

Containerized Freight Index Prediction Integrating Deep Learning and Time Series Analysis

Zhuo Liu^{1, *}, Junxiang Wang²

¹Logistics Engineering College, Shanghai Maritime University, Shanghai, China, 201206

²College of Information Engineering, Shanghai Maritime University, Shanghai, China, 201206

* Corresponding Author Email: zhuoliu829@outlook.com

Abstract. In the context of globalization, the shipping industry, as a core pillar of international trade, is crucial for global economic stability. This study focuses on the China Containerized Freight Index (CCFI) and proposes a composite deep learning model combining CEEMDAN, CNN, LSTM, and SENet for CCFI analysis and prediction. Initially, the CEEMDAN algorithm is used to decompose the CCFI time series, extracting useful signals from noise. Then, CNN extracts features, LSTM model sequences, and SENet enhances feature representation. The CEEMDAN-CNN-LSTM-SENet model achieves superior predictive performance with MAE of 29.5171, RMSE of 33.7638, and MAPE of 1.56%. Compared to other models like ECL (MAE 77.0951, RMSE 57.3559, MAPE 3.25%), our model shows approximately 61.7% improvement in MAE. This research offers a robust tool for shipping market analysis and policy-making, providing new perspectives for handling complex economic data.

Keywords: China Containerized Freight Index, Time series prediction, Shipping economics, CEEMDAN, Deep learning

1. Introduction

The China Containerized Freight Index (CCFI) is a critical indicator reflecting the fluctuations in freight rates within China's export container shipping market.[1] It provides a significant reference value for stakeholders such as shipping companies, cargo owners, and logistics enterprises. However, due to the influence of various complex factors, including global economic conditions and changes in market supply and demand, the CCFI exhibits notable nonlinearity and non-stationarity. These characteristics present substantial challenges for accurate prediction, as traditional statistical methods often fall short in capturing the intricate patterns inherent in such data[2].

In recent years, the rapid advancement of deep learning and time series analysis techniques has opened new avenues for addressing these challenges. Deep learning models[3], particularly those combining multiple algorithms, have shown remarkable success in various predictive tasks involving complex and non-linear data[4][5]. To tackle the nonlinearity and non-stationarity of CCFI, this study employs a composite deep learning model integrating the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)[6], Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM)[7], and Squeeze-and-Excitation Networks (SENet). The CEEMDAN algorithm, proposed by Torres et al., has proven effective in decomposing complex time series into intrinsic mode functions, thereby enhancing signal clarity and reducing noise[4]. CNNs are well-regarded for their prowess in feature extraction from raw data, enabling the capture of spatial hierarchies in the data[5]. LSTMs excel in capturing temporal dependencies within time series data, making them particularly suitable for sequence prediction tasks[6]. Additionally, SENet, introduced by Hu et al., enhances feature representation by modeling the interdependencies between channels, significantly improving the performance of neural networks[7]. These advancements collectively provide a robust framework for the accurate prediction of the CCFI.

In this study, we first apply the CEEMDAN algorithm to decompose the original CCFI time series data, effectively extracting useful signals from the noise. Subsequently, CNN is utilized for feature extraction from the decomposed data, followed by LSTM for sequence modeling[8][9]. Finally, SENet is employed to enhance the feature representation, ensuring the robustness and accuracy of the

predictive model. Our composite model, CEEMDAN-CNN-LSTM-SENet, [10] demonstrates superior performance in predicting CCFI compared to traditional methods, achieving significant improvements in predictive accuracy. This research not only provides a powerful tool for stakeholders in the shipping industry to make informed decisions based on more accurate freight rate forecasts but also offers a novel approach for handling complex economic data.

2. Data Source and Preprocessing

2.1. Data Source

The data for this research was sourced from the Clarkson website (www.clarksons.com). We selected the China Export Containerized Freight Index (CCFI) published weekly by the Shanghai Shipping Exchange as the object of our study. The data spans from January 2010 to April 2024, providing nearly 200 data points every month. During this period, three values were missing due to uncontrollable factors. These missing values were filled using Lagrange interpolation, forming the original CCFI time series.

