

## Research Article

# A Hybrid LSTM-Based Ensemble Learning Approach for China Coastal Bulk Coal Freight Index Prediction

Wei Xiao <sup>1,2,3</sup> Chuan Xu <sup>1,2,3</sup> Hongling Liu <sup>1</sup> and Xiaobo Liu <sup>1,2,3</sup>

<sup>1</sup>School of Transportation and Logistics, Southwest Jiaotong University, No. 111 Erhuanlu Beiyiduan, Chengdu 610031, China

<sup>2</sup>National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu, Sichuan, China

<sup>3</sup>National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Chengdu 611756, Sichuan, China

Correspondence should be addressed to Chuan Xu; [xuchuan@swjtu.edu.cn](mailto:xuchuan@swjtu.edu.cn)

Received 19 February 2021; Revised 13 April 2021; Accepted 13 May 2021; Published 26 May 2021

Academic Editor: Dong-Kyu Kim

Copyright © 2021 Wei Xiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

China Coastal Bulk Coal Freight Index (CBCFI) reflects how the coastal coal transporting market's freight rates in China are fluctuated, significantly impacting the enterprise's strategic decisions and risk-avoiding. Though trend analysis on freight rate has been extensively conducted, the property of the shipping market, i.e., it varies over time and is not stable, causes CBCFI to be hard to be accurately predicted. A novel hybrid approach is developed in the paper, integrating Long Short-Term Memory (LSTM) and ensemble learning techniques to forecast CBCFI. The hybrid LSTM-based ensemble learning (LSTM-EL) approach predicts the CBCFI by extracting the time-dependent information in the original data and incorporating CBCFI-related data, e.g., domestic and overseas thermal coal spot prices, coal inventory, the prices of fuel oil, and crude oil. To demonstrate the applicability and generality of the proposed approach, different time-scale datasets (e.g., daily, weekly, and monthly) in a rolling forecasting experiment are conducted. Empirical results show that domestic and overseas thermal coal spot prices and crude oil prices have great influences on daily, weekly, and monthly CBCFI values. And in daily, weekly, and monthly forecasting cases, the LSTM-EL approaches have higher prediction accuracy and a greater trend complying ratio than the relevant single ensemble learning algorithm. The hybrid method outperforms others when it works with information involving a dramatic market recession, elucidating CBCFI's predictable ability. The present work is of high significance to general commerce, commerce-related, and hedging strategic procedures within the coastal shipping market.

## 1. Introduction

Nowadays, 90% of global trades are completed via sea transportation [1]. Sea transport is critical to the business system on the globe and also in the domestic trade system. China Coastal Bulk Coal Freight Index (CBCFI) [2] is built for timely reflecting how coastal coal transporting market's freight rates in China are fluctuated, by complying with the present mechanism of China Coastal Bulk Freight Index (CBFI) [3]. This system can publicize the complex index and spot ratios in terms of various routes/kinds pertaining to vessels of the coastal coal service market on a day-to-day basis. China (Coastal) Bulk Freight Index Panelist offers CBCFI freight data per weekday (Shanghai Shipping

Exchange publicizes CBCFI at the official website and as well as <http://www.chineseshipping.com.cn> at 15:00 (Beijing Time) on each index publication day).

CBCFI represents the voyage-charter freight rate conditions pertaining to the market of bulk coal shipping in coastal areas. This indicates the unstable property pertaining to the coal bulk shipping market, as well as reflecting the developing states pertaining to the China economic condition and domestic business tendency. Thus, it refers to the “weatherglass” pertaining to the market of bulk coal shipping in coastal areas. Due to the mentioned feature, numerous internal personnel and specialists attempt at estimating the subsequent tendency pertaining to the market of bulk coal shipping in coastal areas by accurately

predicting the bulk coal freight index for guiding company strategy decisions. Additionally, forecasting CBCFI value enables operating personnel and decision-making personnel for managing market trends and avoiding risks inside the coastal coal shipping market. Furthermore, it helps industries and manufactories with the domestic shipping system.

Previous studies indicate that the shipping freight index usually presents complicated instability features to be uncertain, cyclical, and nonlinear properties [4–6]. To achieve the satisfactory freight index prediction, a variety of freight indices predicting methods developed previously, where the econometric time series predicting approaches exhibit drawbacks in nonstationary and nonlinear properties, while the single neural network (NN) methods show disadvantages in overfitting, local minimum point, and parameter selection problems. Because both single econometric and NN approaches are limited in freight index prediction, recently, empirical studies consistently demonstrate the adaptability of hybrid techniques. Hybrid approaches are capable of combining individual approaches and make the respective advantage remedy others' deficiencies.

Recently, ensemble learning algorithms are extensively employed for analyzing the multivariable prediction. Ensemble learning algorithms (e.g., random forest (RF) and gradient boosting regression tree (GBRT)) are effective in determining the essential variables of a time series and investigate the inner relations among variables [7]. Over the past few years, ensemble learning algorithms have been extensively employed for studying the mentioned time series to be stock prices, Baltic Dry Index, and traffic flow, leading to the production of essential outcomes [8–10].

Ensemble learning algorithms' fast advancement presents one novel idea for the way of utilizing multidata and improves the readability from the original data. We here combine AI and ensemble learning algorithms for formulating one emerging hybrid approach, an ensemble learning (LSTM-EL) approach (e.g., LSTM-GBRT and LSTM-RF) by exploiting Long Short-Term Memory (LSTM) for CBCFI forecasting. Inside the approach, the LSTM layer obtains details dependent on time within the data, and the GBRT/RF layer shows great robustness of ensemble learning in terms of the training of approach.

The present work has the following organization. Section 2 presents the review of the analyses of the market of shipping freight. Section 3 presents the designed approaches. Section 4 introduces the data collection. In Section 5, the approach performance receives the comparison and analysis. Lastly, Section 6 gives the concluding remarks.

## 2. Literature Review

Since freight rates are uncertain and unstable, the approaches to quantitatively analyze the rates arouse long concern from the shipping industry. As a result, increasing literature proposes approaches to predict freight rates, many of which use the Baltic Dry Index (BDI) [11]. Both BDI and CBCFI are shipping indices that effectively assess the current situation or shipping market, and they have some similar features. First, CBCFI and BDI are both daily number-issued

leading indicators that measure the costs of shipping raw materials. Specifically, BDI mainly measures shipping costs for dry bulk commodities, including coal, grain, iron ore, and metals, while CBCFI mainly measures shipping costs for coal; second, in terms of the composition, BDI takes 23 shipping routes measured on a time charter and voyage basis while CBCFI takes 14 shipping routes measured on a voyage basis; third, as for the type of vessel, BDI looks at the ships that can carry 15,000 deadweight tons (DWT)–80,000 DWT of cargo (90 percent of the global fleet), while CBCFI focuses on the ships that can carry 15,000 DWT–60,000 DWT [3, 12, 13]. As the BDI and CBCFI share these similar features, we refer to the abundant research of the BDI to obtain a clear view of current prediction approaches. The analytic methodology of BDI among others is split into three types.

Traditional econometric approaches are initially covered, including vector error correction (VEC), generalized autoregressive conditional heteroskedasticity (GARCH), vector autoregression (VAR), and the autoregressive integrated moving average (ARIMA) approaches. Cullinane et al. [14] first developed a Baltic Dry Bulk Index (BDI) study approach method through the ARIMA approach. Kavusanos and Alizadeh-M [15] developed one season-related ARIMA approach for one independent variant as well as one VAR approach for investigating seasonality in the dry bulk shipping market. Batchelor et al. [16] discussed the performances of VAR and ARIMA as well as vector equilibrium correction (VECM) approaches for the prediction of spot and forward freight rates. To improve the accuracy of BDI forecasts, Tsioumas et al. [17] developed a multivariate vector autoregressive approach with exogenous variables (VARX) approach and the results demonstrate that the VARX approach outperforms the ARIMA approach. Chen et al. [18] applied the ARIMA and the VAR approaches to predict the spot rates of several dry bulk routes and found out that the VAR approach performed better than the ARIMA approach in test-sample forecasts. Adland et al. [19] showed a cointegration-based method for analyzing regional ocean freight rates' dynamic properties.

According to Stopford [20], sea prediction is difficult for statistical and traditional econometrical approaches for capturing the nonlinearity of dry bulk freight rates [21]. Accordingly, a second type of analytic approach currently employs several nonlinearity and artificial intelligence (AI) approaches, comprising artificial neural network (ANN), machine learning algorithm, and nonlinear methods. Li and Parsons [22] applied neural systems to predict monthly tanker freight rates from the short to the long term and found the neural systems outperforming the ARIMA time sequence approach for prediction in the long run. Yang et al. [23] investigated and predicted the freight rate instability alarming pertaining to China Coastal Bulk Freight Index (CCBFI), China Containerized Freight Index (CCFI), and Baltic Freight Index (BFI) by exploiting support vector machine (SVM). Thalassinou et al. [24] adopted a chaos approach for predicting the BDI using the invariable coefficients pertaining to the strange attractor under reconstruction, governing the system's evolving process. Guan

et al. [25] developed one SVM-based multistep prediction approach to predict weekly Baltic Supermax Index (BSI) data. Şahin et al. [26] develop three ANN approaches based on BDI data and results show that their performances are close, whereas the highest consistency pertains to the ANN by exploiting the past two weekly observations of the BDI data. Among the deep-learning approaches, the long short-term memory (LSTM) neural network is regarded as a practical technique for handling time series problems [27–29]. For example, Nelson et al. [27] employed the LSMT network to predict the future trends of stock prices based on the historical price, alongside technical analysis indicators; Duan et al. [30] implemented the LSTM model for multistep ahead travel time prediction. The above studies demonstrated that NN methodologies generate more effective forecasting performance than conventional time series and econometric approaches.

Though NNs are capable of handling nonlinearity and show better robustness, it is hard to determine the configuration of the NN algorithm; in addition, it tends to fall into over or lacked training, easily resulting in local minimum trapping [11]. This prompts one shift into the third category hybrid methodologies, which usually integrates one noise-reducing method by adopting one NN-based algorithm. Leonov and Nikolov [31] proposed a hybrid approach of wavelets and neural networks to investigate the fluctuation in the freight rates of the Baltic Panamax routes 2A and 3A. Bulut et al. [32] established one vector autoregressive fuzzy combined logical predicting approach in terms of time charter rates. Zeng et al. [33] propose a hybrid approach of empirical mode decomposing process (EMD) as well as ANN. Uyar et al. [34] presented one trained recurrent fuzzy neural system with the use of a genetic algorithm for improving long-term dry cargo freight rates prediction to be more accurate. As an efficient strategy to improve the forecasting ability of a single model, ensemble learning has been widely used to improve the model performance [35–38]. For example, Kamal et al. [35] developed a deep ensemble recurrent network of recurrent neural network (RNN), long-short-term memory (LSTM), and gated rectified unit neural network (GRU) to improve the BDI predictive performance, and results showed that the ensemble method outperforms the single deep-learning approach. Tan et al. [37] proposed an LSTM-based deep ensemble learning model that combined bagging, random subspace, and boosting to forecast ultra-short-term industrial power demand, and they found out that the proposed model obtains higher accuracy and robustness than LSTM, eXtreme Gradient Boosting (XGBoost), and other time series methods. Liu et al. [36] proposed a deep air pollution model for forecasting PM<sub>2.5</sub> concentrations based on a wind-sensitive attention mechanism, LSTM, and XGBoost, and experiments illustrate that the proposed approach is superior to a single multilayer perceptron (MLP), SVM, LSTM, and XGBoost. The above studies indicate that models combining machine learning and neural networks usually obtain better predictive results than any single model.

Although much research has been conducted to solve the shipping indices forecasting issues, most do not consider

different spot rates or other related factors information. Accordingly, these methods cannot effectively reflect the critical factors that contribute to strengthening the predictive performance. To address this issue, we propose a hybrid model of LSTM and ensemble learning algorithm to handle the CBCFI forecasting problems. The LSTM approach has the ability to acquire the data determined by time and noticeably impacts the predicting process of time series, whereas this approach fails to appropriately mine the implicit relations between exogenous variables in predicting inflection point data. Thus, to better utilize feature information, it is necessary to incorporate an applicable ensemble learning algorithm to optimize the feature combination in order to construct the feature set that reflects the short/long-term trend of the CBCFI. For the ensemble learning part, boosting and bagging are two of the important ideas; they both combine a set of weak learners to create a strong learner that obtains better performance. In all kinds of machine learning methods, Gradient Boosting Regression Tree (GBRT) [39] and Random Forest (RF) [40] have received much attention and are often used as representatives of ensemble learning algorithms. GBRT and RF are decision-tree-based ensemble learning algorithms that use a boosting and bagging framework, respectively. Because of the implementation of a gradient boosting algorithm or a bootstrap sampling method, GBRT or RF can handle variables fast, making it suitable for complicated tasks. Accordingly, this study adopts GBRT and RF as ensemble learning algorithms for CBCFI forecasting.

### 3. Data Description

This section firstly presents the data source. As this dataset includes both historical freight index and other impacted variables, variable correlations are introduced and conducted in this section as well.

*3.1. Data Source.* All datasets in the present study are supported by the Wind Economic Database [41]. Wind pairs more than 1.3 million macroeconomy-related and industry time series based on effective graphics and data study equipment for elucidating China's economy to finance-related professionals.

