ANN with GARCH model to forecast the copper price volatility and significantly promote the forecast precision [17]. In 2022, a hybrid neural network combining Bayesian optimization and wavelet transform was applied to copper price prediction by Liu Kailei et al [18]. According to the advantage of each model, the hybrid model was fully utilized to capture the fluctuation characteristics of the historical data, whereas the optimal configuration of the hyperparameters was insufficient.

Undoubtedly, a majority of efforts have been made to heighten the effectiveness of predicting non-ferrous metals price, but a more stable benchmark model should be proposed. In the light of the consideration of the network degradation phenomenon, the purpose of this study is to establish a novel model based on the ResNet and BiLSTM neural network. At the same time, optimizing the hyper-parameters is a time-consuming process, but it is very imperative for the stability of the model. Previously, the optimization of hyperparameters can be performed by grid search or random search, but neither method was satisfactory. On the one hand, grid search solutions will consume a lot of computing resources. On the other hand, although random search will generally be faster than grid search, the global optimal value is not guaranteed. In addition, due to the complexity of the raw sequence, it is essential to make the data preprocessing. In view of the above problems, we construct the hybrid BiLSTM-ResNet model with GWO algorithm and discrete wavelet transform to forecast the non-ferrous metals price. Here, the wavelet transform is selected to remove noise and then the processed data input to the mentioned hybrid network.

The arrangement of this paper is presented as follows. The methodology including the wavelet transform, grey wolf optimization and normalization are introduced in section 2. We describe the basic theories of the related BiLSTM and ResNet model in section 3. Experiment results and its analyses are presented in section 4. Section 5 summarizes the entire paper, puts forward conclusions. The direction of the future work is given in the final section.

2. Methodology

2.1. Wavelet transform

The wavelet transform algorithm has been widely used in the fields of science, engineering and so on [19]. It is a significant time-scale and time-frequency localization analysis method that both time window and frequency window can be changed. There are two forms of wavelet transform including continuous wavelet transform (CWT) as well as discrete wavelet transform (DWT) [18,20].

The CWT is defined as shown in Eq. (1).

$$CWT_h(m,\tau) = \frac{1}{\sqrt{|m|}} \int_R h(t)\phi(\frac{t-\tau}{m})dt \tag{1}$$

where m, τ, ϕ and h(t) indicate the scaling factor, translating factor, the mother wavelet and the raw signal, respectively.

The DWT is defined as shown in Eq. (2).

$$DWT_h(p,q) = (\frac{1}{\sqrt{2}})^p \int_R h(t)\phi(2^{-p}t - q)dt \tag{2}$$

where p called scale is a parameter characterizing the frequency; q is a parameter characterizing the temporal or spatial location.

The DWT is applied to remove noise Gaussian interference from the raw sequence, with the aim to reduce the adverse effects of complex nonlinear signal. To make the reconstructed signal not distorted, the Daubechies of order 2 (db2) is choosen as the

benchmark wavelet. Then, the key step is the wavelet inversion reconstruction to return the denoised data to the time-frequency domain. Ultimately, a threshold function should be selected to filter the Gaussian noise coefficient, such as the soft threshold or the hard threshold method.

The hard threshold method is shown as follows:

$$\overline{\mu}_{i,j} = \begin{cases} \mu_{i,j} & if \quad |\mu_{i,j}| \geq \alpha, \\ 0 & if \quad |\mu_{i,j}| < \alpha. \end{cases} \tag{3}$$

The soft threshold method is shown as follows:

$$\overline{\mu}_{i,j} = \begin{cases} sgn(\mu_{i,j})|\mu_{i,j} - \alpha| & if \quad |\mu_{i,j}| \geq \alpha, \\ 0 & if \quad |\mu_{i,j}| < \alpha. \end{cases} \tag{4}$$

where $\mu_{i,j}$, α is the wavelet coefficient after multi-layer decomposition and threshold, respectively. In this research, the soft threshold method is used.

2.2. *Grey wolf optimizer*

The grey wolf optimizer (GWO) inspired by the prey hunting activity of grey wolves, is a group intelligent optimization algorithm developed by scholar Mirjalili et al. of Griffith University in Australia [21].

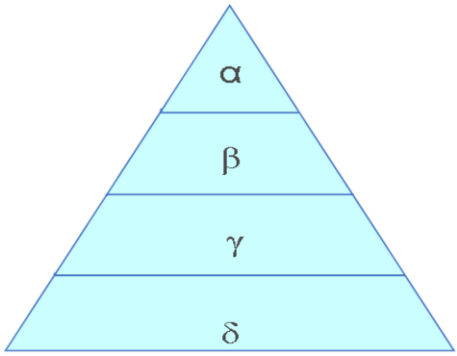


Fig. 1. The top-down dominance hierarchy from grey wolf.

The grey wolf belongs to social canids. They strictly adhere to a top-down social dominance hierarchy, as shown in Fig.1. Where the head wolf in the wolf pack is recorded as α, and α wolf is mainly responsible for making decisions on predation, habitat, work and rest time and other activities. Moreover, β wolf, it obeys and assists α wolf in making decisions, but at the same time β wolf can control wolves at other social levels. After α wolf dies or aging, β wolf will be the best candidate for α wolf. The γ wolf generally consists of juvenile, sentinel, hunting, aged and nursing wolves. It obeys α and β wolf, while dominating the remaining hierarchy of wolves. The δ wolf, it usually needs to obey other social levels of wolves. Although it seems that δ wolf plays little role in wolves, δ wolf exists for reasonable team building, such as avoiding infighting problems.

