Proceedings of The 6th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/2026.30497

Research on Commodity Futures Price Fluctuation Prediction Based on CNN-BiLSTM-Attention

Yixin Wu

Xinjiang University, Urumqi, China
1909954126@qq.com

Abstract. To meet the high-precision demand for short-term forecasting in the commodity futures market and address the challenges of sharp price fluctuations and nonlinear characteristics, this study takes seven representative commodity futures varieties in China (including rebar, iron ore, etc.) as research objects and constructs a prediction system for price fluctuations in the next 30 minutes based on 1-minute high-frequency main contract data. In the data preprocessing stage, multiple rounds of cleaning, outlier removal, sliding window sampling, and construction of multidimensional temporal features are conducted to ensure the information content and stability of the input sequence. The model is constructed with a progressive design: LSTM is used to establish a baseline, CNN-LSTM-Attention is adopted to enhance the ability of extracting key patterns, and finally the Reversible Instance Normalization (RevIN) mechanism is integrated to form the RevIN-CNN-LSTM-Attention main model. Experimental results show that the model performs excellently on the test set, with an average R^2 increase of more than 10% compared with the basic model and a significant decrease in RMSE, and the trend fitting effect is outstanding in varieties such as ferrosilicon and rebar. This paper also analyzes the advantages and limitations of the model, clarifies its applicable boundaries, and proposes subsequent optimization directions. The research indicates that the proposed model has good stability, adaptability, and generalization ability, providing an effective solution for high-frequency commodity futures prediction and reference value for relevant financial engineering modeling.

Keywords: Commodity Futures Prediction, Short-Term Price Fluctuation, LSTM, Attention Mechanism, RevIN Normalization

1. Introduction

With the rapid development of high-frequency trading technology and intelligent investment research methods, the demand for short-term forecasting accuracy in the commodity futures market is constantly increasing. Faced with the sharp fluctuations and nonlinear characteristics of commodity prices, how to construct a short-term prediction model with stability, adaptability, and generalization ability has become an important issue in financial modeling and machine learning research. Taking China's mainstream commodity futures market as the background, this paper selects seven representative varieties (including Rebar (RB), Hot-Rolled Coil (HC), Iron Ore (I), Coking Coal (JM), Silicomanganese (SM), Ferrosilicon (SF), and Stainless Steel (SS)), and

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0
(https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.30497

constructs a sequence prediction system for forecasting price fluctuations in the next 30 minutes based on 1-minute granularity high-frequency main contract data.

In terms of data preprocessing, this paper conducts multiple rounds of cleaning and splicing on the original data, removes abnormal fluctuation intervals, and unifies the scale; constructs supervised learning samples through a sliding window mechanism, and builds multidimensional temporal features by combining technical indicators, statistics, lag structures, and other methods to ensure that the input sequence has stability while taking into account information content. In terms of model construction, three types of progressive structures are designed in turn: the basic LSTM model is used to establish a baseline for time-dependent modeling; the enhanced CNN-LSTM-Attention model introduces convolution and attention mechanisms to improve the ability of extracting key patterns; finally, the Reversible Instance Normalization (RevIN) mechanism is introduced to construct the main model RevIN-CNN-LSTM-Attention integrating normalization and attention, which is used for unified modeling of complex distributions and heterogeneous data.

In the experimental part, independent models are trained for the seven varieties respectively, and the prediction effect is quantitatively evaluated on the test set using two indicators: R^2 and RMSE. The results show that RevIN and the attention mechanism have a significant promoting effect on model performance. The optimized model achieves an increase in R^2 (an average increase of more than 10%) and a decrease in RMSE in most varieties, especially in SF, RB, and SM, with strong trend fitting ability and good error convergence. In addition, this paper analyzes the advantages and limitations of the model from multiple perspectives, including structural contribution analysis, attribution of variety prediction differences, and indicator interpretation contradictions, clarifies the applicable boundary of the model on trend-based commodities, and puts forward a number of subsequent optimization suggestions including Transformer structure transformation, external feature enhancement, and loss function reconstruction.

In summary, the RevIN-CNN-LSTM-Attention model proposed in this paper has good transferability and practicability in the price fluctuation prediction task of multiple commodity futures, has certain reference value for high-frequency financial data modeling, time series structure optimization, and error control strategies, and can provide theoretical basis and practical templates for relevant financial engineering modeling.