The China (Coastal) Bulk Coal Freight Index (CBCFI) is a compound weekday index (null on holiday and weekend) that considers 14 shipping routes [2]. It takes 1st September 2011 as the base period, and the base index is 1,000 points. Now, CBCFI is extensively employed by practitioners of industry and considered a vital economic indicator in coastal coal transportation. Figure 1 displays the composite routes of CBCFI and its corresponding weights. It can be seen that CBCFI mainly describes the shipping market of transporting coal from the north to the south. According to the histogram at the bottom of Figure 1, we notice that six routes whose weights are over 10% hold dominant positions, in which most of them depart from Qinhuangdao Port. For easily reading, in Figure 1, the dominant routes are plotted in red, and for those weight values less than 10% but nonzero,

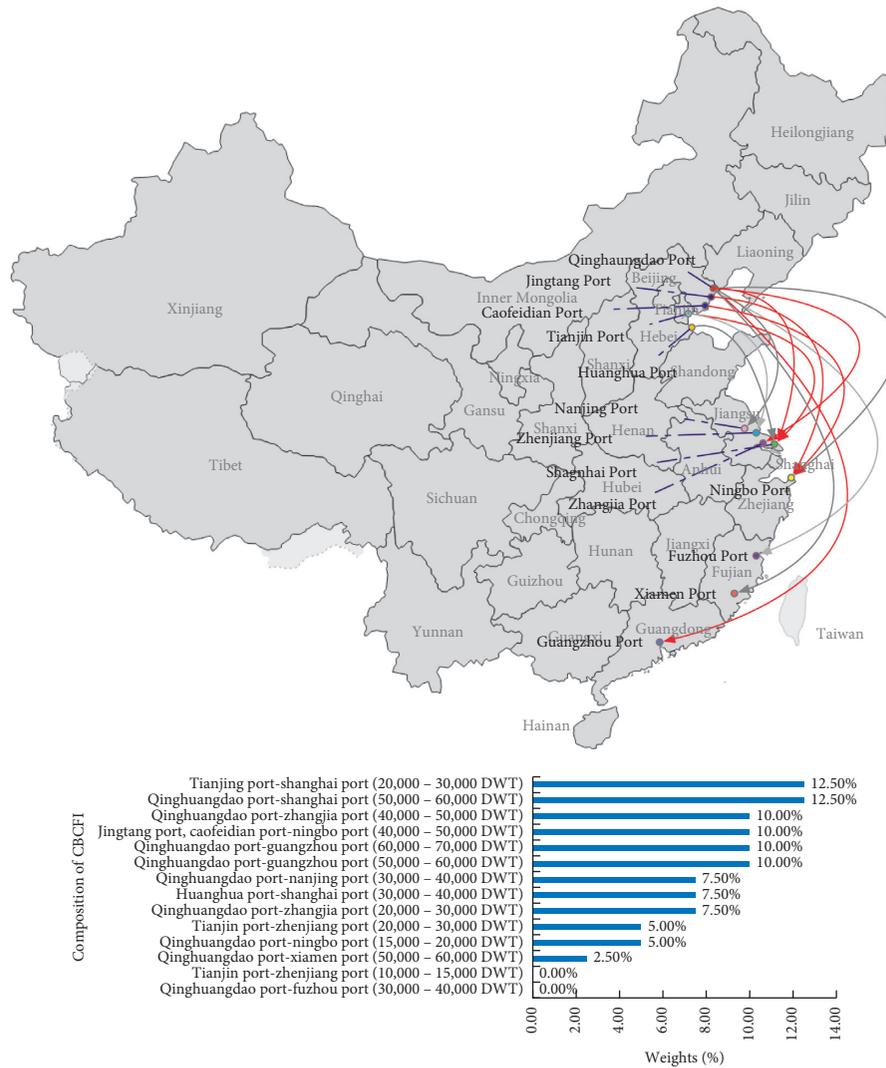


FIGURE 1: Illustration of the CBCFI composition and the corresponding shipping routes (DWT in bracket represents the “deadweight tonnage (DWT)” of the ship).

routes are marked in dark grey, while those with zero weights are in light grey.

Previous studies of traditional shipping index (e.g., BDI) generally discuss the daily, weekly, and monthly forecasting cases to evaluate the performance of approaches in short-, mid-, and long-term predictions [11, 16, 22]. Thus, in this paper, daily, weekly, and monthly forecasting experiments are conducted and the full available CBCFI data from January 2012 to October 2020 are employed as the experimental datasets, with 2129 daily observations, 422 weekly observations, and 10 yearly observations.

According to Figure 2, the CBCFI fluctuations exhibit irregular properties and high frequencies as well as large amplitudes. In the long run, from 2012 to the first half of 2016, CBCFI had a few inflection points. And in the fourth quarter of 2013, CBCFI reached the highest value of 1450.39 since its release and then returned to the low level at the end of the year. From the second half of 2016 to now, CBCFI has experienced a large fluctuation range, especially from 2017 to 2020. Compared with the previous years, the rise and fall of

freight rates have been much larger, and the frequency of rise and fall also increases obviously, which reflects the changeable market situation of coastal coal transport. Moreover, the global Coronavirus Disease 2019 (COVID-19) outbreak has generated a public health crisis that started in December 2019, which has certain consequences for the economics of the shipping industry. For seasonal fluctuations, from the trend of each year, except for 2014, CBCFI shows the phenomenon that its index rises in summer and winter in other years, but the increase rate is different each year. The main reason is that demand for coal tends to rise in summer and winter, traditionally the peak seasons for electricity use, leading to a seasonal rise in coal transport prices. Although not every summer or winter in the past 7 years has seen an increase in rates, historical statistics indicate a high probability of rates rising in both summer and winter each year.

To justify the uncertain and nonlinear characteristics of the CBCFI series, descriptive statistics of the daily CBCFI and its rate of change are provided, and the results are listed

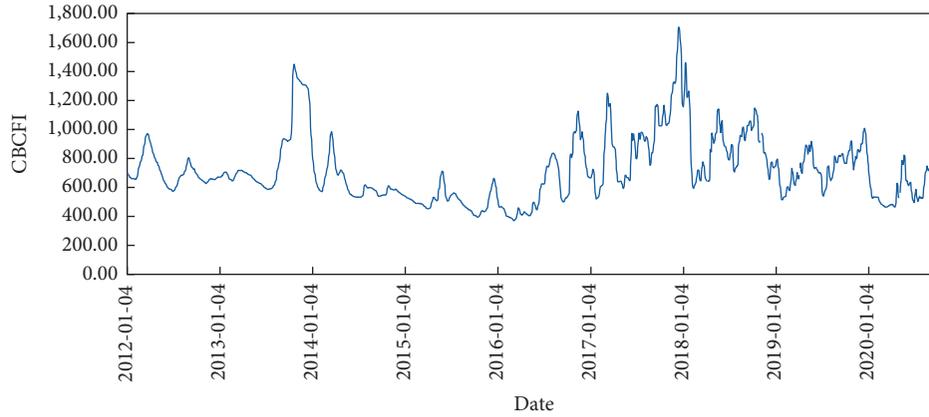


FIGURE 2: Tendency of daily CBCFI dataset adopted for predicting process.

in Table 1. The rate of change for the CBCFI is  $R_t = ((y_t - y_{t-1})/y_{t-1})$ .

From the above statistics, it can be observed that, in terms of the skewness and kurtosis, the skewness of CBCFI series and its rate of change series are greater than zero, and the kurtosis of each is either less than or greater than three. Therefore, the daily CBCFI and its rate of change series are left-skewed and right-skewed, respectively, which indicates that they have the characteristics of a sharp peak and a thick tail. In addition, the J-B test results indicate that both time series do not follow a normal distribution because they reject the hypothesis of the Jarque-Bera (J-B) statistic at 5% significant level. Moreover, both daily CBCFI and its rate of change series are indicated to be nonstationary, based on the results of the augmented Dickey-Fuller (ADF) unit root test; specifically, the  $t$ -statistics values summarized in Table 1 are less than the value of the critics ( $-2.86$ ). Furthermore, the Ljung-Box statistics to the squared residuals (also known as Q-Statistics of Square Residuals) is applied to test the nonlinearity of the CBCFI and its rate of change series. For both series at lag 6, we reject the hypothesis of no autocorrelation at the 5% significance level. Autocorrelated squared residuals are indications of nonlinearity [2]. The above statistics indicate that the CBCFI has the uncertain, nonstationary, and nonlinear properties, and those complicated features justify the need for the proposed ensemble approach.

**3.2. CBCFI Impact Factor Data.** Within other markets, freight rates inside the shipping industry receive the formation based on the interacting processes pertaining to various factors, such as the prices of the transported cargo, the demand and supply, and the cost. Thus, if these elements influence the CBCFI, then they should be investigated. The demand for coastal transport arises from the need of exporters and importers to transport the coal to specifically domestic destinations. The “derived” demand is mainly affected by domestic economy and trade, such as the Import and Export Trade (IET), the industry production, domestic coal inventory, the contract rates, and spot rates of domestic thermal coal. As the domestic economy is improved,

TABLE 1: Descriptive statistics of daily CBCFI and rate of change of CBCFI.

Statistic	Daily CBCFI	Rate of change for CBCFI
Mean	722.6500	-0.0002
Median	675.4800	-0.0012
Standard deviation	223.8494	0.0228
Maximum	1,706.2000	0.1186
Min	370.9900	-0.1337
Kurtosis	1.7801	5.6084
Skewness	1.2186	0.2550
J-B	658.9484	626.4104
$p$ value	0.000	0.000
ADF [lags]	-4.5298 [12]	-11.3960 [12]
$p$ value	0.0130	0.0164
Q [lags]	26.571 [6]	15.95 [6]
$p$ value	0.000	0.000
Observation	2129	2128

*Note.* J-B represents the Jarque-Bera (J-B) test in which a goodness-of-fit test is performed for examining whether the time series follows a normal distribution [42]. ADF represents the augmented Dickey-Fuller (ADF) test with a null hypothesis of the existence of autocorrelation in the sample series [43]. Q represents the Ljung-Box statistics to the squared residuals (also known as Q-Statistics of Square Residuals) in which a nonlinearity test [44] under the null hypothesis of linearity and residuals of a properly specified linear model should be independent [45].

domestic trade will be promoted, and shipping transport will be more demanded. Moreover, random shocks based on emergency events (e.g., 2015 Tianjin explosions and 2019 COVID-19 outbreak) and cyclical and seasonal market movements of the coal transported by sea further substantiate that the demand for shipping transport depends on macroeconomic factors. Besides, transportation consumption and costs constitute other CBCFI determinates. On the promise of collecting data, the relevant data of the CBCFI are considered maximally, which could roughly be classified into three categories:

- (i) *Domestic and Overseas Thermal Coal Spot Prices.* It is worth mentioning that Qinhuangdao Port is the greatest coal export port on the globe, and the routes of CBCFI are the dominant carriers for thermal coal transportation from north to south. Thus, the spot

price of thermal coal is considered as one of the critical factors affecting the fluctuation of CBCFI. Furthermore, with the enlarged opening of the market, the influence of the international market on the domestic market will be increasingly intensified, and the correlation between the domestic and foreign markets will be further enhanced. Even at this stage, the price of thermal coal in the international market will affect the export of coal, which will be reflected through the price of thermal coal.

- (ii) *Coal Inventory.* Coal inventory refers to a critical element that impacts the coal price based on the economic new normal [46]. Coal inventory is the result of coal production, transportation, consumption, and other factors. It is basically the same as the factors of price formation and has a leading effect on the price change of thermal coal. In recent years, Qinhuangdao coal inventory has become the weathervane of thermal coal price change [47].
- (iii) *Fuel Oil and Crude Oil Prices.* On the one hand, fuel oil prices have a significant impact on coal prices as fuel oil prices influence the cost of shipping. When the price of fuel oil increases, the shipping cost will increase; when the price of fuel oil declines, the shipping cost will decrease. On the other hand, coal and crude oil are the most basic energy sources, and the sharp rise in oil prices has also contributed to the rise in thermal coal prices [48]. The influence of crude oil price on thermal coal price is reflected in the following aspects. First, crude oil, that is, one type of vital fossil energy source, refers to a dominant substitute to coal. Thus, crude oil price instability can impact coal demand and price in a corresponding manner [49]. When crude oil rises in price, the demand for coal would rise as the substitute. The fluctuation of coal price will influence the bulk coal shipping cost somehow and thus influence the CBCFI values. Note that the “prices” discussed in the present study contain the spot prices and futures prices (a spot price is an offer to complete a commodity transaction immediately, while a future contract locks in a price for future delivery). Table 2 summarizes the abovementioned factors in the basic feature set and will be optimized later. Variables are numbered sequentially for the feature description.

**3.3. Variables Correlation Analysis.** To give a clear and simple view of the correlations between CBCFI and relevant impact factors, the Pearson correlation coefficient [50] is calculated and the results are presented in Table 2. Domestic and overseas thermal coal spot rates ( $u_1 \sim u_8$ ) show obvious correlation with CBCFI, particularly, “Qinhuangdao Port-Q5000 Index-FOB,” “Jintang Port-Q5000 Index-FOB,” and “Guangzhou Port-Q5500 Index (Indian Coal)-EXT” with the largest Pearson correlation values; on the other hand, coal inventories ( $u_9 \sim u_{12}$ ) show negative correlations with the CBCFI series. Based on the results, we remove two

variables with the smallest Pearson coefficient value ( $u_9$  and  $u_{11}$ ), and the rest of the 22 variables are selected as the input variables.

In addition, to demonstrate the necessary property for selecting the ensemble learning algorithm to deal with the CBCFI prediction problem, the relationships between the 24 variables and CBCFI are checked. Then, one color-coded Pearson correlation matrix is generated. The numerical value one with the expression of dark blue indicates one overall positive linear correlation of two characteristics, whereas chartreuse indicates zero, demonstrating no linear correlation. As is shown in Figure 3, there is an interrelation of different degree between the 24 variables. For example, fuel oil spot and futures prices ( $u_{14} \sim u_{19}$ ) are highly correlated with crude oil spot and futures prices ( $u_{20} \sim u_{24}$ ). Therefore, the ensemble learning algorithms dealing with multidata relations are considered to solve the CBCFI prediction problem.

## 4. Methodology

In this section, we first give a problem statement to present an overview of the prediction problem researched in this work. Then, the core concept and the flowchart with algorithm pseudocode of the proposed hybrid model structure are presented. At last, the prediction accuracy measurements are described.

**4.1. Problem Statement.** The goal of this work is to predict the CBCFI values of the next day given historical data. We define the historical observations of the target CBCFI as  $Y = (y_1, \dots, y_t, \dots, y_T)^T \in \mathbb{R}^T$ , where  $T$  represents time window size and  $y_t$  is the CBCFI value at time  $t$ . Similarly, auxiliary factors are represented by  $X = (X_1, \dots, X_t, \dots, X_T)^T \in \mathbb{R}^{T \times D}$ , where  $D$  specifies the number of related factors.  $X_t \in \mathbb{R}^D$  is the values of all  $D$  related factors at time  $t$ , and  $X^d \in \mathbb{R}^T$  is the value of the  $d^{\text{th}}$  factor in time window  $T$ . Thus, the prediction target  $y_{T+1}$  could be defined as follows:

$$y_{T+1} = F(y_1, \dots, y_t, \dots, y_T; X_1, \dots, X_t, \dots, X_T), \quad (1)$$

where  $F(\cdot)$  is the mapping function we are aiming to learn.