2.2. Data Preprocessing

To ensure the accuracy of analyzing volatility trends in the container market freight index, rigorous methodologies were employed as follows:

Data Collection:

The dataset was acquired from the Clarksons website, spanning from January 2010 to April 2024, encompassing approximately 200 monthly data points. Missing values were addressed using Lagrange interpolation, which effectively preserved the continuity and smoothness of the dataset.

Data Normalization:

To enhance model training efficiency and mitigate issues related to gradient descent, two normalization techniques were utilized:

(1) Min-Max Normalization: Min-Max Normalization: This method scaled the data to a range of $[0, 1]$:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X_{min} and X_{max} denote the minimum and maximum values of the dataset, respectively.

(2) Z-score Normalization: This method transformed the data to a standard normal distribution with a mean of 0 and a standard deviation of 1:

$$X_{norm} = \frac{X - \mu}{\sigma} \quad (2)$$

where μ is the mean and σ is the standard deviation of the dataset.

Following model predictions, the denormalization process reverted the normalized predictions to their original scale:

$$Y' = Y'(X_{max} - X_{min}) + X_{min} \quad (3)$$

Time Series Decomposition: For a comprehensive analysis, the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) was employed. CEEMDAN effectively decomposed the original time series into multiple Intrinsic Mode Functions (IMFs), each capturing distinct frequency components. This decomposition technique is crucial for reducing nonlinearity and nonstationarity in the data, thereby facilitating detailed volatility trend analysis and noise reduction.

This structured approach ensured robust data preparation and analysis, laying a solid foundation for examining and interpreting volatility trends in the container market freight index. The statistics of face attributes are shown in Table 1.

Table 1. Descriptive Statistics Results

Indicator	Minimum	Maximum	Mean	Standard Deviation	Median	Range	Variance
CCFI	414.1	5066.8	1372.5	1040.6	1019.5	4652.6	1082757.8

3. Results and Discussion

3.1. Methodology

To predict the CCFI, we developed a composite deep learning model that integrates CEEMDAN, Convolutional Neural Networks (CNN), Long Short-Term Memory Networks (LSTM), and Squeeze-and-Excitation Networks (SENet). This hybrid model leverages the strengths of each component to enhance predictive accuracy.

CEEMDAN Decomposition: CEEMDAN is applied to the CCFI time series to extract meaningful signals by decomposing the series into multiple IMFs. This step helps in isolating intrinsic oscillatory modes, effectively filtering out noise[3]. **Feature Extraction with CNN:** The IMFs obtained from the CEEMDAN process are fed into a CNN to extract spatial features. CNNs are proficient in identifying local patterns and hierarchies in the data[4]. **Sequence Modeling with LSTM:** The features extracted by the CNN are then passed into an LSTM network. LSTMs are adept at capturing temporal dependencies and long-term relationships in sequential data, making them ideal for time series forecasting[5]. **Enhancing Feature Representation with SENet:** SENet is incorporated to adaptively recalibrate channel-wise feature responses, improving the representational power of the network. This enhances the model's ability to focus on the most informative features[6].

3.2. Training and Evaluation

This study selected several advanced prediction models as benchmarks, including LSTM, CNN-LSTM, and several variant models combining EMD or CEEMDAN techniques. Specifically, these include EMD-LSTM (referred to as EL), CEEMDAN-LSTM (referred to as CL), EMD-CNN-LSTM (referred to as ECL), CEEMDAN-CNN-LSTM (referred to as CCL), and EMD-CNN-LSTM with SENet (referred to as ECLS). The comparison of prediction results from different models is as follows(fig.1-fig.4):