2. Data preprocessing and feature construction

To effectively predict the price fluctuation of commodity futures in the next 30 minutes, this paper first systematically preprocesses the provided 1-minute frequency main contract data to ensure the accuracy, continuity, and consistency of the original data.

The entire preprocessing process mainly includes six steps: data decompression, merging, cleaning, time period screening, yield calculation, and segmented saving.

Data Decompression and Merging. The original data is provided in .csv.gz format. First, the Python standard library is used to batch decompress all compressed files. Then, multiple CSV files of the same variety are merged into a single file in chronological order. In this process, the timestamp field is uniformly parsed and converted to datetime type, and sorted by time to ensure the temporal continuity of the data.

Time Period Screening and Trading Session Extraction. Considering the obvious time period interference in futures market transactions, we only retain data from the following three main trading sessions: morning session: 09:00–11:30, afternoon session: 13:30–15:00, night session: 21:00–23:00. Data from each session is extracted separately using pandas' between_time function and merged into a new valid dataset.

DOI: 10.54254/2755-2721/2026.30497

Outlier Detection and Cleaning. To improve the quality and stability of modeling data, we design a set of multi-indicator-based outlier detection mechanisms covering core variables such as price, trading volume, and open interest [1,2]. Specifically include:

Opening price (open): Calculate the z-score based on logarithmic yield, and remove abnormal fluctuations exceeding 3 times the standard deviation;

Highest price and lowest price (high, low): Detect outliers based on the price range (high - low);

Closing price (close): Identify local anomalies using the logarithmic yield fluctuation (z-score) of a rolling window (length = 20);

Trading volume (volume): Detect abnormal trading volume records using the interquartile range (IQR) method;

Open interest (open interest): Construct a rolling standard deviation based on the rate of change to detect significant mutation points. All outlier detection methods are based on statistical criteria (such as z-score, IQR), combined with the dynamic characteristics of rolling average and long-distance windows to ensure that structural anomalies rather than short-term fluctuations are removed.

To improve the quality and stability of model training samples, we design a set of statistical rule-based multidimensional outlier detection mechanisms covering fields such as opening price, highest price, lowest price, closing price, trading volume, and open interest. Except for trading volume, all other fields use the z-score method (i.e., standard deviation rule) for outlier removal; among them, the trading volume data presents a skewed distribution, so the IQR (interquartile range) method is more robust.

Specifically, the outlier detection of trading volume is based on the following IQR rule:

$$\text{outlier} = \{x \mid x < Q1 - 1.5 \cdot IQR \text{ or } x > Q3 + 1.5 \cdot IQR\}$$

where Q1 and Q3 are the 25th and 75th percentiles respectively, and $IQR = Q3 - Q1$. For other fields, the following methods are adopted:

For opening price and closing price, logarithmic yield is used for standardization;

For highest price and lowest price, detect whether they deviate significantly from the interval;

For open interest, z-score detection is performed using volatility (price fluctuation range).

All outliers are uniformly processed and removed, and the final retained data accounts for 90% of the total, as shown in Table 1, ensuring the integrity and representativeness of the data.

Table 1. Statistics of data cleaning for each variety

Variety	Original Data Volume	Cleaned Data Volume
SM	337,260	308,412
HC	508,980	467,439
I	509,606	465,878
JM	509,006	466,387
RB	509,070	466,471
SF	338,910	276,448
SS	349,977	309,186

The results show that most varieties retain more than 90% of the original data during cleaning, indicating that the overall data quality is high, and the outlier removal process effectively improves data quality without excessive damage to the sample structure. For varieties with a relatively low

DOI: 10.54254/2755-2721/2026.30497

cleaning ratio (such as SF and SS), it may be related to their market liquidity differences or price fluctuation characteristics, which will be paid attention to in subsequent model design.

Yield Calculation and Processing. Based on the cleaned data, we calculate the yield (price fluctuation range) for the next 30 minutes per minute according to the following formula [3]:

$$\text{Price Fluctuation Range} = \frac{\mathrm{P_{t+30}} - \mathrm{P_t}}{\mathrm{P_t}} \times 100\%$$

where $\mathrm{P_t}$ is the closing price of the current minute, and $\mathrm{P_{t+30}}$ is the closing price 30 minutes later. To avoid future data leakage and modeling bias, we set the price fluctuation rate of data within the first 30 minutes of each trading day as missing values (NaN), including 09:00–9:30, 13:30–14:00, and 21:00–21:30.

Group Saving by Variety. After completing the above steps, we uniformly merge all variety data into a single data table and save them in groups according to contract identifiers (such as "SS", "SF", etc.), facilitating subsequent models to separately process the feature distribution and prediction modeling of each variety.