**4.2. Preparing Data.** Given a CBCFI time series  $\tilde{s}$  with length  $N$ ,  $\tilde{s} = \{s(t_i)\}_{i=1}^N$ , the time series firstly should be pre-processed. Data cleaning and data normalization are the key data preparation tasks for further forecasting tasks. As the given time series has no missing value and to preserve the characteristics of real-world data, no noise reduction or data smoothing was performed on data. In this study, we only normalized the data with min-max normalization algorithm. The time series  $\tilde{s}(t_i)$  is normalized and the resulting normal data are expressed as  $s = \{s(t_i)\}_{i=1}^N$ :

$$s(t_i) = \frac{\tilde{s}(t_i) - \min(\tilde{s})}{\max(\tilde{s}) - \min(\tilde{s})}, \quad (2)$$

TABLE 2: Impact factors description and the corresponding Pearson coefficient values.

Variable category	Variable description	Feature number	Pearson coefficient values
Domestic and overseas thermal coal spot rates	Qinhuangdao Port-Q5500 Index-FOB (a spot price is an offer to complete a commodity transaction immediately, while a futures contract locks in a price for future delivery. It leaves its point of origin)	$u_1$	0.550
	Qinhuangdao Port-Q5000 Index-FOB	$u_2$	0.575
	Qinhuangdao Port-Q5500K Index-FOB	$u_3$	0.512
	Qinhuangdao Port-Q5000K Index-FOB	$u_4$	0.530
	Jintang Port-Q5500 Index-FOB	$u_5$	0.554
	Jintang Port-Q5000 Index-FOB	$u_6$	0.571
	Guangzhou Port-Q5500 Index (Australian Coal)-EXT (EX-tank (EXT) refers to the price of coal shipped from the warehouse, including the price before the coal is put into the warehouse and the warehouse usage fee)	$u_7$	0.545
	Guangzhou Port-Q5500 Index (Indian Coal)-EXT	$u_8$	0.571
Coal inventories	Qinhuangdao Port	$u_9$	0.007
	Caofeidian Port	$u_{10}$	-0.157
	Guangzhou Port	$u_{11}$	-0.021
	Tianjin Port	$u_{12}$	-0.295
	Jintang Port	$u_{13}$	-0.197
Fuel oil prices	Fuel Oil 180 Singapore-FOB	$u_{14}$	0.202
	Fuel Oil 380 Singapore-FOB	$u_{15}$	0.205
	China fuel oil futures closing price (a closing price is the final price at which it trades during regular market hours on any given day) (continuous contract (a continuous contract is a reinsurance contract that does not have a fixed contract end date, which will continue to be renewed and be in effect until one of the parties in the contract terminates it. Continuous contracts are different from standard reinsurance contracts in that they do not provide coverage for only a fixed period of time))	$u_{16}$	0.265
	China fuel oil futures closing price (active contract (an active contract means that this future contract can be traded for a specific amount of time))	$u_{17}$	0.272
	China fuel oil futures settlement price (the settlement price is the average price at which a contract trades, calculated at both the open and close of each trading day) (continuous contract)	$u_{18}$	0.266
	China fuel oil futures settlement price (active contract)	$u_{19}$	0.273
Crude oil prices	Brent crude oil spot price	$u_{20}$	0.152
	West Texas Intermediate (WTI) crude oil spot price	$u_{21}$	0.157
	Dubai crude oil spot price	$u_{22}$	0.152
	WTI crude oil futures settlement price	$u_{23}$	0.181
	Brent crude oil futures settlement price	$u_{24}$	0.187

where  $\min(\tilde{s})$  is the minimum value of  $\tilde{s}$  and  $\max(\tilde{s})$  is the maximum value of  $\tilde{s}$ .

**4.3. LSTM-Ensemble Learning (LSTM-EL) Approaches for CBCFI Forecasting.** Traditional freight indices prediction methods usually use the historical time series data of the target with ignorance of other impact factors. Generally speaking, the trend of CBCFI is reflected in two ways: historical CBCFI information and impact factor information. The historical time series information sometimes is sparse and thus not enough to produce accurate prediction, while some close impact factor information could reflect the movement of CBCFI from different aspects to a certain extent and with the support of the powerful database of Wind Economic Database (Wind Economic Database, refer to <https://www.wind.com.cn/en/edb.html>).

Our proposed LSTM-EL model is composed of two layers: in the first layer, a cluster of LSTMs is constructed to

generate the embedding features, and in the second layer, an ensemble learning method for final CBCFI prediction.

Figure 4 illustrates the overall framework of the bilevel hybrid LSTM-EL configuration; it consists of an LSTM method and two parallel ensemble learning methods. Note that the LSTM-EL model includes two different hybrid models, LSTM-GBRT and LSTM-RF. GBRT behaves similarly to RF in the manner of fitting multiple trees, but it instead fits them in a sequential manner, and hence another expectation of this paper aims to explore which ensemble algorithm fits the CBCFI forecasting problem better. For detailed descriptions of single RF, GBRT, and LSTM approaches, see Appendix A.

In the first layer, the dataset is first split into the in-sample and the out-of-sample. The preliminary embedded LSTM focuses on extracting the time-dependency information from variables of the in-sample and generates embedding features from the last LSTM layer, GBRT/RF is taken as an ensemble learning method to make the final

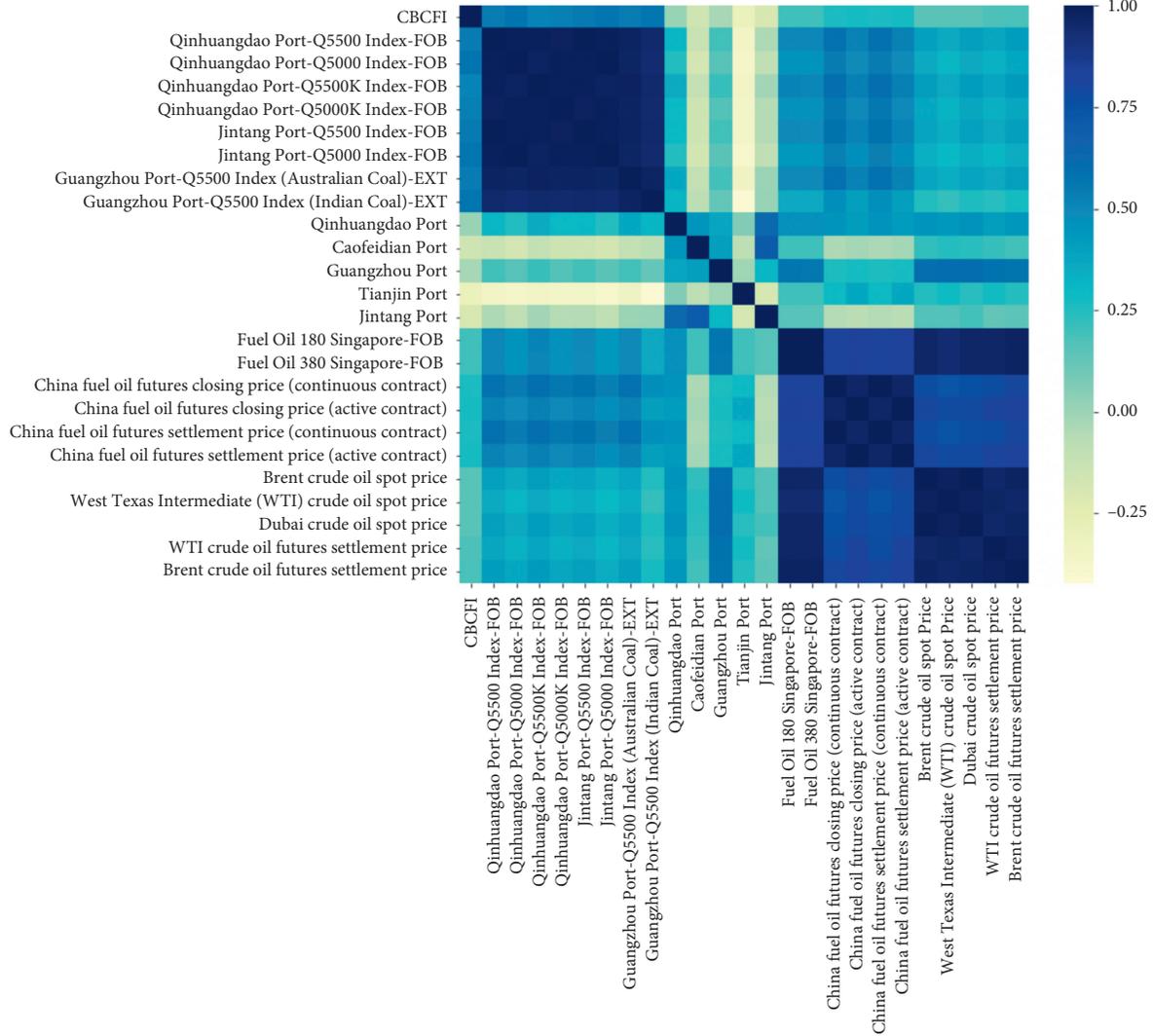


FIGURE 3: The Pearson correlation matrix of all features.

predictions by combining preliminary embedding features from LSTM, and the prediction values of each individual tree are summed up to get the final value. The details of the proposed LSTM-EL model are illustrated in Algorithm 1.

The idea behind GBRT is that each iterator is used to reduce the previous residual. To reduce these residuals, a new tree in the direction of the gradient descent of the loss function is created. After LSTM forms the training samples, the recursive form regression tree is as the equation to calculate  $F_{g,m}(s_{t+1}')$  in Algorithm 1.

Before building the bilevel LSTM-EL prediction architecture, several hyperparameters should be determined. For the upstream model LSTM, the LSTM network with optimization of multiple hyperparameters has achieved acceptable performance when applied on sequence data [51]. For the time series problem, the key hyperparameters include the number of LSTM layers, the number of nodes in each LSTM layer, the number of fully connected layers, and the time-lags, and for ensemble learning algorithm, the number of trees and the maximum depth of a tree are the

most essential parameters. In our work, the time-lag and embedding size are the most important hyperparameters.

- (1) Time-lag: the time-lag parameter has a significant impact on the performance of time series forecasting [52], as it determines the length of the historical sequence that should be included for the training.
- (2) Embedding size: that is, the number of neurons for the last layer in the LSTM network represents the input-data dimension of the downstream ensemble learning models and further determines the complexity of GBRT and RF. If the embedding size is very high, then the LSTM will be overfitting on training instances and increase the training difficulties of the downstream models, and if its size is too small, then it will be unable to memorize the time-dependency information collected from the time-lag sequences.

However, to the best of our knowledge, there are no general rules to choose the time-lag and the hidden layers' size. Therefore, we investigated the effect of key parameters

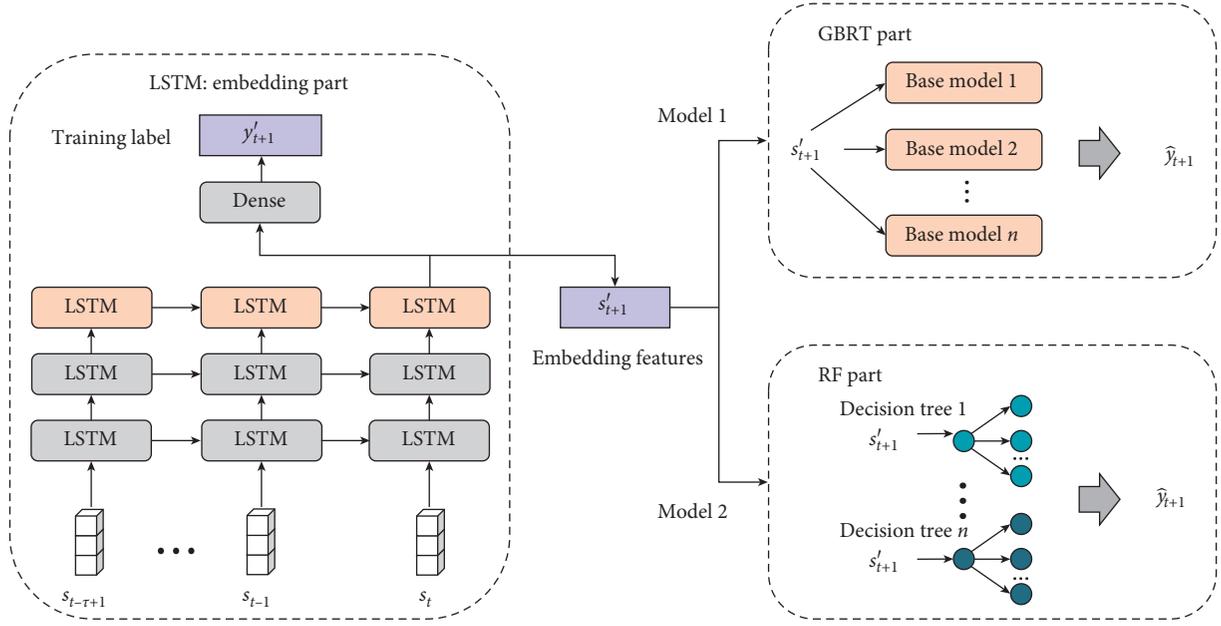


FIGURE 4: Structure diagram of LSTM-EL approach.

**Input:** historical observations:  $\{X_1, \dots, X_t, \dots, X_T; Y_1, \dots, Y_t, \dots, Y_T\}$ ,  $X_t = (x_t, x_{t-1}, \dots, x_{t-\tau+1})$ ,  $Y_t = (y_t, y_{t-1}, \dots, y_{t-\tau+1})$ ,  $X_t \in \mathbb{R}^{\tau \times D}$ ,  $Y_t \in \mathbb{R}^{\tau \times 1}$ ,  $x_t \in \mathbb{R}^D$ ,  $y_t \in \mathbb{R}$ ; Length of input sequence (time-lag):  $\tau$ ; Feature size:  $D$  (i.e., 22 in this paper).

**Output:** learned LSTM-EL model.

//construct training instance

for all available time interval  $t (1 \leq t \leq T)$  do:

$S_t = (X_t; Y_t)$ ,  $S_t \in \mathbb{R}^{\tau \times (D+1)}$  //Embedding part- LSTM model to extract the features with time-dependent information.