The algorithm considers the optimal solution as a prey which is constantly approached by the grey wolf pack, as follows: (a)The hunting space of the grey wolf pack is set to $m \times n$ Euclidean space, where m is the number of grey wolves and n is the dimension of the prey. The location of each wolf can be expressed as $Z_i = (z_1, z_1, ..., z_n)(i = 1, 2, ..., m)$. Nextly, the distance between each grey wolf and its prey was judged according to the hyperparameters of the location. Assuming α, β and γ wolf have stronger ability to identify potential prey location, the rest is the δ wolf. (b)Under the guidance of the three strongest wolves, the search activity is updated according to the location information of the four wolves. The mathematical model is represented by Eqs. (5)-(7) as follows:

$$\begin{cases} \overrightarrow{d}_\alpha = |\overrightarrow{c}_1 \bullet \overrightarrow{Z}_\alpha - \overrightarrow{Z}_t|, \\ \overrightarrow{d}_\beta = |\overrightarrow{c}_2 \bullet \overrightarrow{Z}_\beta - \overrightarrow{Z}_t|, \\ \overrightarrow{d}_\gamma = |\overrightarrow{c}_3 \bullet \overrightarrow{Z}_\gamma - \overrightarrow{Z}_t|. \end{cases} \tag{5}$$

$$\begin{cases} \overrightarrow{Z}_1 = \overrightarrow{Z}_\alpha - \overrightarrow{A}_1 \bullet d_\alpha, \\ \overrightarrow{Z}_2 = \overrightarrow{Z}_\beta - \overrightarrow{A}_2 \bullet d_\beta, \\ \overrightarrow{Z}_3 = \overrightarrow{Z}_\gamma - \overrightarrow{A}_3 \bullet d_\gamma. \end{cases} \tag{6}$$

$$\begin{cases} \overrightarrow{C} = 2\overrightarrow{r}_2, \\ \overrightarrow{A} = 2\overrightarrow{a} \bullet \overrightarrow{r}_1 - \overrightarrow{a}, \\ \overrightarrow{Z}_{t+1} = \frac{\overrightarrow{Z}_1 + \overrightarrow{Z}_2 + \overrightarrow{Z}_3}{3}. \end{cases} \tag{7}$$

where $\overrightarrow{d}_\alpha$, $\overrightarrow{d}_\beta$ and $\overrightarrow{d}_\gamma$ indicate the distance between the current candidate wolf and the optimal three wolves, respectively. Z_α, Z_β, and Z_γ indicate the positional vectors of α, β and γ wolf in the current population, respectively. Z_t represents the location vector of the grey wolf in the t iterations, $\overrightarrow{a}$ drops linearly from 2 to 0 throughout the iteration; r_1 and r_2 are random vectors in the range [0,1], and '•' represents the Hadamard product.

When $|A| < 1$, then the next moment position of search agent can be at any position between the current grey wolf and the prey. For the establishment of the dispersion model with $|A| > 1$, the search method enables the GWO to make a global search. Besides, the vector $\overrightarrow{C}$ composed of random values over the interval range [0,2], which is a nonlinear decreasing search coefficient that provides random weights to the prey [22].

2.3. Data normalization

It is significant that a standardization processing can improve training efficiency, which makes gradient descent almost along the same direction. We select MinMaxScaler method to normalize the original price data, calculated as in Eq. (8). Finally, all test results need to make inverse normalization so that we can be easily able to observe the proximity to the original data.

$$\widehat{x_t} = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \tag{8}$$

Where, $\widehat{x_t} \in [0, 1]$, x_t is the normalized and the actual value at time t, severally; $x_{\min}$, $x_{\max}$ is minimum and maximum value of the original sequence, respectively.

3. Deep model

3.1. BiLSTM

Recurrent neural network (RNN) plays an important role in the processing of a time-series. However, it tends to appear gradient disappearance or gradient burst [13,20]. LSTM coming into being, it was proposed by Schmidhuber et al, seen in literature [23]. General structure of deep LSTM can be seen in Fig.2.

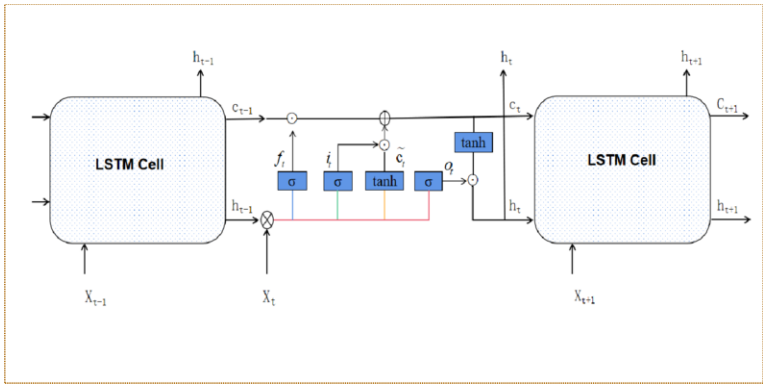


Fig. 2. The overall framework of LSTM neural networks.