Temporal Feature Construction and Observation of Data Sparsity. In the process of processing 1-minute-level trading data of commodity futures, we notice that its trading structure not only has obvious intraday segmented characteristics but also has sparse record phenomena in time granularity. These characteristics have an important impact on subsequent feature construction.

First, the trading day of the futures market is strictly divided into three main sessions: morning session (09:00–11:30), afternoon session (13:30–15:00), and night session (21:00–23:00). This time period structure requires us to label timestamp features for the model to identify differences in market behavior at different times.

In fact, although the data is labeled as "recorded once per minute", we observe that some time points (such as 09:15:00, 13:45:00, 22:00:00) do not exist in the data, indicating that data records are highly correlated with actual trading activities. Therefore, when constructing time-related features (such as sliding mean, delayed yield, etc.), priority should be given to aligning and synchronizing existing time recording points instead of mechanically traversing by minute.

Based on the above temporal features, we uniformly construct the following features using a 30-minute sliding window and a 30-minute lagged value:

• Sliding statistical features (historical window behavior):

-mean_close_30min: Mean closing price in the past 30 minutes

-std_close_30min: Standard deviation of closing price in the past 30 minutes

-mean_volume_30min, mean_oi_30min: Sliding mean of trading volume and open interest

• Lag features (delayed behavior):

-close_lag30: Closing price 30 minutes before the current moment

-return_lag30: Yield 30 minutes before the current moment

-volume_lag30, oi_lag30: Lagged values of trading volume and open interest

The above features are all constructed within the sliding time window, combined with the real timestamp of the current data record to ensure time consistency. In the case of missing records, some window features can also fallback to using longer windows instead. In addition, retaining record gaps also has modeling significance to prevent the introduction of misleading prediction signals.

The construction of such temporal features is the key to the short-term price fluctuation prediction task, which can provide the model with rich and intuitive behavioral context information.

DOI: 10.54254/2755-2721/2026.30497

3. Data feature analysis and statistical testing

To comprehensively understand the statistical characteristics and structural attributes of the closing price sequence of commodity futures, after data cleaning, this paper analyzes and verifies the data from three levels: basic distribution characteristics, stationarity characteristics, and nonlinear structural characteristics. The analysis results provide a theoretical basis for feature engineering and modeling strategies [4,5].

3.1. Descriptive statistical analysis

We conduct descriptive statistics on the closing prices of the seven commodity varieties, including indicators such as mean, standard deviation, skewness, kurtosis, extreme values, and value range. The statistical results are shown in the following table:

Table 2. Descriptive statistical metrics of closing prices for each variety

Variety	Start Date	Sample Size	Mean	Standard Deviation	Skewness	Kurtosis	Minimum Value	Maximum Value
HC	2017/01/03	509,680	3969.72	621.40	1.3817	1.8072	2815.00	6723.00
I	2017/01/03	465878,	727.66	177.36	0.5903	0.9690	450.00	1350.50
JM	2017/01/03	466,387	1556.11	459.11	1.4381	2.1484	850.00	2481.00
RB	2017/01/03	466,471	3705.70	562.82	0.3264	1.2460	2613.00	4930.00
SF	2017/02/23	276,448	7101.96	1184.93	0.4243	0.8553	4607.00	10349.00
SM	2017/02/23	308,412	7012.49	1163.13	0.2020	0.7007	4577.00	10000.00
SS	2019/09/25	309,186	12507.43	1163.13	0.1927	0.9499	11705.00	23845.00

The analysis results show that most varieties have a positive skewness in price distribution and certain fat-tail characteristics (e.g., SF has a kurtosis of 6.98), reflecting the potential risk of price jumps or extreme fluctuations in commodity futures. At the same time, skewness reflects the difference between the median and the mean, and the extreme value interval provides an important reference basis for risk control.

To intuitively present the structural trends, trading activity, and position changes of each variety of futures, we draw financial data visualization charts for the seven varieties. Each chart contains three subgraphs, representing the historical price change trend, trading volume distribution, and open interest evolution process of the variety in turn. The time span in the charts covers from 2017 to 2024, basically including the main operation cycles of the futures market in recent years [6] [7].