Given a training instance  $S_t$ :

**Step 1.** Embedding features:

$s'_{t+1} = \text{LSTM}(S_t)$  //embedding features produced by LSTM, which fused the temporal and internal correlations of  $S_t$  into a new vector  $s'_{t+1}$  with lower dimension,  $s'_{t+1} \in \mathbb{R}^{d'}$  ( $d'$  equals to the number of units in the last LSTM layer and the last dense layer only has one unit).  $s'_{t+1}$  is then being used to predict  $y_{t+1}$  in the second layer.

**Step 2.** LSTM prediction:

$y'_{t+1} = \text{Dense}(\text{LSTM}(S_t))$  //  $y'_{t+1} \in \mathbb{R}$  is the prediction value given by LSTM part. Note that we just use  $y'_{t+1}$  to optimize the parameters of the embedding part and the final prediction of our method is obtained by the downstream methods, GBRT and RF.

**Step 3.** Optimization:

$\min_{\theta} \|y_{t+1} - y'_{t+1}\|_2^2$  //The embedding part is trained by minimizing the objective function shown above and its parameters  $\theta$  are updated via backpropagation. //GBRT or RF is taken as downstream method to make the final predictions.

//**Model 1: GBRT part**

Given the embedding features  $s'_{t+1}$  generated from  $S_t$ :

**Input:** training set  $\{s'_{t+1}\}_{t=1}^T$ , differentiable loss function  $L(y_{t+1}, F_g(s'_{t+1}))$ , and the maximum number of trees  $M$ . //  $F_g$  is the decision tree mapping function, and the optimal  $F_g$  can be obtained through minimizing the loss function.

For  $m=1$  to  $M$ , do

For  $t=1$  to  $N$ , do

$F_{g,m}(s'_{t+1}) = F_{g,m-1}(s'_{t+1}) + \text{lr} * \rho_m g_m(s'_{t+1})$ . //  $\rho_m$  is the step-size,  $g_m$  is the base learner, and lr is the learning rate. For each step, a new decision tree is aimed at correcting the error made by its previous base learner.

Output model  $\hat{y}_{t+1} = \text{GBRT}(s'_{t+1})$  //

//**Model 2: RF part**

**Step 4:** given the embedding features  $s'_{t+1}$  generated from  $S_t$

**Input:** training set  $\{s'_{t+1}\}_{t=1}^T$ , the number of trees in forest  $M$

For  $m=1$  to  $M$ , do

$D_k = \text{Bootstrap}(D)$  //  $D$  is the training set

$h_k = \text{Decision Tree with random feature selection}(D_k, f)$  //  $f$  denotes the number of attributes to use at each node, picked uniformly at random new features for every split

Prune tree to minimize out-of-bag error

Average all  $M$  Trees

Output model  $\hat{y}_{t+1} = \text{RF}(s'_{t+1})$ .

while keeping the other parameters fixed, and grid search [53] was applied to find the optimum hyperparameters. To do so, our model requires some basic settings. In our work, we built our model using two LSTM layers and one fully connected layer. The number of the neurons for the first layer is equal to that of the second layer. The search space of the above four critical parameters is illustrated in Table 3.

For each combination of the time-lag and embedding size, the LSTM-EL model is designed and trained, and the corresponding optimal combination of the number of trees and depth of trees is selected using the grid search. Here, for the sake of brevity, different combinations of the time-lag  $\tau$  and the embedding size  $d'$  are evaluated by root mean square error (RMSE) and mean absolute percentage error (MAPE) and the results are shown in Figures 5 and 6. According to previous studies, Li et al. [54] showed that a small time-lag cannot guarantee enough long-term memory inputs for our LSTM-EL model; thus, the model cannot fully exploit the LSTM for long-term memory modeling. Large time-lags permit an increased number of unrelated inputs, which increases the model's complexity and the difficulty of learning useful features. It can be observed that the effect of the number of nodes in each neuron layer shows that, with the increase in the number of neuron nodes, the prediction performance improves slightly. Thus, we set different time-lag  $\tau$  and embedding size in the successive experiment to optimize both accuracy and time efficiency for both LSTM-GBRT and LSTM-RF, as indicated by the RMSE and MAPE. Based on the results of the experiments, Table 4 summarizes the best parameters of the obtained model.

**4.4. Accuracy Measurement.** For measuring the prediction precision pertaining to the developed approach, several evaluation criteria are implemented, such as the RMSE as well as the MAPE. Generally, with the decline of MAPE and RMSE, the approach will be more precise. However, it is well known that, for a given prediction, actual outcomes above and below the prediction are treated asymmetrically when using MAPE and RMSE [55]. For the mentioned reason, to measure the data fluctuations (e.g., upward, stable, or downward), direction matching rate (Dsta) is employed. In addition, we utilize mean absolute scaled error (MASE) to assess if the developed prediction approach outclasses the naïve prediction method [56]. For detailed description of MAPE, RMSE, Dsta, and MASE criteria, see Appendix B.

## 5. Empirical Results

**5.1. Daily, Weekly, and Monthly CBCFI Forecasting.** After determining the best network architecture for the prediction task, the training set was utilized to train our LSTM-EL model until convergence. Evaluations were conducted using the test set. To analyze the generality of the hybrid LSTM-EL structure, we use a dataset with day-to-day, week-to-week, and month-to-month bases. Specifically, the weekly and monthly data are calculated as the average of daily CBCFI. In addition, to avoid overfitting problem, early stopping and validation sets are utilized in the present study, and the

TABLE 3: Value specified for key hyperparameters.

Hyperparameters	Search range
Time-lag $\tau$	[2, 4, 6, 8, 10]
Embedding size $d'$	[16, 32, 64, 128]
Number of trees	[1, 1000]
Depth of trees	[1, 50]

percentages for training, testing, and validating sets are 60%, 20%, and 20%, respectively.

In our prediction of each forecasting approach, LSTM-RF, LSTM-GBRT, GBRT, and RF models are applied, all of which are evaluated by calculating MAPE, RMSE, Dsta, and MASE. A rolling approach is implemented to conduct a next-day/weekly/monthly CBCFI forecast. The approach uses the actual value of the predictor variable in the previous period for making a prediction in the testing set. Note that the time-lag is fixed, and new data are added for further  $t+1$  prediction. Figure 7 displays how the rolling approach works.

Note that, in weekly forecasting, each point represents the weekly CBCFI value and a new weekly CBCFI value is calculated by every new 5 daily CBCFI values (only workdays data). Likewise, in monthly forecasting, each point represents the monthly CBCFI value and a new monthly CBCFI value is calculated by the working days in each new month, automatically excluding weekends (Saturday and Sunday).

We next conduct the daily, weekly, and monthly CBCFI forecasting experiments, respectively. The CBCFI data from January 2012 to October 2020 are sample data. To evaluate the predictive performance of LSTM-EL models, we split the data into training data, validating data, and testing data. The ratio for each dataset is 6:2:2. Figure 8 shows the learning curves of mean square error (MSE) for the validation data and training data for 100 epochs in daily, weekly, and monthly forecasting cases. The learning curves show a good fit because of the decrease in training and validation loss to one stable data with a minimal gap between the two final loss values.

Table 5 compares the predictive performance of two hybrid LSTM-EL approaches with the corresponding single EL approaches in the rolling forecasting approaches. For daily CBCFI forecasting cases, we found that the predictive performances for orientation matching and errors are enhanced through the introduction of the proposed hybrid structure. And all the values of MASE less than 1 indicate that four approaches outperform the average one-step naïve forecast. Note that the predictive performance in MAPE, RMSE, and Dsta is improved by considering hybrid structure forecasting. Compared to the predictive performances of corresponding single approaches, the improvement percentages (the improvement percentage is calculated by the difference between the evaluation index value of the hybrid model and the conventional model over that of the conventional model) in MAPE of LSTM-GBRT and LSTM-RF are 22.47% and 24.59%, respectively, of RMSE are 41.54% and 24.73%, respectively, and of Dsta are 72.84% and 69.36%, respectively. The above results indicate that LSTM-GBRT exhibits the most significant improvement level.

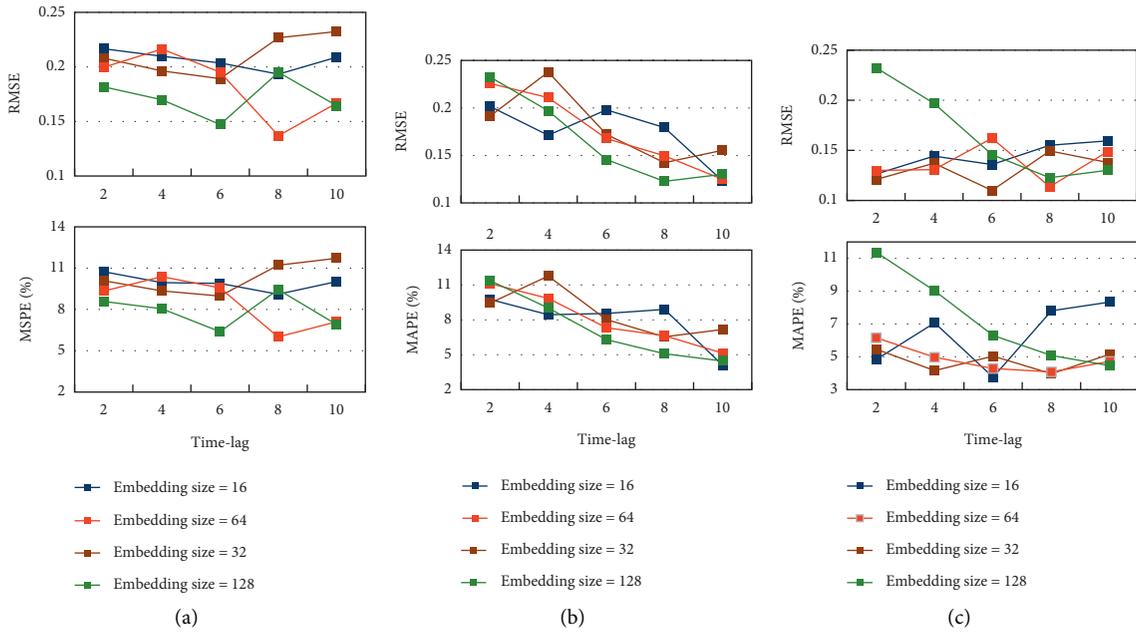


FIGURE 5: Key parameter sensitivity analysis for LSTM-GBRT under three time scales: (a) daily, (b) weekly, and (c) monthly CBCFI forecasting.

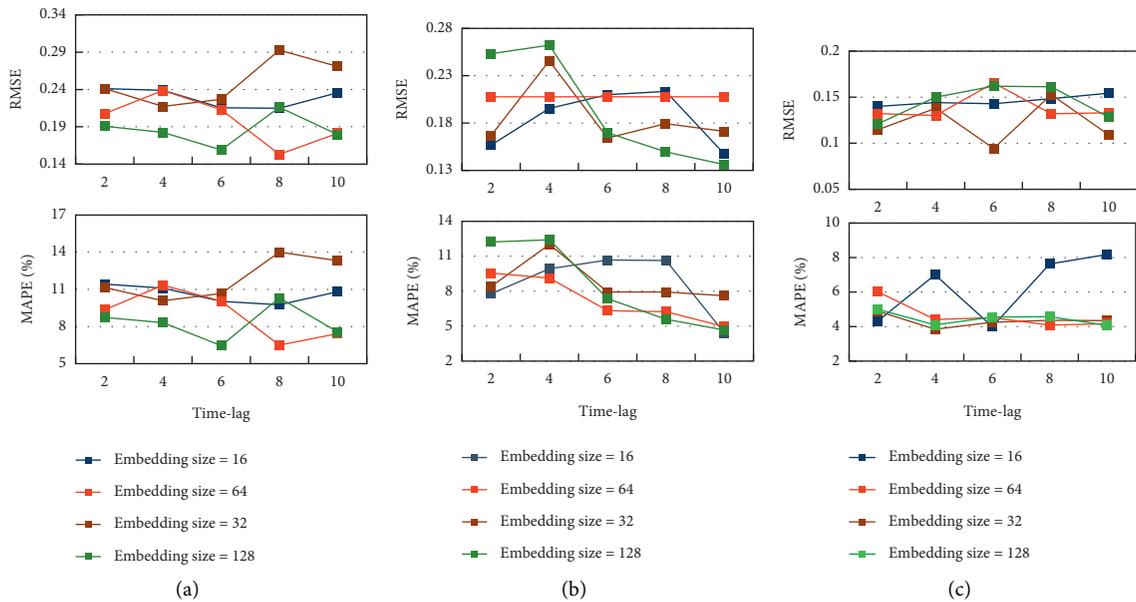


FIGURE 6: Key parameter sensitivity analysis for LSTM-RF under three time scales: (a) daily, (b) weekly, and (c) monthly CBCFI forecasting.

For weekly CBCFI forecasting cases, consistent with daily CBCFI predicting processes, performance enhancement rates are positive for both errors and Dsta. The improvement percentages in MAPE of LSTM-GBRT and LSTM-RF are 54.62%, 39.48%, respectively, of RMSE are 32.12% and 11.77%, respectively, and of Dsta are 50.91% and 67.17%, respectively. Moreover, the improvement in MAPE for all two LSTM-EL approaches shows higher significance for the weekly CBCFI predicting processes than the daily prediction. Thus, using a hybrid structure to extract the

time-dependent characteristics between features in LSTM-based prediction enhances accuracy. In comparison with daily information, the hybrid approach shows significant improvements for the weekly CBCFI predicting process.

Since predicting long-term CBCFI with low-frequency time-scale data raises several challenges, this study examines how the hybrid approach promotes monthly data forecasting to be accurate. Consistent with daily and weekly forecasting, hybrid approaches outperform the single ensemble learning approaches. Notably, the hybrid structure promotes the

TABLE 4: The best configurations of the proposed model and benchmark models.