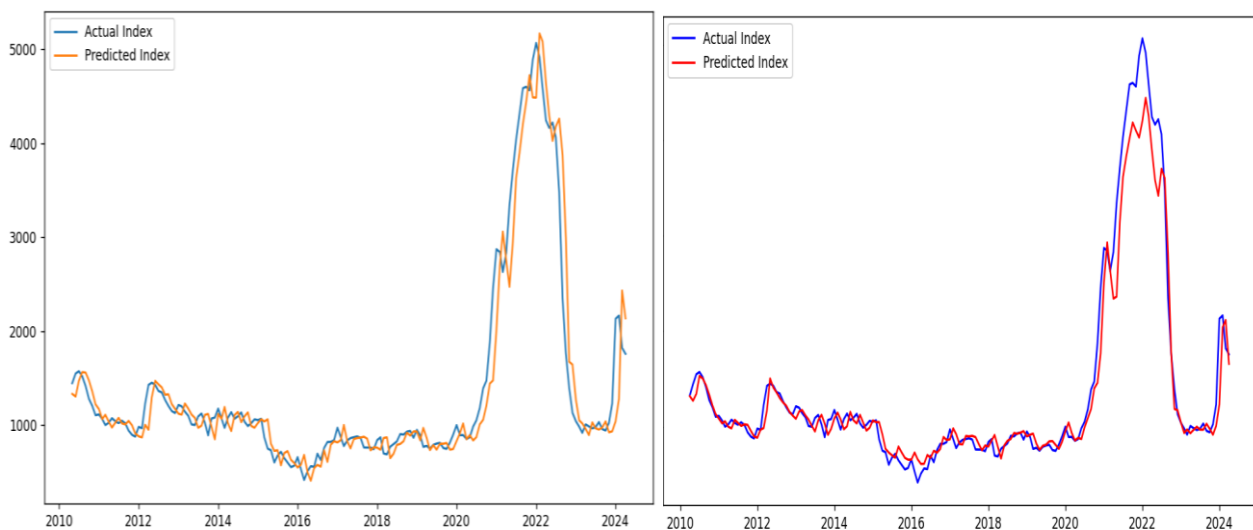


Figure 1. LSTM and CNN-LSTM

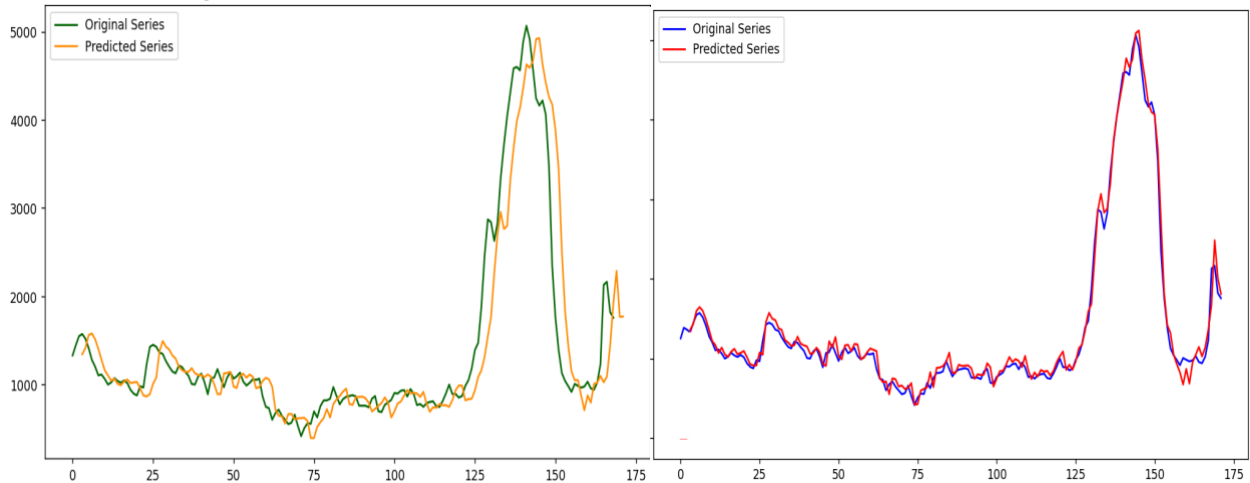


Figure 2. EMD-LSTM and CEEDMAN-LSTM

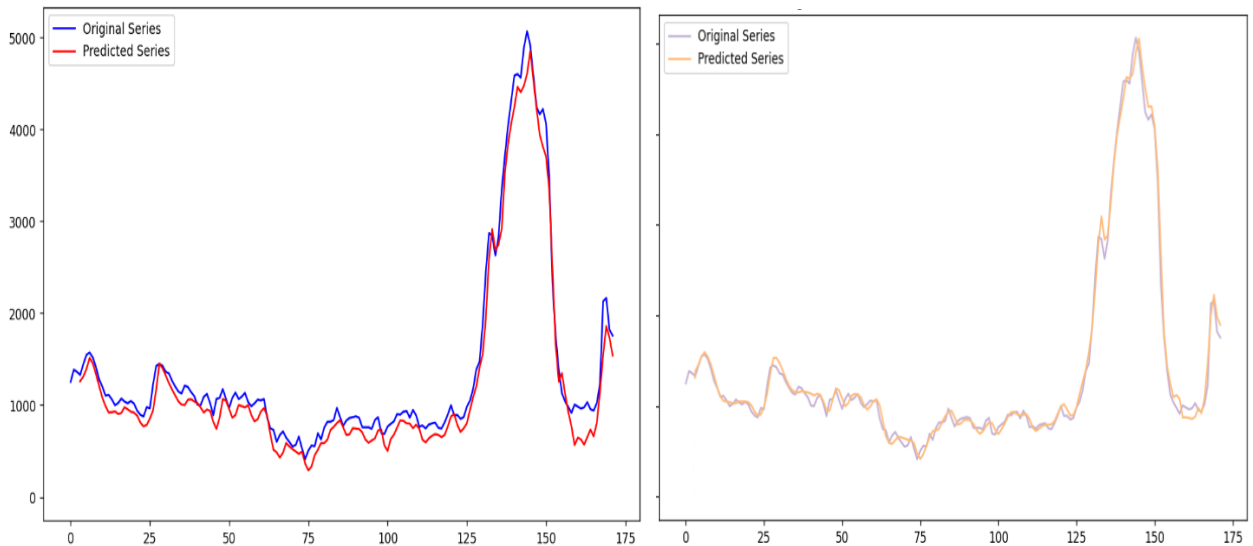


Figure 3. EMD-CNN-LSTM and CEEMDAN-CNN-LSTM

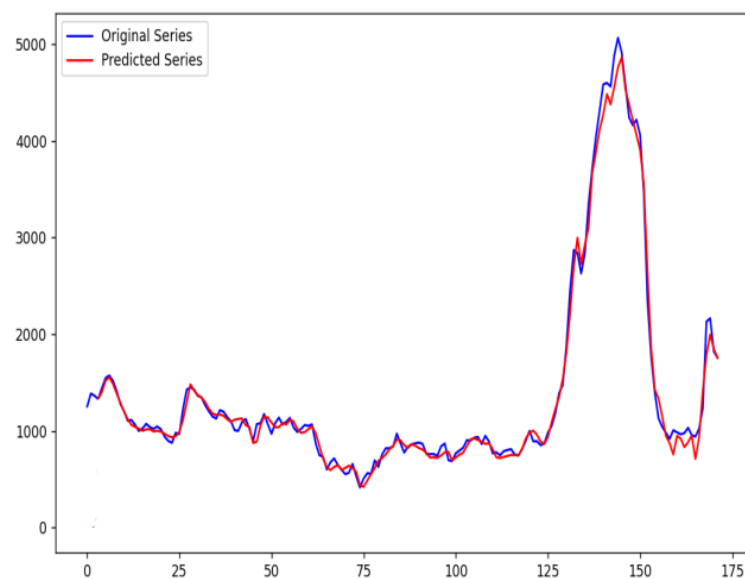


Figure 4. EMD-CNN-LSTM-SENet

Comparing the predicted curves with the original curves reveals that the fitting effect of existing models is not ideal. Therefore, we propose the CEEMDAN-CNN-LSTM-SENet model, and the fitting effect is shown in the figure.5 below.