DOI: 10.54254/2755-2721/2026.30497

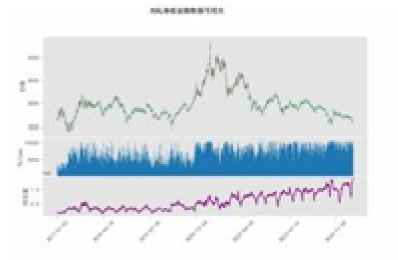

Figure 1. Hot-Rolled Coil

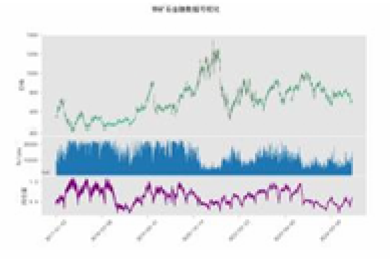

Figure 2. Iron Ore

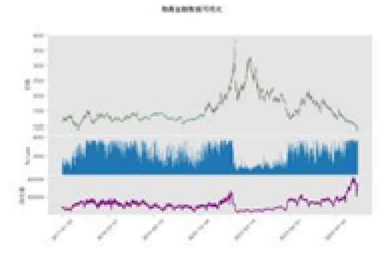

Figure 3. Coking Coal

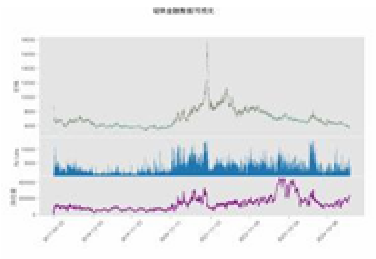

Figure 4. Ferrosilicon

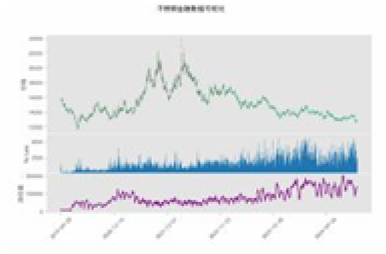

Figure 5. Stainless steel

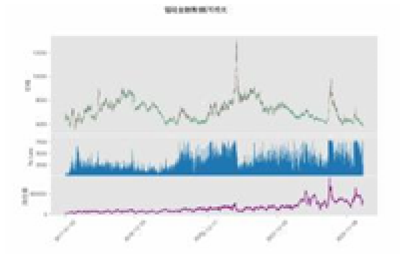

Figure 6. Silicomanganese

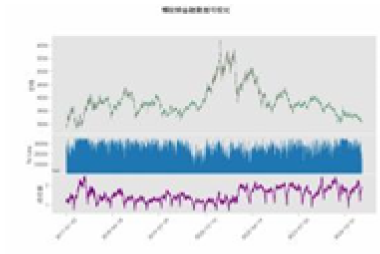

Figure 7. Rebar

The overall results of data visualization show that different varieties have significant differences in structural volatility. Steel varieties represented by Rebar (RB) and Hot-Rolled Coil (HC) show obvious cyclical fluctuation structures, and the changes in trading volume and open interest are relatively synchronized, indicating active trading and stable driving factors; while alloy varieties such as Ferrosilicon (SF) and Silicomanganese (SM) have price highs, but their trading volume and fluctuations are more intense in most periods, reflecting market uncertainty; Stainless Steel (SS) shows an approximate trend but discrete phenomenon in price and trading structure, reflecting the phased impact of macroeconomic fluctuations on the demand for this variety in recent years [8,9].

Some varieties (such as Coking Coal (JM) and Iron Ore (I)) have experienced sharp rises and falls within the price range, especially the extremely large fluctuation range between 2021 and 2022, which puts forward high requirements for the model's ability to handle non-stationarity. The graphical characteristics of trading volume and open interest also suggest some implicit features, such as the agglomeration of trading activity around price peaks for some varieties, which provides a verifiable pattern basis for the model's feature extraction mechanism [10].

To further explore the correlation structure of yields among various varieties, this paper constructs an annular cluster heatmap as shown in Figure 8. The darker the color in the figure, the stronger the correlation between different varieties. Hierarchical clustering is used to organize the correlation structure between samples to improve the overall visual effect.

DOI: 10.54254/2755-2721/2026.30497

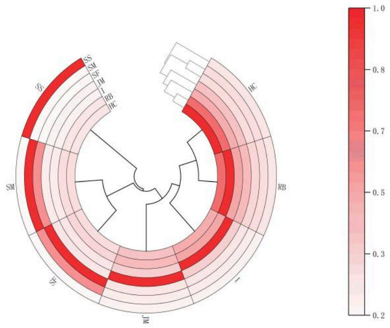

Figure 8. Cluster heatmap of yield correlation of each variety

It can be observed that some varieties have significant clustering characteristics, such as SS and SM, RB and HC, SF and JM, etc., which have a high degree of synchronization in yield structure, showing obvious industrial chain linkage or homogeneous trading behavior characteristics. This structural information provides strong support for subsequent model input design, risk control, and prediction error explanation.