Algorithm	Best configurations	Forecasting cases		
		Daily	Weekly	Monthly
LSTM-GBRT	LSTM layer			
	Time-lag	8	10	6
	Number of hidden layers	2	2	2
	Number of units in the hidden layers	64	16	32
	GBRT layer			
LSTM-RF	Number of trees	101	31	41
	Depth of trees	1	1	1
	LSTM layer			
	Time-lag	8	10	6
	Number of hidden layers	2	2	2
GBRT	Number of units in the hidden layers	64	16	32
	RF layer			
	Number of trees	41	11	21
	Depth of trees	11	11	1
	Number of trees	191	11	11
RF	Depth of trees	21	1	1
	Number of trees	11	31	11
LSTM	Depth of trees	1	1	1
	Time-lag	8	10	8
	Number of hidden layers	2	2	2
	Number of units in the hidden layers	64	32	64

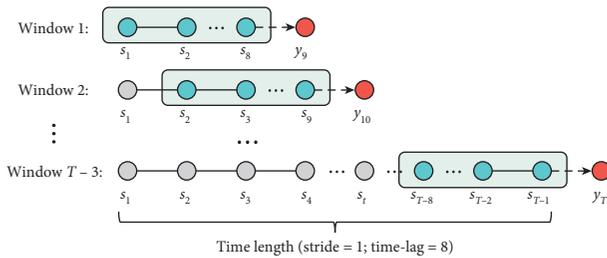


FIGURE 7: Illustration of rolling forecasting.

precision noticeably, MAPE improvement for GBRT and RF are 53.72% and 61.40%, respectively. RMSE improvements for LSTM-GBRT and LSTM-RF are 50.32% and 57.83%, respectively. Moreover, the MASE of the monthly data of hybrid approaches is less than 1, indicating that hybrid approaches in this approach outperform the average one-step naïve forecast.

Table 6 compares the general predictive performances of hybrid LSTM-EL approaches over three time scales. The MAPE and RMSE values indicate that hybrid LSTM-EL approaches perform best in weekly CBCFI forecasting and achieve the most obvious improvement of accuracy in monthly forecasting. And the performance indicators of different approaches (Table 5) show that the LSTM-GBRT approach outperforms LSTM-RF in daily forecasting while LSTM-RF achieves a higher accuracy in weekly and monthly forecasting, which indicates that LSTM-GBRT is better capable of dealing with high-frequency data while LSTM-RF is more suitable for mid- or low-frequency data.

Figure 9 shows the comparisons between the real CBCFI values and the predicted values produced by LSTM-EL approaches and their corresponding single EL approaches.

The inset square of Figure 9(a) shows how the hybrid and single EL approaches perform at the uptrend, bottom, and downtrend. It is found that four approaches are capable of predicting the daily, weekly, and monthly tendency of CBCFI, but the outputs from LSTM-EL approaches derive less than from the actual CBCFI values. In the three situations, single EL predictions show obvious fluctuations that actual CBCFI values do not have. For example, in situation I beginning in January 2019, the GBRT and RF predictions display heavily up and down while they do not happen in the actual CBCFI trend, LSTM-GBRT and LSTM-RF still keep close to the real CBCFI values, and similar phenomena are observed in situations II and III.

Figures 9(b) and 9(c) show that single GBRT and RF models produce large errors in weekly and monthly forecasting, especially at the bottom or top of the trendline. In contrast, the hybrid LSTM-EL forecasting approach reproduces CBCFI trends and generates relatively small errors. For example, in weekly forecasting beginning in December 2019, the CBCFI dropped from approximately 1,000 points to a historic low of roughly 450 points, since the outbreak of COVID-19 hits the economy across the globe. And the CBCFI was kept at a low point until May 2020, despite the epidemic was gradually under control. In this case, only LSTM-GBRT and LSTM-RF predictions reproduce the CBCFI trend, while other approaches display large fluctuations not existing in the actual CBCFI itself. Likewise, in the monthly forecasting, the single EL approaches GBRT and RF predictions deviate, and the LSTM-GBRT and LSTM-RF predictions are very close to the actual CBCFI.

5.2. Diebold–Mariano (DM) Test. To evaluate whether there is any statistically significant difference between the hybrid

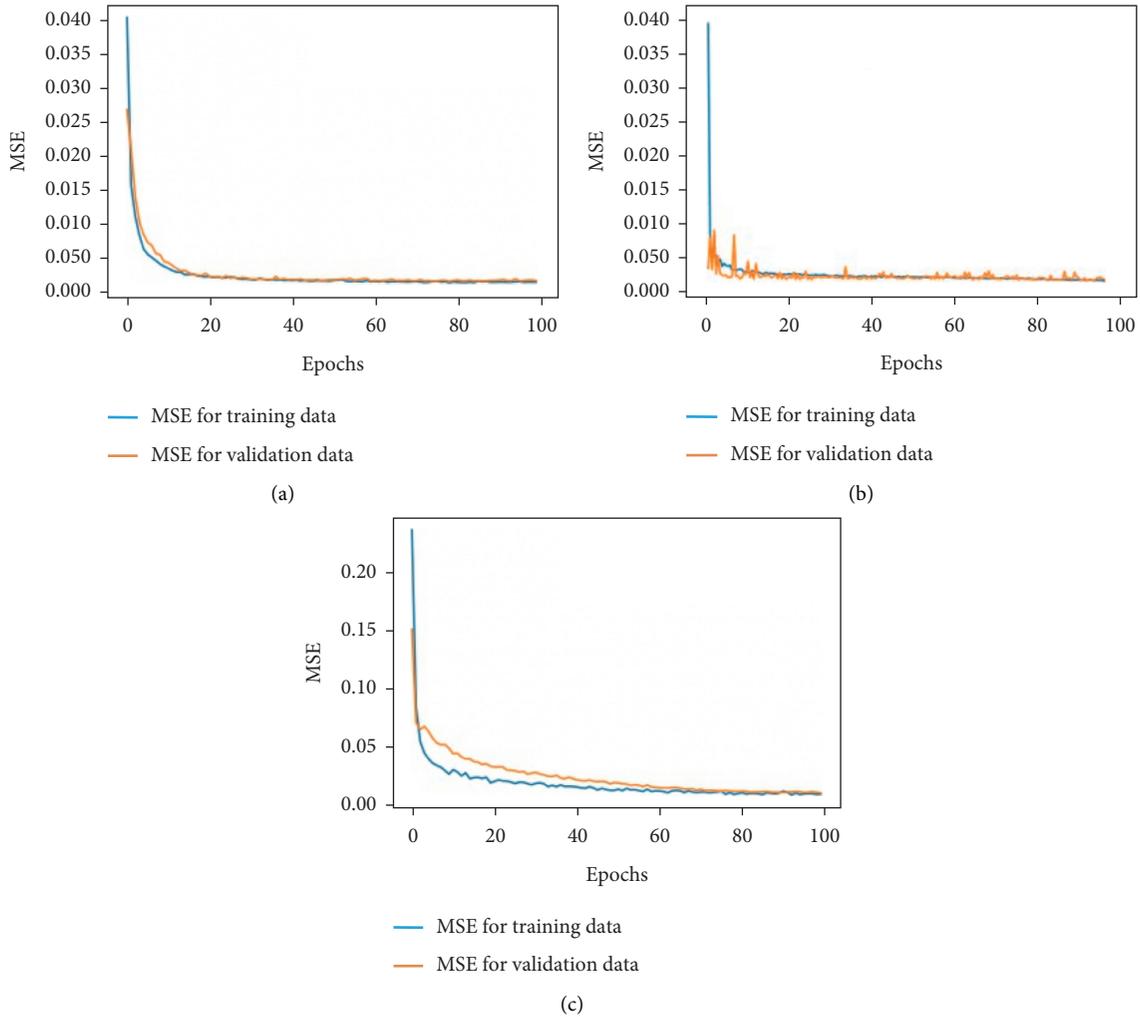


FIGURE 8: MSE for validation and training data for (a) daily, (b) weekly, (c) monthly forecasting cases.

and conventional models, the Diebold–Mariano (DM) test [57] is implemented to compare the testing-sample prediction results. The DM test is widely used in determining whether the differences of time series predicting accuracy by different models are substantially crucial from a statistical perspective [58]. Table 7 summarizes the results of the DW test for the daily, weekly, and monthly CBCFI datasets, respectively.  $p$  value less than 0.05 indicates the rejection of the null hypothesis that there is no difference between the two compared forecasting models. It can be seen that LSTM-EL models have significantly different accuracies when compared to other benchmarks. For the daily dataset, LSTM-RF and LSTM-GBRT present statistical differences in predictive performance when compared with other models but there is no statistical difference in accuracies between LSTM-RF and LSTM-GBRT. Similarly, in weekly and monthly forecasting cases, the predictive improvement offered by incorporating ensemble learning algorithms is statistically significant while the predictive improvements between different conventional ensemble learning algorithms or LSTM-based hybrid models do not display statistically predictive performances.

**5.3. Feature Importance Analysis.** Moreover, Figure 10 presents the features ranked by complying with the clarified variance the respective feature facilitates the LSTM-GBRT approach. In this case, the features are plotted against their relative importance, that is, the percent significance regarding the critical feature. For brevity, Figure 10 only presents the top 6 features with the sum of feature importance over 98%. It can be clearly seen that coal inventory at Tianjin Port influences the monthly CBCFI values but less important for daily and weekly data, which indicates that coal inventory is more likely to impact the long-term forecasting but not short- or midterm forecasting. In addition, domestic and overseas thermal coal spot rates and crude oil prices have obvious impacts on daily, weekly, and monthly CBCFI values, while coal inventory and fuel oil price are less important for daily and weekly CBCFI. Specifically, Guangzhou Port-Q5500 Index (Australian Coal)-EXT shows a great impact on daily, weekly, and monthly CBCFI values. Moreover, WTI crude oil spot price has an obvious impact on daily CBCFI values, while weekly CBCFI is more sensitive to Dubai crude oil spot price. This may result from these two crude oil indices that mainly serve

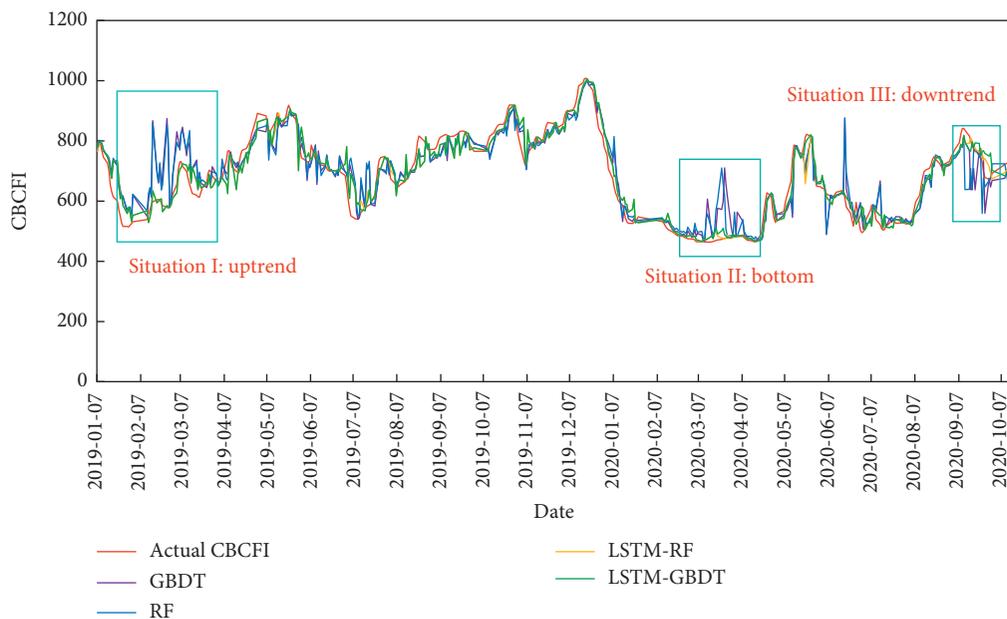
TABLE 5: Predictive performance of hybrid LSTM-EL and benchmark models for CBCFI prediction.

Methods	MAPE (%)	RMSE	MASE	Dsta
<i>Daily CBCFI forecasting</i>				
<i>Training period: 2012/1/4~2017/4/10</i>				
<i>Testing period: 2019/1/7~2020/10/13</i>				
LSTM	9.15	0.2419	1.1532	0.5703
GBRT	7.84	0.2342	0.8915	0.5145
LSTM-GBRT	6.00	0.1369	0.9132	0.8893
RF	8.62	0.2034	0.8370	0.5042
LSTM-RF	6.50	0.1531	0.9560	0.8539
<i>Weekly CBCFI forecasting</i>				
<i>Training period: 2012/1/9~2017/3/24</i>				
<i>Testing period: 2019/1/21~2020/9/28</i>				
LSTM	9.47	0.1931	1.6532	0.4303
GBRT	8.99	0.1815	0.6173	0.5324
LSTM-GBRT	4.08	0.1232	0.9835	0.8034
RF	7.32	0.1673	0.5525	0.4853
LSTM-RF	4.43	0.1476	0.8941	0.8113
<i>Monthly CBCFI forecasting</i>				
<i>Training period: 2012/1/4~2017/3/4</i>				
<i>Testing period: 2019/4/1~2020/10/13</i>				
LSTM	11.53	0.2571	1.0126	0.5863
GBRT	10.89	0.2212	0.7940	0.5932
LSTM-GBRT	5.04	0.1099	0.9203	0.9024
RF	11.01	0.2236	0.7864	0.6072
LSTM-RF	4.25	0.0943	0.8825	0.9043

TABLE 6: Predictive performance of hybrid LSTM-EL over three time scales.

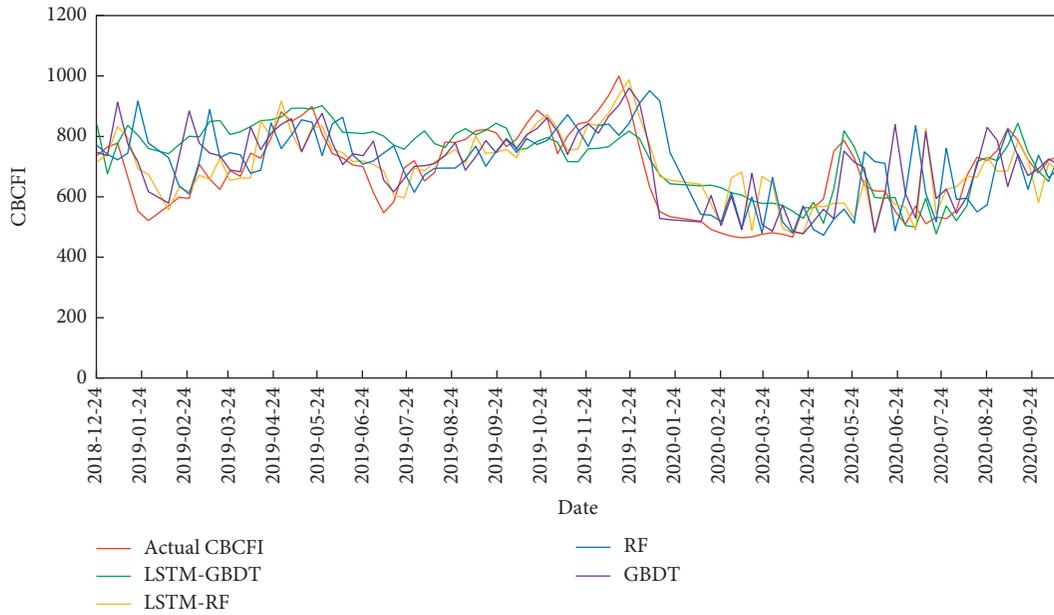
	Forecasting accuracy			Improvement percentage		
	Daily forecasting	Weekly forecasting	Monthly forecasting	Daily forecasting	Weekly forecasting	Monthly forecasting
<u>MAPE</u>	6.25%	4.26%	4.65%	23.53%	47.05%	57.56%
<u>RMSE</u>	0.1450	0.1354	0.1021	33.14%	21.95%	54.08%

Note. MAPE and RMSE represent the average values of MAPE and RMSE. Each average MAPE/RMSE is calculated by the average MAPE/RMSEs of LSTM-GBRT and LSTM-RF from Table 5.