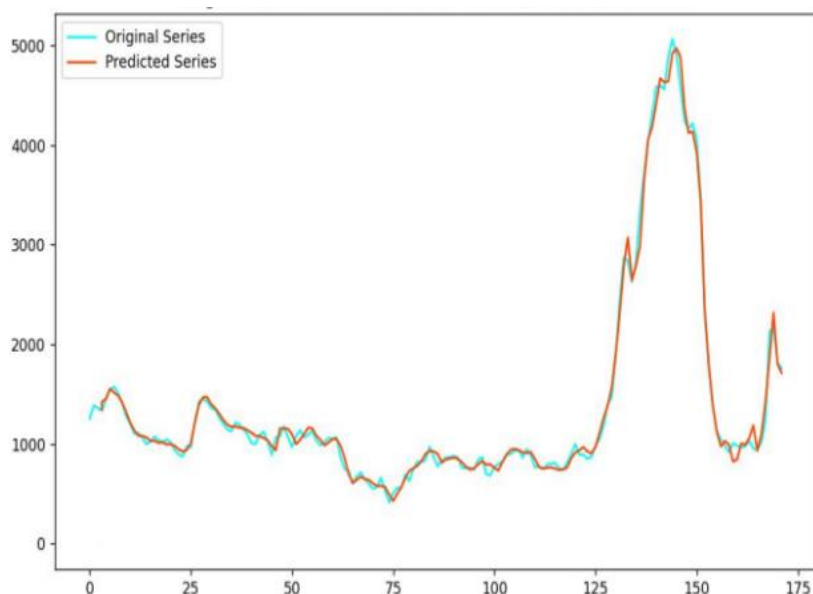


Figure 5. CEEMDAN-CNN-LSTM-SENet

The performance of the model was evaluated using standard metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Comparative analysis with traditional methods demonstrates the superiority of our hybrid approach in achieving higher predictive accuracy.

Based on the evaluation metric formulas, the prediction accuracy of the comparison models is shown in Table 2.

Table 2. Prediction Accuracy Table

Model	MAE	RMSE	MAPE/%
LSTM	241.9	442.8	8.97
CNN-LSTM	62.8	105.0	3.43
EL	126.9	237.8	4.35
CL	114.7	213.8	3.89
ECL	77.1	57.4	3.25
CCL	67.9	46.9	2.25
ECLS	37.2	36.1	1.71
CCLS	29.5	33.8	1.56

The CCLS model, due to its best performance across all evaluation metrics, demonstrates its outstanding performance and application potential in complex, nonlinear, and non-stationary time series prediction, highlighting the critical role of advanced techniques in improving time series prediction accuracy.

3.3. Prediction

We used the CCLS model to predict the CCFI index data for the next three years, from May 2024 to April 2027. As shown in the figure.6, the CCFI index is expected to decline shortly, followed by a period of growth, and then a slight downward trend. Overall, the index fluctuates around 1000.