As a forward network, LSTM can only excavate the characteristic of forward historical price data [24]. Notwithstanding, price data is not only dependent on information from the past, but also related to future financial conditions. BiLSTM is the expansion of LSTM that combines the forward and backward processing to capture useful information [25,26]. Hence, it can be performed to make the forecast more precise. The framewok of BiLSTM is shown as in Fig.3:

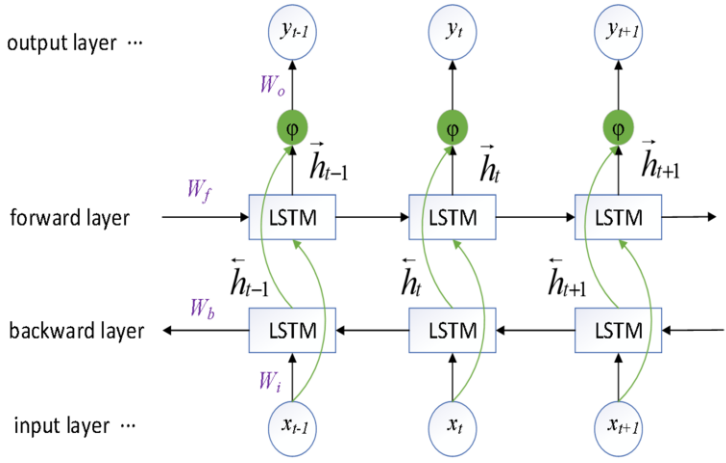


Fig. 3. The overall framework of deep BiLSTM network.

Concretely, the forward hidden layer state $\overrightarrow{h_t}$ is received via the forward network unit where the historical price series $X = [x_1, x_2, ..., x_T]$ is input. Historical price series X, as input to the LSTM is employed to forecast future price series $X' = [x'_1, x'_2, ..., x'_T]$ as an input to the reverse network unit, it obtains the reverse hidden-layer state $\overleftarrow{h}_t$. The output y_t of the network can be obtained through $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$, and BiLSTM of the cells are calculated as follows:

$$\overleftarrow{h}_t = W_{x\overleftarrow{h}} x_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{t-1} + \overleftarrow{b}_h \tag{9}$$

$$\overrightarrow{h}_t = W_{x\overrightarrow{h}} x_t + W_{\overrightarrow{h}\overrightarrow{h}} \overrightarrow{h}_{t-1} + \overrightarrow{b}_h \tag{10}$$

$$y_t = \varphi(W_{\overrightarrow{h}y} \overrightarrow{h}_t + W_{\overleftarrow{h}y} \overleftarrow{h}_t + b_y) \tag{11}$$

Where, φ represents nonlinear activation function. $W_{x\overleftarrow{h}}$ represents the weight from the input x of the current neuron to $\overleftarrow{h}$ at this moment, $W_{\overleftarrow{h}\overleftarrow{h}}$ represents the weight from the state at the previous moment to the current state, and $\overleftarrow{b}_h$ represents the threshold value. Each of $W_{x\overrightarrow{h}}$, $W_{\overrightarrow{h}\overrightarrow{h}}$ and $\overrightarrow{b_h}$ has a similar meaning to the above specific representation. Additionally, b_y denotes the offset value, $W_{\overrightarrow{h}y}$ denotes the weight from $\overrightarrow{h}_t$ to the output, and $W_{\overleftarrow{h}y}$ denotes the weight from $\overleftarrow{h}_t$ to the output.

3.2. ResNet

Usually, neural network demonstrate a powerful learning ability to various features. However, the deeper networks trend to appear overfitting or underfitting and lead to low accuracy. Although we can take a regularization approach to handle it, network degradation phenomenon is along. In 2015, He et al. proposed a deep residual network through making innovative works to the traditional convolutional neural network that appended the shortcut connection [27], which greatly eliminates the thorny problem of too deep network for training. Assuming $f(x)$ is a nonlinear unit, and $h(x)$ is an objective function. When the dimensions of $f(x)$ and input vetor x are consistent, the raw mapping $h(x)$ can be directly represented as $f(x) = h(x) - x$. Otherwise, we need to expand through the autoencoder matrix $\overline{v}$ and then convert to shortcut connection form [28,29], which is represented mathematically as follows:

$$h(x) = f(x) + \overline{v}x \tag{12}$$

where $\overline{v}$ is required to satisfy the following condition:

$$\overline{v}, z = \underset{\overline{v}\in R^{m\times n}, z\in R^{n\times m}}{\arg\min} \|x - z(v(x))\|_2^2 \tag{13}$$

In this work, we employ a special residual connection, as is depicted in Fig.4(a). It can be expressed as:

$$H_i = F_i(G_{i-1}) \otimes ReLU(T(G_{i-1})) \otimes x \tag{14}$$

where $T(r)$ denotes the output of 1D CNN input for r, F_i is a mapping of hideeen layer i. Concretely, the range of filters for 1D CNN is [16,32], the kernel size of 1D CNN is 2, the stride is 1. For the concatenation compution, it means if $\eta^{(1)} = (\eta_1, .., \eta_s)$, $\eta^{(2)} = (\theta_1, ..., \theta_t)$, then $\eta^{(1)} \otimes \eta^{(2)} = (\eta_1, .., \eta_s, \theta_1, ..., \theta_t)$. Particularly, the rectified linear unit (ReLU) is defined:

$$ReLU(t) = \begin{cases} t \ if & t \geq 0, \\ 0 \ if & t < 0. \end{cases} \tag{15}$$

3.3. BiLSTM-ResNet model

In contrast, the hybrid networks have greater generalization ability than single networks. Regarding time-series prediction, the LSTM is one of the most widely used models. In this work, we construct the BiLSTM-ResNet network, considering BiLSTM as a benchmark framework. As shown in Fig.4(b), behind the second hidden layer, the low and high dimensional features are fused to prevent the degradation of the model.