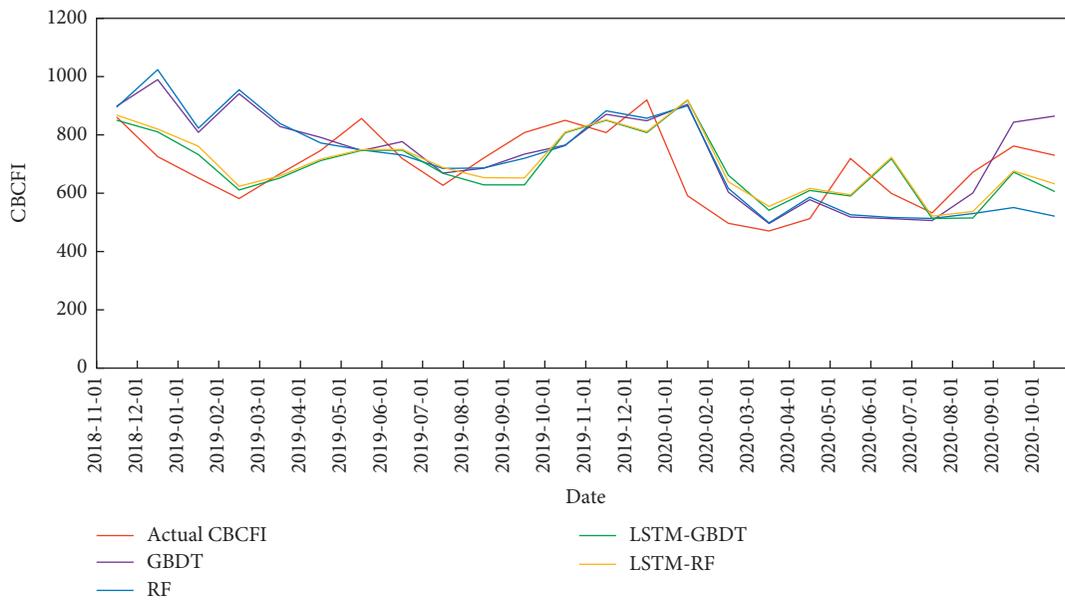


(a)

FIGURE 9: Continued.



(b)



(c)

FIGURE 9: Actual CBCFI and forecasting results through LSTM-EL approaches and corresponding EL approaches for (a) daily, (b) weekly, and (c) monthly data.

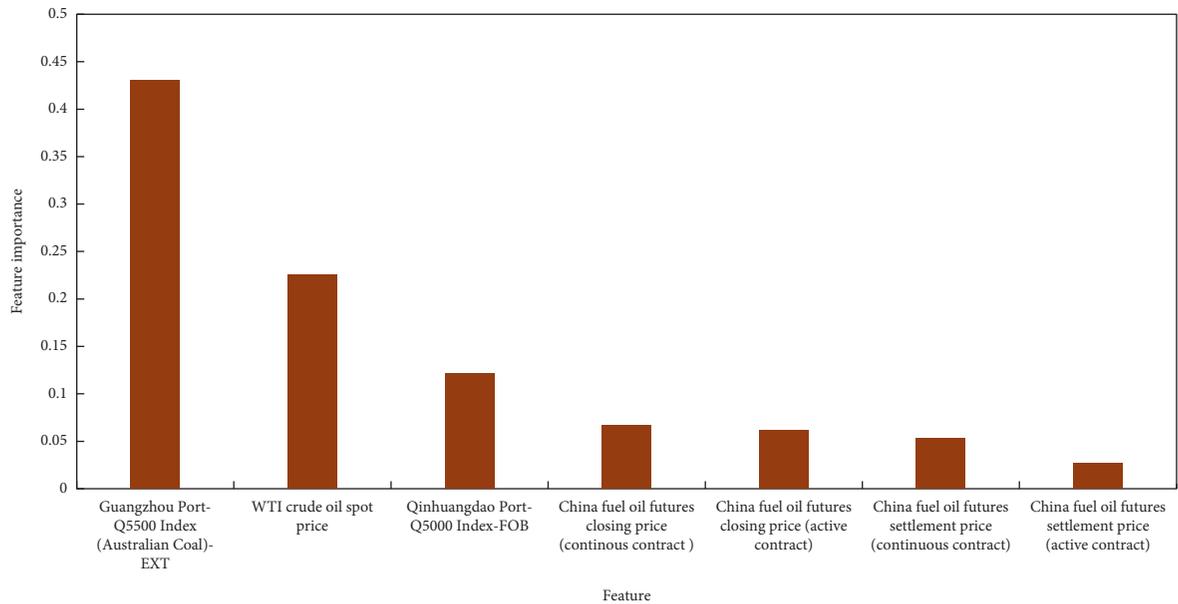
TABLE 7: DM test results for hybrid models and the benchmarks.

Data type	Tested model	Reference model				
		LSTM	GBRT	RF	LSTM-RF	LSTM-GBRT
Daily CBCFI forecasting	LSTM	—	—	—	—	—
	GBRT	2.4328**	—	—	—	—
	RF	2.4103**	-1.6784	—	—	—
	LSTM-RF	2.5123**	2.2341**	2.1231**	—	—
	LSTM-GBRT	2.2763**	2.2910**	2.8723**	-1.2432	—

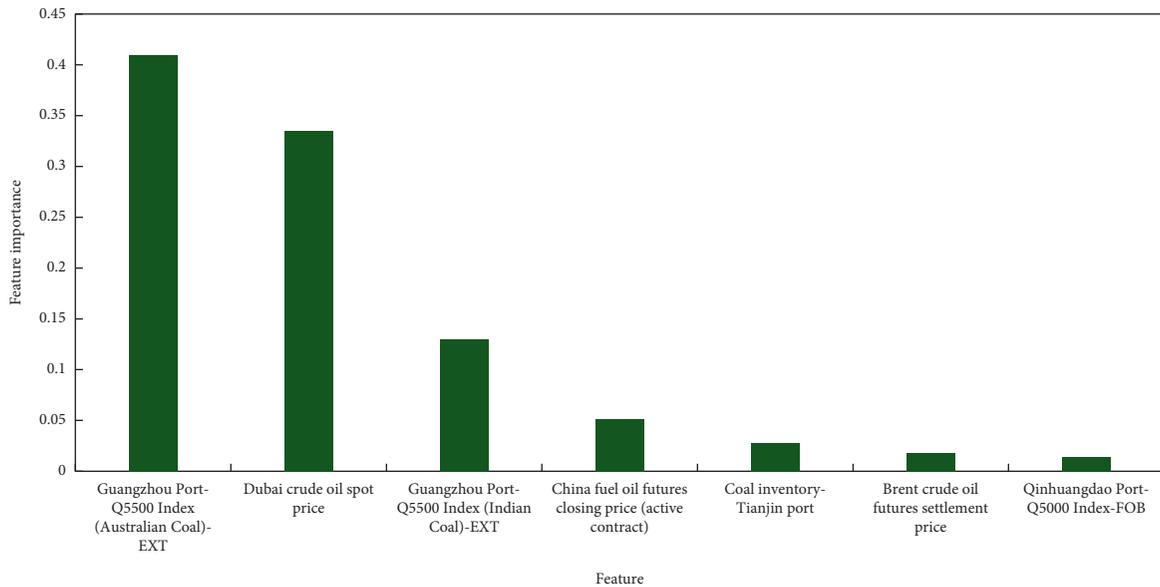
TABLE 7: Continued.

Data type	Tested model	Reference model				
		LSTM	GBRT	RF	LSTM-RF	LSTM-GBRT
Weekly CBCFI forecasting	LSTM	—	—	—	—	—
	GBRT	-2.7084**	—	—	—	—
	RF	2.2034**	-1.6110	—	—	—
	LSTM-RF	2.2361**	2.3414**	2.6535**	—	—
	LSTM-GBRT	2.2276**	2.3012**	2.5541**	2.6287	—
Monthly CBCFI forecasting	LSTM	—	—	—	—	—
	GBRT	-4.5123**	—	—	—	—
	RF	-5.2341**	-1.9883	—	—	—
	LSTM-RF	-5.2011**	-6.1094**	-5.3312**	—	—
	LSTM-GBRT	-5.1998**	-6.2014**	-6.4234**	-2.6536	—

Note. \*\*The value is significant at 5%.



(a)



(b)

FIGURE 10: Continued.

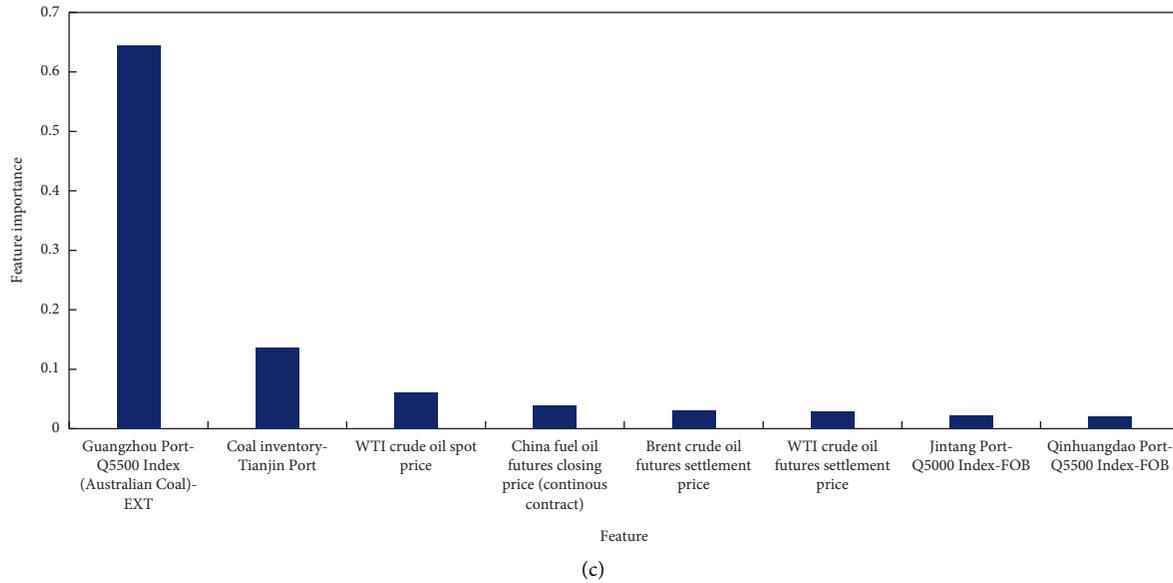


FIGURE 10: Feature importance for (a) daily, (b) weekly, and (c) monthly CBCFI forecasting cases.

different regions. China, the world's biggest oil importer, has ramped up purchases of American crude for years [59]. WTI crude oil refers to oil extracted from wells in the US and sent via pipeline to Cushing, Oklahoma, which is the main benchmark for oil consumed in the United States [60]. Crude oil as a substitute for coal and thus Chinese domestic crude oil price would be more sensitive to the daily changes of the US crude oil prices. In addition, because crude oil is a substitute for coal, the fluctuation of coal price will influence the bulk coal shipping cost somehow, and thus the CBCFI values will be sensitive to the WTI crude oil prices in the short time scale. On the other hand, as Dubai crude oil price index is the main reference for Persian Gulf oil delivered to the Asian market and roughly half (44.8%) of Chinese imported crude oil originates from nine Middle Eastern nations [61], the CBCFI index will be affected by the fluctuation of Dubai crude oil price as well. Moreover, Nanovsky [62] introduced an oil price-distance interaction variable to explain how global trade behaves as a result of oil price changes. He found that when oil prices increase, international trade becomes more localized in that countries begin trading relatively more with their neighbors. In contrast, when they decrease, trade becomes more dispersed in that the distance between countries becomes less relevant. Thus, the price of crude oil usually presents volatility in one week, and thus, the weekly Chinese coastal shipping cost may look at the price of the Asian crude oil market more.

**5.4. The Impact of the Supply and Demand on CBCFI.** As CBCFI is an index for shipping price, in addition to the factors discussed above, supply and demand are essential factors. For CBCFI, the supply should be available bulk fleet and the demand should be the amount of coal that needs to be shipped.

On the supply side, specifically, according to the routes and ports of CBCFI, the supply should be the available bulk

fleet at Tianjin Port, Jintang Port, Qinhuangdao Port, Caofeidain Port, and Huanghua Port. However, we did not find any available open-source fleet data.

On the demand side, specifically, the CBCFI represents the coal transportation need mainly from northern China to southern China. The national statistics (Figure 11) indicate that most of the coal was used for thermal power generation (56.40%) and steel-making (18.1%). Therefore, the southern thermal power generation and steel production are considered as indexes of the CBCFI-related demand. However, we only found the above data on a nationwide basis and cannot get the data only for southern China. To estimate the demand, we use the sum of utility electricity consumption of the southern coastal provinces with at least one port city included in the CBCFI routes, including Shanghai, Jiangsu, Zhejiang, Fujian, and Guangdong, as a substitution to the southern thermal power generation. For the steel production, we use the total domestic steel production as a substitution index of southern provinces' steel production.