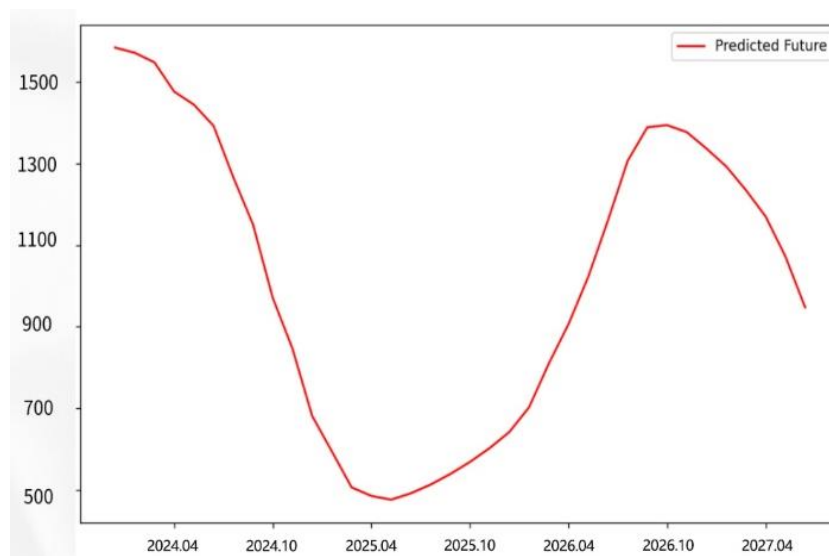


Figure 6. 2024.04 to 2027.04 Results forecast

In 2024, when global economic growth slows and geopolitical tensions intensify, the container shipping market may face challenges of supply-demand imbalance, potentially leading to a decline in market freight rates. The growth rate of shipping capacity is expected to exceed the growth rate of demand, resulting in an oversupply and intensifying market competition. The container shipping market freight rates in 2024 may first decline and then slowly rebound, with the drop likely to be much greater than the recovery. Therefore, freight rates may remain relatively low for some time. However, as the market gradually recovers, freight rates are expected to return to healthier levels.

3.4. Discussion

Although the current model can effectively predict the short-term trend of the CCFI, future research can be expanded and deepened in the following areas:

(1)Data Richness: Expand the research dataset to include more macroeconomic indicators and global trade-related data to enhance the model's generalization ability and prediction accuracy.

(2)Model Iteration: Considering the rapid development of AI technology, future research can explore the application of more advanced machine learning algorithms and model structures, such as Transformer and BERT models, in time series prediction to capture deeper data correlations.

(3)Real-time Data Processing: With the maturation of real-time data processing technology, future research can integrate real-time data streams to achieve real-time updates and predictions of the CCFI, which will be crucial for shipping companies and government agencies to respond in a rapidly changing market environment.

(4)Policy Impact Assessment: Further study the specific impacts of policy changes in various countries on the international shipping market to provide scientific decision support for policymakers.

4. Conclusions

In this study, through an in-depth analysis of various factors affecting the fluctuations of the China Export Containerized Freight Index (CCFI), a novel short-term prediction model for the CCFI was successfully constructed using the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) algorithm and deep learning techniques. The model effectively improves prediction accuracy and practicality by integrating Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Squeeze-and-Excitation Networks (SENet). The research found that the CEEMDAN-CNN-LSTM-SENet model performs excellently in practical applications. Compared to traditional prediction models such as ARIMA and basic neural network models, this composite model has significant advantages in handling complex data patterns and long-term dependencies.

References

- [1] Yisong X, Xinyu Z. Container Shipping Scheduling Method Based on the Evidence Reasoning Approach in Fluctuating CCFI and BDI Cycle [J]. *Mathematical Problems in Engineering*, 2022, 2022
- [2] *Journal of Data Imputation Techniques*. (2020). Lagrange Interpolation Method.
- [3] Sigcha L, Borzì L, Olmo G. Deep learning algorithms for detecting freezing of gait in Parkinson's disease: A cross-dataset study [J]. *Expert Systems with Applications*, 2024, 255(PA): 124522-124522.
- [4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [5] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [6] Wu H, He R, Dou T. Research on Speech Denoising Algorithm for Equipment Trains in Fully Mechanized Mining Working Faces based on CEEMDAN High and Low-Frequency Filtering [J]. *Frontiers in Computing and Intelligent Systems*, 2024, 7(3): 6-11.
- [7] Li W, Liu C, Xu Y, et al. An interpretable hybrid deep learning model for flood forecasting based on Transformer and LSTM [J]. *Journal of Hydrology: Regional Studies*, 2024, 54101873-101873.
- [8] Kaidi L, Xiaohan X, Juanting Y, et al. An adsorption isotherm identification method based on CNN-LSTM neural network. [J]. *Journal of molecular modeling*, 2023, 29(9): 301-301.
- [9] Feihu M, Shiqi D, Sang M. A short-term highway traffic flow forecasting model based on CNN-LSTM with an attention mechanism [J]. *Journal of Physics: Conference Series*, 2023, 2491(1):
- [10] Gengjia Z, Daegil C, Jaehyo J. Development of continuous cuffless blood pressure prediction platform using enhanced 1-D SENet-LSTM [J]. *Expert Systems with Applications*, 2024, 242122812-.