The historical price series is as input, and the data transmitting for each hidden layer is as follows. After a mapping F, feature information is passed to the first hidden layer. There is a residual connection behind the first hidden layer to prevent the disappearance of the gradient during the back-propagation, and its output is

$$H_1 = F_1(\bar{v}_1 x + B_1) \tag{16}$$

where B_1, $\bar{v}_1$ denote bias and weight factors, severally.

The network continues to learn, and the next one is mathematically expressed as

$$H_2 = F_2(\bar{v}_2 H_1 + B_2) \tag{17}$$

$$H_3 = F_3(\bar{v}_3 H_2 \otimes ReLU(T(H_1)) \otimes (v_3 x) + B_3) \tag{18}$$

where the low and high dimensional features are fused to prevent the degradation of the model, the concatenation is same as the second hidden layer.

Particularly, in the every hidden layer, the Kaiming-uniform is used to initialize the weights and thresholds. At the same time, we employ the L_2 regularization which adds an item to the objective function to reduce the feature weight, producing a smooth effect [30]. Concretely, the mathematical expression is as follows:

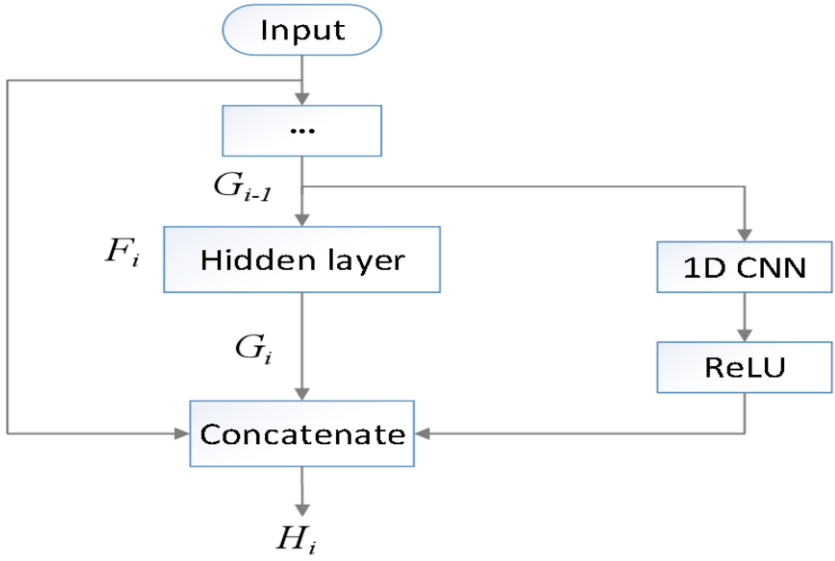
$$G(\bar{v}) = \frac{1}{L} \sum_{l=1}^{L} (\bar{v}^T x_j - y_j)^2 + c\|\bar{v}\|_2^2 \tag{19}$$

where $G(\bar{v})$ is the objective function, c is a smaller attenuation factor, L is the number of layers.

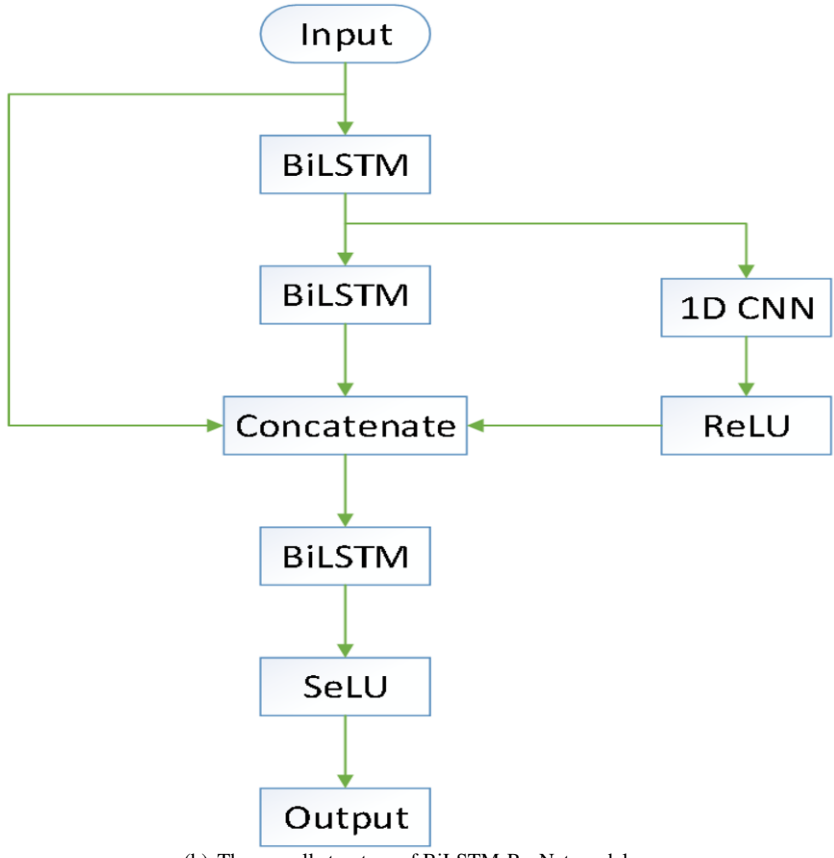
Lastly, the scaled exponential linear unit (SeLU) is choosen to be output activation function of the dense layer [26]. The formula is expressed as in Eq. (20). The predicted value is calculated as shown in Eq. (21).

$$SeLU(t) = \begin{cases} \kappa t & if \quad t > 0, \\ \kappa\mu(e^t - 1) & if \quad t \leq 0. \end{cases} \tag{20}$$

$$y_p = SeLU(\bar{v}_4 H_3 + B_4) \tag{21}$$



(a) The structure of ResNet network



(b) The overall structure of BiLSTM-ResNet model

Fig. 4. The general framework of the hybrid model.