Besides, the coal sources in China consist of 90% self-produced mainly from northern China and 10% coal imports [47]. The coal production and coal imports also may have impacts on the CBCFI-related demand. Specifically, both the coal production increase and coal imports decrease may increase the CBCFI-related demand. Therefore, domestic coal production and coal imports are also considered as the demand-related factors.

Note that the above demand and demand-related factors are only available on a monthly basis. Accordingly, southern utility electricity consumption, domestic steel production, domestic coal production, and coal imports are added to the monthly forecasting model.

Table 8 displays the improvement percentage of monthly predictive performance with adding demand and demand-related factors. It is found that the predictive performances of different models have positive improvements with adding

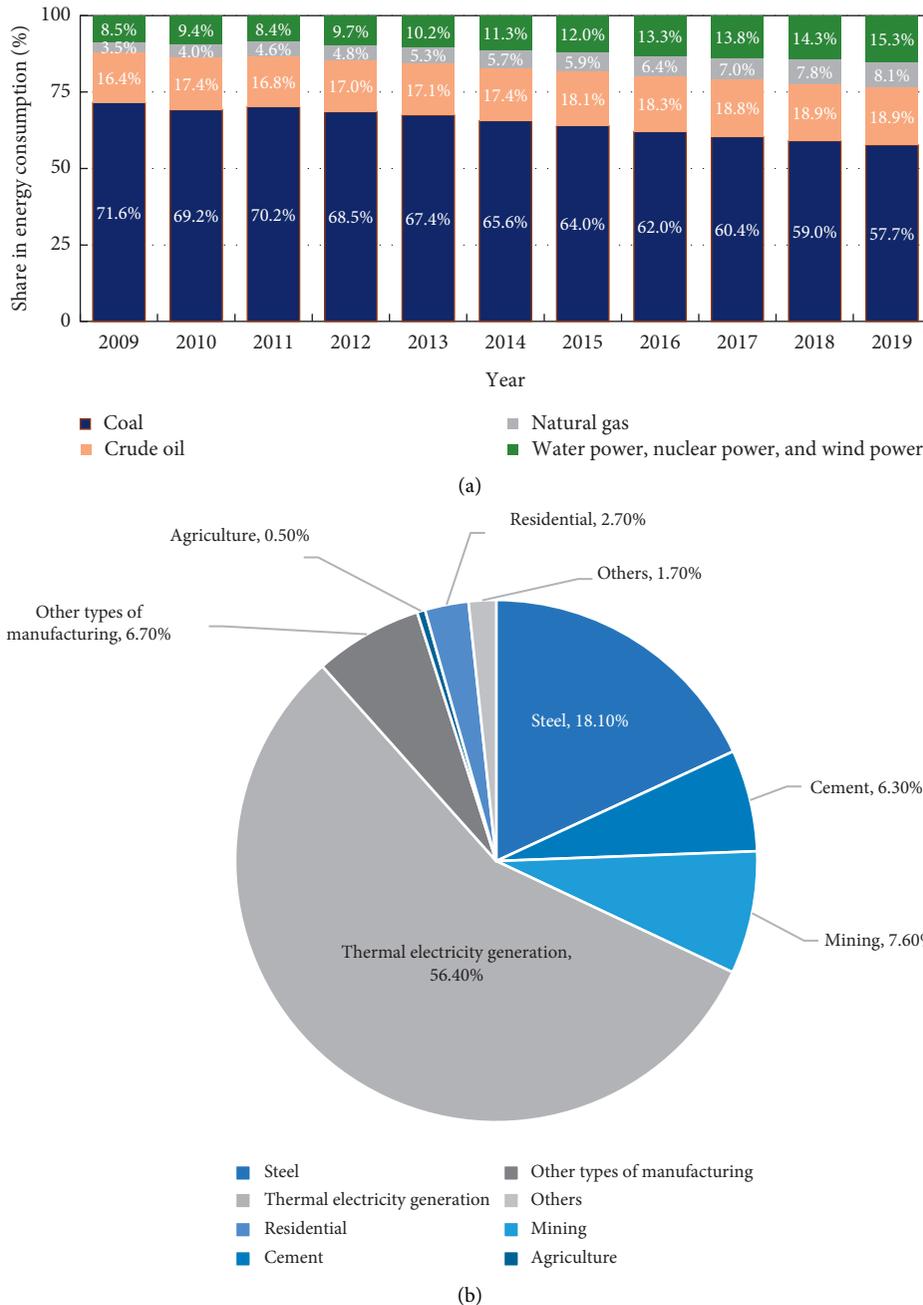


FIGURE 11: Energy consumption in China. (a) China's energy consumption by source (data source: Samantha. W. Energy consumption in China from 2009 to 2019, by source. Available at <https://www.statista.com/statistics/278669/energy-consumption-in-china-by-source/>); (b) China's coal consumption by end market in 2020 (Data source: National Bureau of Statistics. Available at <http://www.stats.gov.cn/tjsj/>).

demand and demand-related factors, particularly an obvious improvement in MAPE (12.67%) and RMSE (10.01%) of LSTM. And for hybrid models, the improvement percentages in MAPE of LSTM-GBRT and LSTM-RF are 5.32% and 5.16%, respectively, and of RMSE are 5.18% and 5.01%, respectively. Figure 12 presents the top 7 features with the sum of feature importance over 98%. Southern utility

electricity consumption ranks the top, followed by Guangzhou Port-Q5500 Index (Australian Coal)-EXT, domestic steel consumption, and coal inventory at Tianjin Port. As not all self-produced coal needs to be transported, domestic coal production shows little importance for CBCFI forecasting. The above results illustrate that demand factors could lead to a higher monthly predictive accuracy.

TABLE 8: Improvement percentage of predictive performance with adding demand and demand-related factors.

Evaluation criteria	Models	Improvement percentage of monthly forecasting (%)
MAPE	LSTM	12.67
	GBRT	9.97
	LSTM-GBRT	5.32
	RF	9.81
	LSTM-RF	5.16
RMSE	LSTM	10.01
	GBRT	8.90
	LSTM-GBRT	5.18
	RF	7.87
	LSTM-RF	5.01

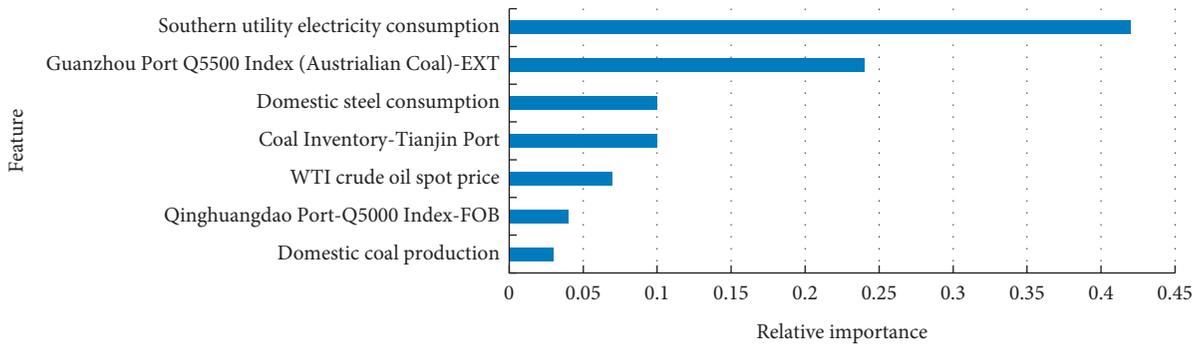


FIGURE 12: Feature importance for CBCFI monthly forecasting with adding demand and demand-related factors.

## 6. Conclusion

The present study attempted at enhancing the forecasting accuracy of the CBCFI by formulating a novel hybrid LSTM-EL approach, which is capable of extracting the useful time-dependent information in the data by combining the LSTM technique and ensemble learning algorithms. A rolling forecasting approach is developed for assessing LSTM-EL's forecasting accuracy in comparison with its corresponding single ensemble algorithms. Furthermore, critical factors that influence CBCFI values are discussed and experiments under daily, weekly, and monthly time scales in the rolling forecasting approach are conducted in order to test the performance generality of LSTM-EL approaches.

The major intellectual advantages here consist of the emerging method by exploiting artificial neural network and ensemble learning methods to be the useful approach to obtain the shipping freight market's nonlinear and non-stationary features. According to the empirically achieved outcomes, domestic and overseas thermal coal spot rates and crude oil prices have obvious impacts on daily, weekly, and monthly CBCFI values, while coal inventory and fuel oil price are less important for daily and weekly CBCFI. In terms of forecasting accuracy, LSTM-EL approaches outperform the single EL models in three time-scale forecasting cases and generate better results than the naïve forecasts. Moreover, the accuracy improvement by LSTM-EL approaches for different CBCFI time-scale datasets is validated. Results indicate that hybrid LSTM-EL approaches perform best in daily CBCFI forecasting but achieve the most obvious

improvement of accuracy in weekly forecasting. In addition, a DM test is implemented to evaluate whether there is any statistically significant difference between the hybrid and conventional models, and the results illustrate that LSTM-based hybrid models present statistical difference in predictive performance when compared with other models but there is no statistical difference in accuracies between LSTM-RF and LSTM-GBRT, so do the weekly and monthly forecasting cases. Overall, the LSTM-EL method has a high prospect to predict the CBCFI index in an accurate manner.

The mentioned emerging method is capable of acting to be one effective tool to make the decisions regarding chartering and shipping based on uncertain properties and further being incorporated into management toolkit by shipping industry practitioners. The developed method and outcomes widen freight rates forecasting study and indicate probable subsequent study in relevant aspects fields.

## Appendix

### A. Basic Model Theory

The deep-learning approach exhibits one prominent performance as opposed to the conventional statistics-related approach since it is capable of mapping the initial information for a nonlinear approach, which generates more effective influence. And long short-term memory (LSTM) based on the concept of recurrent neural network (RNN) presents an outstanding ability in time series predictions. On the other hand, ensemble learning methods refer to machine

learning technique that combines several bases approaches in order to minimize the causes of error in learning approaches, such as noise, bias, and variance, for improving the overall predictive performance of the approach. In this paper, two prevailing approaches are focused on, (i) Random Forest (RF) and (ii) gradient boosting regression tree (GBRT).

### (1) Long short-term memory (LSTM)

LSTM, developed by Hochreiter and Schmidhuber, refers to on e special type of recurrent neural networks (RNN) [63]. LSTM consists of a unique set of memory cells that trains the data through a time back-propagation algorithm, capable of solving the long-term dependence problems of RNN, and thus is suitable for time series problems. The schematic diagram of the LSTM approach is displayed in Figure 13.

LSTM has options for adding or deleting the data of its cell condition, as achieved with the use of cell gates. The standard LSTM can be expressed as follows. The respective step  $t$  and its corresponding input sequence are denoted as  $X = \{x_1, x_2, \dots, x_t\}$ , and the three types of gates are input gate  $i_t$ , output gate  $o_t$ , and forget gate  $f_t$ . The passed information can be determined whether to be remembered or forgotten by the output of the hidden layer  $h_{t-1}$  and input  $x_t$  of the current layer. The activation function for limiting the outputted data inside the range of  $[0, 1]$  is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \quad (\text{A.1})$$

The input gate aims at determining the appropriate input information  $(h_{t-1}, x_t)$  to the cell. Its activated function is to set the forgetting gate.  $\tilde{C}_t$  denotes a "candidate" hidden state determined from the previous hidden state and the current input.  $C_t$  expresses the internal memory of the unit. For achieving the purpose of retaining the corresponding information,  $C_t$  integrates the previous memory, under the multiplication from the gate that is forgotten, and the newly hidden state below under the multiplication from the input gate.

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \tilde{C}_t &= \tan h(W_c \cdot [h_{t-1}, x_t] + b_c), \\ C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t. \end{aligned} \quad (\text{A.2})$$

The output gate controls the data able to be outputted. Likewise, the activation function aims at setting the gate this is forgotten. After the memory cell state gets updated by the  $\tan h$  activation function, the multiplication of dot determines the to-be-output information.

$$\begin{aligned} o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= o_t \cdot \tan h(C_t), \end{aligned} \quad (\text{A.3})$$

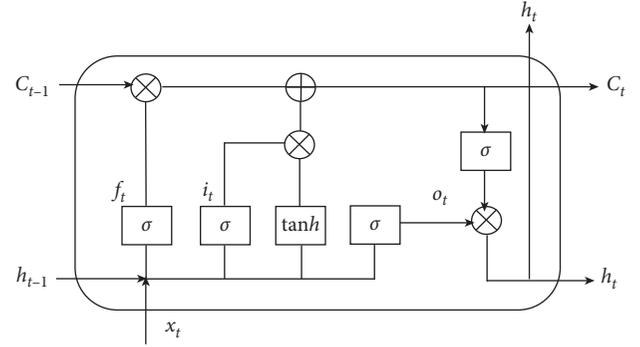


FIGURE 13: LSTM structure diagram.

where  $W_i$ ,  $W_f$ , and  $W_o$  denote the matrices of weight and  $b_i$ ,  $b_f$ ,  $b_c$ , and  $b_o$  express biased vectors.

### (2) Random Forest (RF)

RF expresses one combined learning algorithm containing decision trees, introduced by Breiman [40]. Thus, each tree of RF receives the training in a separate manner on an independent training set under the selection in a random manner. As opposed to the conventional decision tree, RF superiority has the reflection within two aspects. On the one hand, the trees built show inconsistency as generated in various training sets of a bootstrap subsampling and various random subdivided sets pertaining to characteristics for the split in terms of the respective tree node. On the other hand, the subsets of features are selected randomly. In this case, RF can achieve both low bias and low variance output and is not easy to fall into overfitting [64]. For regression, the final prediction is the average of the predictions from the set of decision trees. In RF, out-of-bag (OOB) error rate generally serves to be one way for measuring evaluation indices' significance [64].

### (3) Gradient boosting regression tree (GBRT)

In terms of an established set  $D$  with  $n$  examples and  $m$  features  $D = \{(x_i, y_i)\} (|D| = n, x_i \in \mathfrak{R}^m, y_i \in \mathfrak{R})$ , a tree ensemble approach applies  $K$  additive functions for predicting the output.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in \Gamma, \quad (\text{A.4})$$

where  $\Gamma = \{f(x) = \omega_{q(x)}\} (q: \mathfrak{R}^m \rightarrow T, \omega \in \mathfrak{R}^T)$  denotes the regression tree space. Specifically,  $q$  denotes the configuration of the respective tree mapping one instance to the relevant leaf index,  $T$  expresses the number of leaves in the tree, and the respective  $f_k$  represents one single tree structure  $q$  and leaf weight  $\omega$ . Inconsistent with decision trees, the respective regression tree covers one continuous score on the respective leaf, and  $\omega_i$  is used for representing the score on the  $i$ th leaf.