4. Experiment and analysis

4.1. Problem description

In the financial market, the time series cannot be considered as absolutely continuous due to a certain time interval in the trading range [31,32]. If a discrete stock index price sequence is recorded as: $y_1, y_2, ..., y_n$, the nonlinear fluctuation process f can be simulated by building a appropriate prediction model in the light of the observed historical data and selecting the optimal parameters. The final goal of predicting the y_t is reached, as shown in formula (22).

$$y_t = f(y_{t-d}, y_{t-d-1}, ..., y_{t-d-m+1}) + p(t) \tag{22}$$

where, d, m and $p(t)$ is the delay duration, the time span to be considered and the noise in the observed data at time t, severally.

In this study, we select three price datasets of nonferrous metals(including aluminum, copper and zinc) on Shanghai Futures Exchange (SHFE) to validate the stability of the established model. The data are obtained from the website https://al.iyunhui.com/market/analysis/, as shown Fig.5. For a more ideal fitting effect, all data are normalized. The data set is divided according to ratio 9:1, and transaction data for the first 21 days corresponds to a price on the next day. We carry out all the experiments via employing TensorFlow 2.1.0 under the running environment of Windows 10.

4.2. Evaluation indicators of the forecast results

The common Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are employed to estimate the effectiveness of proposed model [13]. They are calculated by the following formula:

$$RMSE = \sqrt{\frac{1}{N_{test}} \sum_{n=1}^{N_{test}} (\widehat{y}_n - y_n)^2} \tag{23}$$

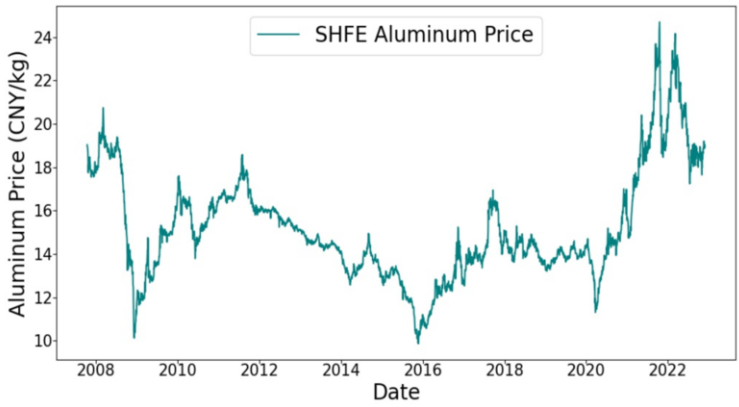
$$MAE = \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} |\widehat{y}_n - y_n| \tag{24}$$

$$MAPE = \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} |\frac{\widehat{y}_n - y_n}{\widehat{y}_n}| \times 100\% \tag{25}$$

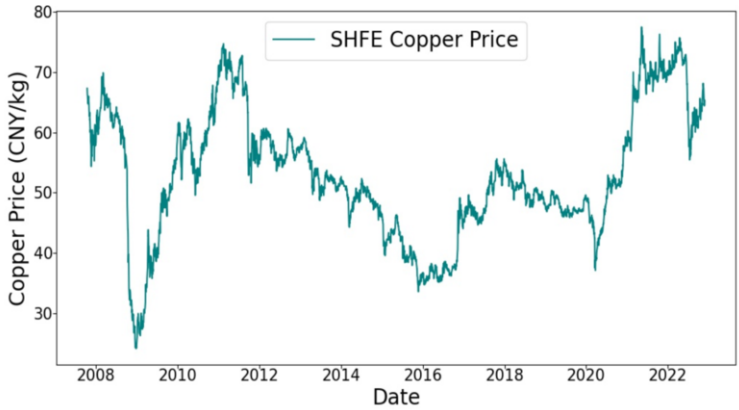
Where y_n, $\widehat{y}_n$ are the raw data and the predicted value, respectively; N_{tset} indicates the size of the test set.

4.3. The determination of parameters

In the experiment, we make the initialization of weights and bias through Kaiming uniform distribution. In addition, the batch size is firstly determined to be 16, the initial



(a) The change of aluminum price



(b) The change of copper price

(c) The change of zinc price

Fig. 5. The nonferrous metals price on SHFE from 16th October 2007 to 25th November 2022.

epochs is 100, the initial attenuation factor of L_2 regularization is 0.01. While the range of hyperparameters is given in Table 1.

Reference to the relevant literature, the number of hidden neurons for the BiLSTM in hybrid model is assumed to be N_1-N_2-N_3. In the beginning, N_1=N_2=N_3=32 and the learning rate is 1×10^{-3}. The iterative search trial is performed based on the GWO algorithm. Take aluminum price as example, we can see that the mentioned model achieves the convergence point after 4 iterations in Fig.6. Furthermore, it is very friendly to judge that the gray wolf optimization model works well. In the end, the optimal parameter combination is $N_1 = 16$, $N_2 = 22$, $N_3 = 19$, learning rate $= 2.74 \times 10^{-3}$, decay rate $= 6.07 \times 10^{-3}$, shown in Table 2.

Table 1. The range of hyperparameters in GWO algorithm

Specific parameter	Range
neurons in each hidden layer	16-32
L_2 regularization	0.01-0.1
learning rate	0.001-0.01
Nadam decay	0.0001-0.01

Table 2. Related parameters of the proposed model

Parameters	Value
neurons in the hidden layer	16-22-19
Nadam learning rate	2.74×10^{-3}
Nadam decay rate	6.07×10^{-3}
L_2 regularization factor	0.01
wavelet base	db2
epochs	100
time-step	21
batch-size	16

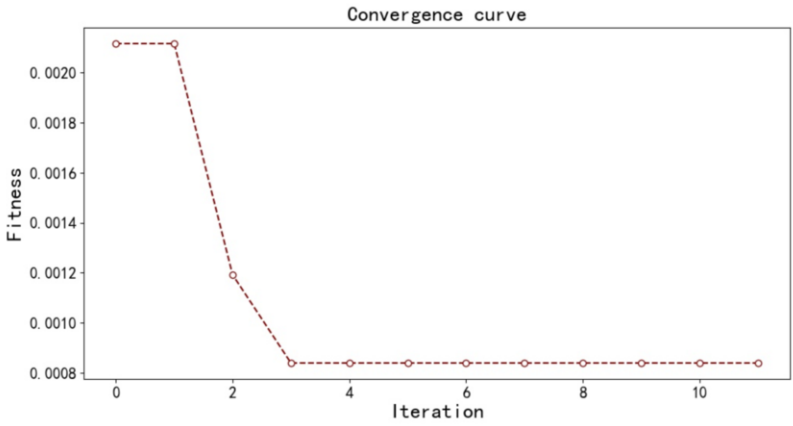


Fig. 6. Convergence diagram of gray wolf optimization model.

4.4. Data preprocessing

Take aluminum price as example, we use the wavelet transform to remove outliers in the original data, as shown in Fig.7. The original sequence is decomposed into three parts: one trend component and two detail components. Obviously, the denoising data curve is exceedingly smooth better than the original data curve. Here, the detail parts are still high frequency signal containing the details or differences of the signal.

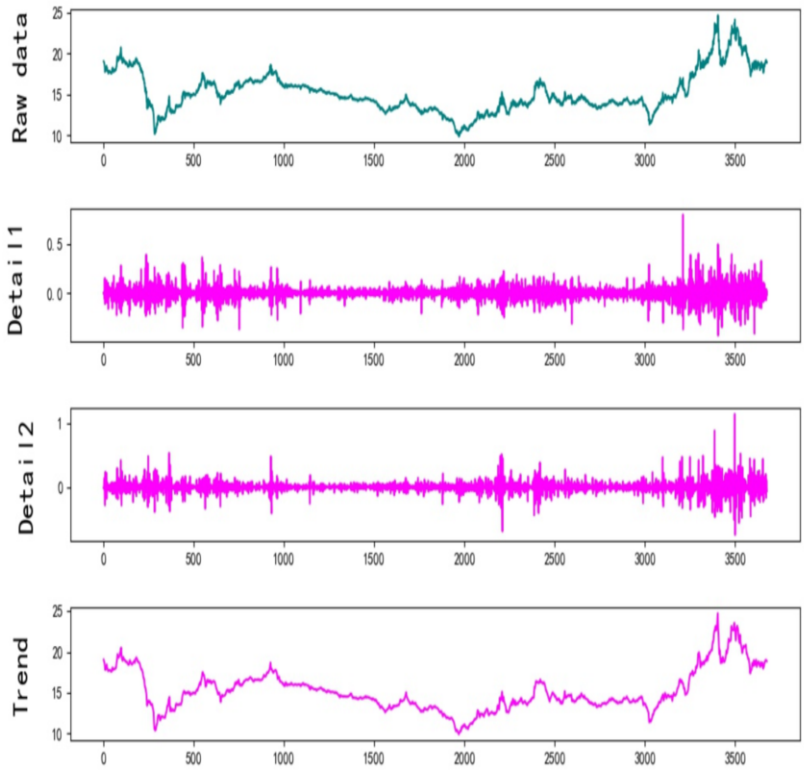


Fig. 7. The wavelet transform result of aluminum price.

4.5. Empirical results and analysis

To demonstrate the validity of the proposed model, we performed a series of experiments, with the forecast schemes as follows: LSTM, BiLSTM, BiLSTM-ResNet, BiLSTM-ResNet-DWT and BiLSTM-ResNet-GWO-DWT.

The Fig.8 illustrate the comparison graph and bar graph of error, which visually show the forecast accuracy. In the case of aluminum price, the single model BiLSTM

performs better than LSTM. However, the hybrid model BiLSTM-ResNet outperforms all single models in RMSE, MAE and MAPE. While the wavelet transform is used to decompose the complex nonlinear price sequence into multiple subsequences, it presents more obvious trend and fluctuation characteristic. Further, if the hyperparameters are optimized, the prediction accuracy is improved again.