## B. Evaluation Criteria

Model performance evaluation criteria MAPE and RMSE have the following formulation:

$$\begin{aligned} \text{MAPE} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_r - Y_p}{Y_r} \right|, \\ \text{RMSE} &= \sqrt{\frac{\sum_{t=1}^n (Y_r - Y_p)^2}{n}}, \end{aligned} \quad (\text{B.1})$$

where  $n$  denotes the size pertaining to the dataset under the testing process and  $Y_r$  and  $Y_p$  represent the actual and forecasted data at time  $t$ , separately. Generally, the lower the RMSE and MAPE data are, the more accurate the approach will be. Nevertheless, specific to a set predicting process, practical results over and under the forecast receive the asymmetrical treat when using MAPE and RMSE [55]. Thus, direction matching rate (Dsta) is used to measure the data fluctuations (e.g., upward, stable, or downward), which is defined by

$$\begin{aligned} \text{Dsta} &= \frac{1}{n} \sum_{t=1}^n u(t), \\ u(t) &= \begin{cases} 1, & (Y_r(t+1) - Y_r(t)) * (Y_p(t+1) - Y_p(t)) \geq 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (\text{B.2})$$

The range of Dsta value is [0, 1]. The more approaching the Dsta data is to 1, the greater the precision pertaining to the direction-related predicting process concerned with the approach will be, and vice versa.

In addition, we utilize mean absolute scaled error (MASE) for assessing whether the proposed predicting approach outperforms naïve forecasting method [56].

$$\text{MASE} = \text{mean} \left| \frac{|e_t|}{(1/(N-1)) \sum_{i=2}^N |Y_p - Y_{p-1}|} \right|, \quad (\text{B.3})$$

where  $e_t$  denotes the prediction error determined to be  $(Y_r - Y_p)$  and  $(Y_p - Y_{p-1})$  refers to the naïve forecast's error. That is, MASE less than 1 indicates that the approach generates a more effective prediction than the calculated naïve predictions.

## Data Availability

The data applied here originate from the public Wind Economic Database, IEA, National Bureau of Statistics of China, and Statista, and the data could receive the assessment from <https://www.wind.com.cn/en/edb.html>, <https://www.iea.org/data-and-statistics>, <http://www.stats.gov.cn/tjsj/>, and <https://www.statista.com/statistics>, respectively.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was financially supported by the Key Research and Development Plan of the Ministry of Science and Technology, China (no. 2018YFB1601402), the Energy Foundation (project: Preliminary Investigation and Research on Freight Industry, no. VQ99QT1136Y17005), and the Open Research Fund for National Engineering Laboratory of Integrated Transportation Big Data Application Technology (no. CTBDAT201912).

## References

- [1] M. Stopford, "How shipping has changed the world and the social impact of shipping," *Global Maritime Environmental Congress*, vol. 7, pp. 1–10, 2010.
- [2] SSE, "About CBCFI," 2020, [https://en.sse.net.cn/indices/introduction\\_cbcfi\\_new.jsp](https://en.sse.net.cn/indices/introduction_cbcfi_new.jsp).
- [3] SSE, *Popularization of CBF*, SSE, Perth, UK, 2020.
- [4] X. Ding, S. Dai, F. Chen et al., "Long memory and scaling behavior study of bulk freight rate volatility with structural breaks," *Transportation Letters*, vol. 10, no. 6, pp. 343–353, 2018.
- [5] Y. Kou, M. Luo, and Y. Zhao, "Examining the theoretical-empirical inconsistency on stationarity of shipping freight rate," *Maritime Policy & Management*, vol. 45, no. 2, pp. 145–158, 2018.
- [6] J. Liu and F. Chen, "Asymmetric volatility varies in different dry bulk freight rate markets under structure breaks," *Physica A: Statistical Mechanics and Its Applications*, vol. 505, pp. 316–327, 2018.
- [7] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *NPJ Computational Materials*, vol. 5, no. 1, p. 83, 2019.
- [8] D. Wolff and U. Neugebauer, "Tree-based machine learning approaches for equity market predictions," *Journal of Asset Management*, vol. 20, no. 4, pp. 273–288, 2019.
- [9] J. Yang, C. Zhao, H. Yu, and H. Chen, "Use GBDT to predict the stock market," *Procedia Computer Science*, vol. 174, pp. 161–171, 2020.
- [10] S. Yang, J. Wu, Y. Du, Y. He, and X. Chen, "Ensemble learning for short-term traffic prediction based on gradient boosting machine," *Journal of Sensors*, vol. 2017, Article ID 7074143, 15 pages, 2017.
- [11] X. Zhang, M. Y. Chen, M. G. Wang, Y. E. Ge, and H. E. Stanley, "A novel hybrid approach to Baltic dry index forecasting based on a combined dynamic fluctuation network and artificial intelligence method," *Applied Mathematics and Computation*, vol. 361, pp. 499–516, 2019.
- [12] BalticExchange, "BDI overview," 2021, <https://www.balticexchange.com/en/index.html>.
- [13] W. Hansen, "Understanding the Baltic dry index. Learning markets," 2021, <https://www.learningmarkets.com/understanding-the-baltic-dry-index/>.
- [14] K. P. B. Cullinane, K. J. Mason, and M. Cape, "A comparison of models for forecasting the baltic freight index: box-Jenkins revisited," *International Journal of Maritime Economics*, vol. 1, no. 2, pp. 15–39, 1999.
- [15] M. G. Kavussanos and A. H. Alizadeh-M, "Seasonality patterns in dry bulk shipping spot and time charter freight rates," *Transportation Research Part E: Logistics and Transportation Review*, vol. 37, no. 6, pp. 443–467, 2001.

- [16] R. Batchelor, A. Alizadeh, and I. Visvikis, "Forecasting spot and forward prices in the international freight market," *International Journal of Forecasting*, vol. 23, no. 1, pp. 101–114, 2007.
- [17] V. Tsioumas, S. Papadimitriou, Y. Smirlis, and S. Z. Zahran, "A novel approach to forecasting the bulk freight market," *The Asian Journal of Shipping and Logistics*, vol. 33, no. 1, pp. 33–41, 2017.
- [18] S. Chen, H. Meersman, and E. Voorde, "Forecasting spot rates at main routes in the dry bulk market," *Maritime Economics & Logistics*, vol. 14, 2012.
- [19] R. Adland, F. E. Benth, and S. Koekebakker, "Multivariate modeling and analysis of regional ocean freight rates," *Transportation Research Part E: Logistics and Transportation Review*, vol. 113, pp. 194–221, 2018.
- [20] M. Stopford, *Maritime Economics*, Routledge, Evanston, IL, USA, 2013.
- [21] S. Dai, F. Chen, Y. Zeng, and X. Zeng, "Scaling behavior of bulk freight rate volatility before and after noise reduction," *Journal of Shanghai Jiaotong University (Science)*, vol. 21, no. 6, pp. 655–661, 2016.
- [22] J. Li and M. G. Parsons, "Forecasting tanker freight rate using neural networks," *Maritime Policy & Management*, vol. 24, no. 1, pp. 9–30, 1997.
- [23] H. Yang, F. Dong, and M. Ogandaga, "Forewarning of freight rate in shipping market based on support vector machine," in *Proceedings of the 6th International Conference of Traffic and Transportation Studies Congress (ICTTS)*, Nanning, China, August 2008.
- [24] E. Thalassinos, M. Haniyas, P. Curtis, and J. Thalassinos, *Marine Navigation and Safety of Sea Transportation: STCW, Maritime Education and Training (MET), Human Resources and Crew Manning, Maritime Policy, Logistics and Economic Matters*, CRC Press, Boca Raton, FL, USA, 2013.
- [25] F. Guan, Z. Peng, K. Wang, X. Song, and J. Gao, "Multi-step hybrid prediction model of Baltic supermax index based on support vector machine," *Neural Network World*, vol. 26, no. 3, pp. 219–232, 2016.
- [26] B. Şahin, S. Gürgen, B. Ünver, and İ. Altın, "Forecasting the Baltic dry index by using an artificial neural network approach," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 26, pp. 1673–1684, 2018.
- [27] D. M. Q. Nelson, A. C. M. Pereira, and R. A. d. Oliveira, "Stock market's price movement prediction with LSTM neural networks," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1419–1426, Anchorage, AK, USA, May 2017.
- [28] C. Wu, C. Lu, Y. Ma, and R. Lu, "A new forecasting framework for bitcoin price with LSTM," in *Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 168–175, Singapore, November 2018.
- [29] Z. Zhao, W. Chen, X. Wu, P. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, pp. 68–75, 2017.
- [30] Y. Duan, L. V. Yisheng, and F.-Y. Wang, "Travel time prediction with LSTM neural network," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, November 2016.
- [31] Y. Leonov and V. Nikolov, "A wavelet and neural network model for the prediction of dry bulk shipping indices," *Maritime Economics & Logistics*, vol. 14, no. 3, pp. 319–333, 2012.
- [32] E. Bulut, O. Duru, and S. Yoshida, "A fuzzy integrated logical forecasting (FILF) model of time charter rates in dry bulk shipping: a vector autoregressive design of fuzzy time series with fuzzy c-means clustering," *Maritime Economics & Logistics*, vol. 14, no. 3, pp. 300–318, 2012.
- [33] Q. Zeng, C. Qu, A. K. Y. Ng, and X. Zhao, "A new approach for Baltic dry index forecasting based on empirical mode decomposition and neural networks," *Maritime Economics & Logistics*, vol. 18, no. 2, pp. 192–210, 2016.
- [34] K. Uyar, Ü. İlhan, and A. İlhan, "Long term dry cargo freight rates forecasting by using recurrent fuzzy neural networks," *Procedia Computer Science*, vol. 102, pp. 642–647, 2016.
- [35] I. M. Kamal, H. Bae, S. Sunghyun, and H. Yun, "DERN: deep ensemble learning model for short- and long-term prediction of baltic dry index," *Applied Sciences*, vol. 10, no. 4, p. 1504, 2020.
- [36] D. R. Liu, S. J. Lee, Y. Huang, and C. J. Chiu, "Air pollution forecasting based on attention-based LSTM neural network and ensemble learning," *Expert Systems*, vol. 37, Article ID e12511, 2020.
- [37] M. Tan, S. Yuan, S. Li, Y. Su, H. Li, and F. H. He, "Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2937–2948, 2020.
- [38] L. Duan and M. Binbasioglu, "An ensemble framework for community detection," *Journal of Industrial Information Integration*, vol. 5, pp. 1–5, 2017.
- [39] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [40] L. Breiman, "Random forests," *Machine LearningMachine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [41] Wind <https://www.wind.com.cn/en/default.html> 2021.
- [42] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters*, vol. 7, no. 4, pp. 313–318, 1981.
- [43] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427–431, 1979.
- [44] N. Davies and J. D. Petrucci, "Detecting non-linearity in time series," *The Statistician*, vol. 35, no. 2, pp. 271–280, 1986.
- [45] R. S. Tsay, "Chapter 4.2 nonlinear tests," in *Analysis of Financial Time Series*, Wiley, Hoboken, NJ, USA, 3rd edition, 2010.
- [46] L. Yang, C. Haozhou, R. Yuqi, and Y. Kunpneg, *Analysis of Factors Influencing Coal Prices under the New Normal of China's Economy*, Reseach Square, Durham, NC, USA, 2020.
- [47] S. Cornot-Gandolphe, *China's Coal Market: Can Beijing Tame "King Coal"?*, Oxford Institute for Energy Studies, Oxford, UK, 2014, <https://www.oxfordenergy.org/wpcms/wp-content/uploads/2014/12/CL-11.pdf>.
- [48] N. Zamani, "The relationship between crude oil and coal markets: a new approach," *International Journal of Energy Economics and Policy*, vol. 6, no. 4, pp. 801–805, 2016.
- [49] X. Guo, J. Shi, and D. Ren, "Coal price forecasting and structural analysis in China," *Discrete Dynamics in Nature and Society*, vol. 2016, Article ID 1256168, 7 pages, 2016.
- [50] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, 2018.
- [51] N. Reimers and I. Gurevych, "Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks," 2017, <https://arxiv.org/abs/1707.06799>.

- [52] G. H. T. Ribeiro, P. S. G. d. M. Neto, G. D. C. Cavalcanti, and I. R. Tsang, "Lag selection for time series forecasting using particle swarm optimization," in *Proceedings of the 2011 International Joint Conference on Neural Networks*, pp. 2437–2444, San Jose, CA, USA, 2011.
- [53] Scikit-learn.org, "Parameter estimation using grid search with cross-validation—Scikit-learn 0.19.1 documentation," 2021, [http://scikit-learn.org/stable/auto\\_examples/model\\_selection/plotgrid\\_search\\_digits.html](http://scikit-learn.org/stable/auto_examples/model_selection/plotgrid_search_digits.html).
- [54] X. Li, L. Peng, X. Yao et al., "Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation," *Environmental Pollution*, vol. 231, pp. 997–1004, 2017.
- [55] J. McKenzie, "Mean absolute percentage error and bias in economic forecasting," *Economics Letters*, vol. 113, no. 3, pp. 259–262, 2011.
- [56] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
- [57] X. Diebold Francis and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 134–144, 1995.
- [58] G. Li, H. Song, and S. F. Witt, "Recent developments in econometric modeling and forecasting," *Journal of Travel Research*, vol. 44, no. 1, pp. 83–99, 2005.
- [59] WorldOil, "China buys millions of barrels of U.S. oil to comply with trade deal," 2021, <https://www.worldoil.com/news/2020/8/18/china-buys-millions-of-barrels-of-us-oil-to-comply-with-trade-deal>.
- [60] Energy Information Administration, "Benchmarks play an important role in pricing crude oil," 2020, <https://www.eia.gov/todayinenergy/detail.php?id=18571>.
- [61] D. Workman, "Top 15 crude oil suppliers to China," 2021, <http://www.worldstopexports.com/top-15-crude-oil-suppliers-to-china/>.
- [62] S. Nanovsky, "The impact of oil prices on trade," *Review of International Economics*, vol. 27, no. 1, Article ID e0001, 2019.
- [63] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [64] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, no. 1, p. 3, 2006.