Quantitative analyses are performed, and the RMSE, MAE, and MAPE for each method on the three datasets are shown in Table 3-5. Take aluminum as an example, the RMSE, MAE and MAPE of BiLSTM-ResNet are 0.3458, 0.2469 and 1.2133%, which is significantly better than BiLSTM and LSTM. This means that BiLSTM-ResNet improves the generalization ability by constructing the residual blocks. The RMSE, MAE and MAPE of BiLSTM-ResNet-DWT are 0.3320, 0.2404 and 1.1681%, respectively. Additionally, the RMSE, MAE and MAPE of BiLSTM-ResNet-GWO-DWT are 0.2672, 0.1935 and 0.9434%, concretely. This is ascribed to the fact that GWO algorithm has optimized the model by searching the optimal parameters for the neural network. Being compared with four other methods, a series of experiments results fully demonstrate the effectiveness of the BiLSTM-ResNet-GWO-DWT method.

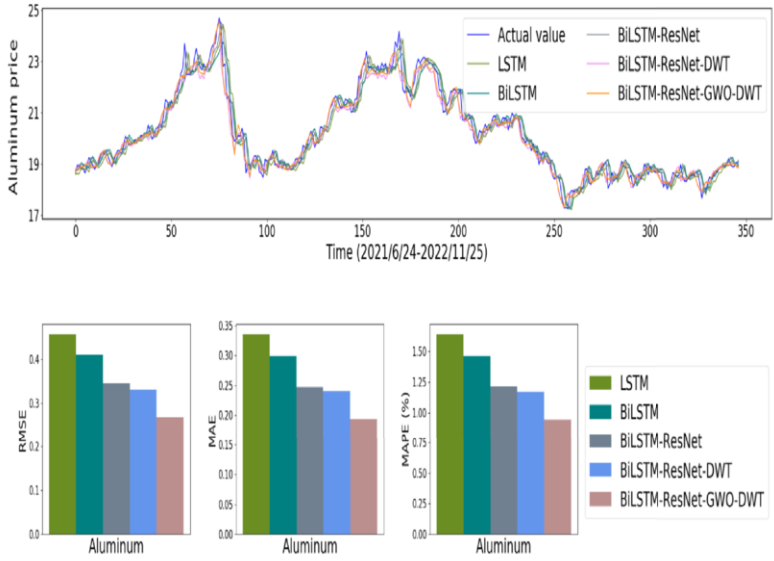


Fig. 8. The prediction result of aluminum price.

Table 3. The metrics of aluminum price prediction with various model.

Models	epochs=100		
	RMSE	MAE	MAPE
LSTM	0.4574	0.3354	1.6413%
BiLSTM	0.4105	0.2989	1.4653%
BiLSTM-ResNet	0.3458	0.2469	1.2133%
BiLSTM-ResNet-DWT	0.3320	0.2404	1.1681%
BiLSTM-ResNet-GWO-DWT	0.2672	0.1935	0.9434%

Table 4. The metrics of copper price prediction with various model.

Models	epochs=100		
	RMSE	MAE	MAPE
LSTM	1.2919	0.9881	1.4632%
BiLSTM	0.9363	0.7246	1.0793%
BiLSTM-ResNet	0.8461	0.6452	0.9598%
BiLSTM-ResNet-DWT	0.6347	0.4670	0.6950%
BiLSTM-ResNet-GWO-DWT	0.5780	0.4209	0.6254%

Table 5. The metrics of zinc price prediction with various model.

Models	epochs=100		
	RMSE	MAE	MAPE
LSTM	0.5326	0.3796	1.5428%
BiLSTM	0.5061	0.3690	1.4972%
BiLSTM-ResNet	0.4399	0.3050	1.2414%
BiLSTM-ResNet-DWT	0.4073	0.2809	1.1392%
BiLSTM-ResNet-GWO-DWT	0.3092	0.2171	0.8839%

5. Conclusion

For the industry, the price forecast of nonferrous metals is a meaningful work. In this work, a deep learning framework based on BiLSTM-ResNet-GWO-DWT is proposed. The qualitative and quantitative analysis results demonstrate that BiLSTM-ResNet-GWO-DWT model is better than other benchmark models. Moreover, because the proposed method takes the irregularities of price sequence into account, it further illustrates the validity of the model.

6. Future works

These prediction methods perform well in nonferrous metals price forecast, but at the same time, we also need to consider some shortcomings of this hyperparameter search

and data preprocessing. Firstly, applying GWO algorithm can reduce falling into the local optimum, while it also faces the trouble of high computational cost. It is believed that with the continuous advancement of research, there will be more new optimization algorithms in the future, and this problem will be improved. Secondly, in spite of the wavelet transform can remove the noise interference from the original data, it pays the cost of losing part of the real information. Hence, we will focus on the preprocessing of original data to fully mine the essential feature and enhance the prediction performance as for the future work.

Acknowledgements

This work was supported by Guizhou Provincial Science and Technology Projects(No. QKHJC-ZK[2023]YB036), Guizhou Provincial Science and Technology Projects (No. QKHJC-ZK[2021]YB017), Guizhou Provincial Education Department Higher Education Institution Youth Science Research Projects (QJJ[2022]098).

Declaration of interest statement

No conflict of interest.

Availability of data

Data available if required.

References

[1] Diego Garcła, Werner Kristjanpoller. An adaptive forecasting approach for copper price volatility through hybrid and non-hybrid models. Applied Soft Computing. 2019, 74:466-478.
[2] Alquist, Ron, Saroj Bhattarai, Olivier Coibion. Commodity-price comovement and global economic activity. Journal of Monetary Economics. 2020, 112:41-56.
[3] Cohen Gil. Algorithmic Strategies for Precious Metals Price Forecasting. Mathematics. 2022, 10(7):1134.
[4] Kriechbaumer, Thomas, et al. An improved wavelet–ARIMA approach for forecasting metal prices. Resources Policy. 2014, 39:32-41.
[5] Lasheras, Fernando Sánchez, et al. Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. Resources Policy. 2015, 45:37-43.
[6] Yaziz Siti Roslindar,Zakaria Roslinazairimah,Suhartono. ARIMA and Symmetric GARCH-type Models in Forecasting Malaysia Gold Price. Journal of Physics: Conference Series. 2019, 1366(1):012126.
[7] Yang Y, Hou M, Luo J, et al. Numerical solution of several kinds of differential equations using block neural network method with improved extreme learning machine algorithm. Journal of Intelligent and Fuzzy Systems. 2020, 38(3):3445-3461.
[8] Carrasco, R., Astudillo, G., Soto, I., Chacon, M., Fuentealba, D.,Forecast of copper price series using vector support machines. In: 2018 7th Int. Conf. Ind. Technol. Manag. 2018, 380-384.
[9] Shaukat K, Luo S, Abbas N, et al. An analysis of blessed Friday sale at a retail store using classification models. The 4th International Conference on Software Engineering and Information Management. 2021, 193-198.

[10] Kumar Chandar Sivalingam,Sumathi Mahendran, Sivanandam Natarajan. Forecasting Gold Prices Based on Extreme Learning Machine. International Journal of Computers Communications and Control. 2016, 11(3).
[11] Shao Bilin, Li Maolin, Zhao Yu, Bian Genqing. Nickel Price Forecast Based on the LSTM Neural Network Optimized by the Improved PSO Algorithm. Mathematical Problems in Engineering, 2019.
[12] Hou Muzhou, Liu Taohua, Yang Yunlei, et al. A new hybrid constructive neural network method for impacting and its application on tungsten price prediction. Applied Intelligence. 2017, 47(1).
[13] Zhanglong Li, Yunlei Yang, Yinghao Chen, et al. A Novel Non-Ferrous Metals Price Forecast Model Based on LSTM and Multivariate Mode Decomposition. Axioms, 2023, 12(7):670.
[14] Chengshi Tian, Yan Hao. Point and interval forecasting for carbon price based on an improved analysis-forecast system. Applied Mathematical Modelling. (2019).
[15] Batool D, Shahbaz M, Shahzad Asif H, et al. A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning. Plants. 2022, 11(15):1925.
[16] Gligoric Z, Gligoric M, Halilovic D, et al. Hybrid stochastic-grey model to forecast the behavior of metal price in the mining industry. Sustainability. 2020, 12(16):6533.
[17] Yan Hu,Jian Ni,Liu Wen. A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction. Physica A: Statistical Mechanics and its Applications. 2020, 557.
[18] Liu Kailei,Cheng Jinhua,Yi Jiahui. Copper price forecasted by hybrid neural network with Bayesian Optimization and wavelet transform. Resources Policy. 2022, 75.
[19] Zhang Chu,Hua Lei,Ji Chunlei,Shahzad Nazir Muhammad,Peng Tian. An evolutionary robust solar radiation prediction model based on WT-CEEMDAN and IASO-optimized outlier robust extreme learning machine. Applied Energy, 2022, 322.
[20] Zhongyu Li , Hongxia Ge , Rongjun Cheng. Traffic flow prediction based on BILSTM model and data denoising scheme. Chinese Physics B. 2022, 31:040502.
[21] Seyedali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis. Grey Wolf Optimizer. Advances in Engineering Software, 2014, 69.
[22] Jinxin Pan, Bo Jing, Xiaoxuan Jiao, Shenglong Wang. Analysis and Application of Grey Wolf Optimizer-Long Short-Term Memory. IEEE Access. 2020, 8.
[23] Sepp Hochreiter, Jürgen Schmidhuber. Long short-term memory. Neural computation. 1997, 9(8):1735-1780.
[24] Wei, Wangyang, Honghai Wu, Huadong Ma. An autoencoder and LSTM-based traffic flow prediction method. Sensors. 2019, 19(13):2946.
[25] Yi Peng et al. Meteorological Satellite Operation Prediction Using a BiLSTM Deep Learning Model. Security and Communication Networks. 2021, 1-9.
[26] Min-seung Ko, Kwangsuk Lee, Chang Woo Hong, et al. Deep Concatenated Residual Network with Bidirectional LSTM for One-Hour-Ahead Wind Power Forecasting. IEEE Transactions on Sustainable Energy. 2020, 12(2):1321-1335.
[27] He Kaiming, Sun Jian, et al. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, 770-778.
[28] Chen Yinghao, Wang Dongdong, Hou Muzhou, et al. Prediction of safety parameters of pressurized water reactor based on feature fusion neural network. Annals of Nuclear Energy. 2022, 166:108803.
[29] Krizhevsky, Alex, Ilya Sutskever, Geoffrey E., Hinton. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 2017, 60(6):84-90.
[30] Xue Yan, Wang Weihan, Miao Chang. Research on financial assets transaction prediction model based on LSTM neural network. Neural Computing and Applications. 2020, 33(1):257-270.
[31] Liu, Huilan, Junjie Ma, Changgen Peng. Shrinkage estimation for identification of linear components in composite quantile additive models. Communications in Statistics-Simulation and Computation. 2020, 49(10):2678-2692.
[32] Pengfei Yu, Xuesong Yan. Stock price prediction based on deep neural networks. Neural Computing and Applications. 2020, 32(6):1609-1628.