3.2. Feature correlation analysis

To further understand the structural differences of each commodity futures variety and the linear relationship between variables, we draw sample feature correlation heatmaps of the seven varieties (HC, I, JM, RB, SF, SM, SS), as shown in Figures 9-15.

The heatmap is based on the Pearson correlation coefficient, with feature variables (including price indicators such as opening price, highest price, etc., trading indicators, and the predicted target variable 30-minute yield) on both the horizontal and vertical coordinates. The closer the color is to red, the stronger the positive correlation; the closer it is to blue, the stronger the negative correlation.

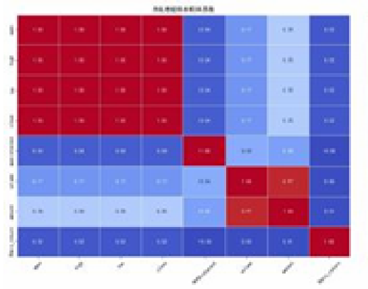

Figure 9. Hot-Rolled Coil

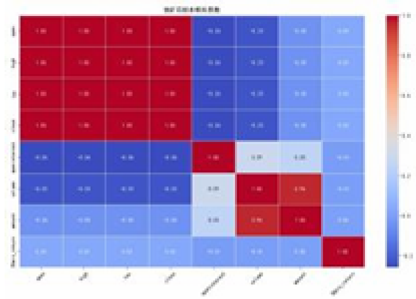

Figure 10. Iron Ore

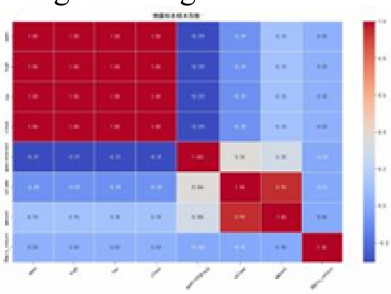

Figure 11. Coking Coal

DOI: 10.54254/2755-2721/2026.30497

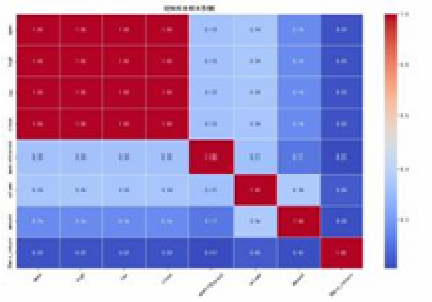

Figure 12. Ferrosilicon

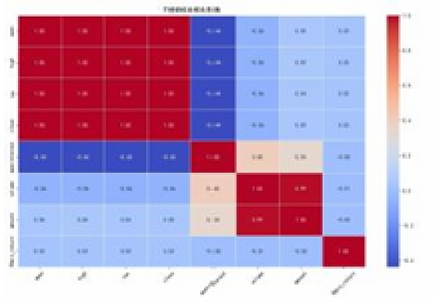

Figure 13. Stainless Steel

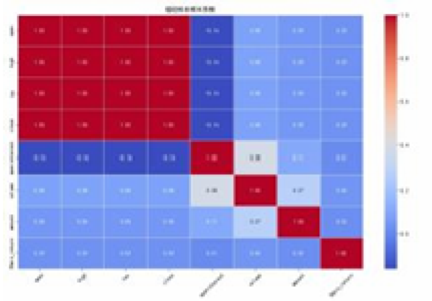

Figure 14. Silicomanganese

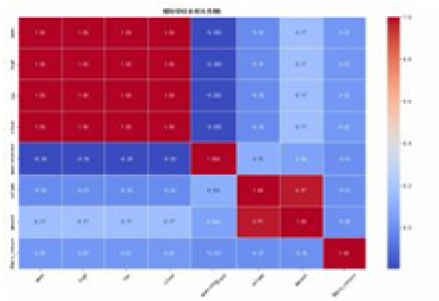

Figure 15. Rebar

It can be observed from the figures that different varieties have significant differences in variable correlation structures: for steel varieties such as HC and RB, price features such as open, high, low, and close are highly collinear (correlation coefficient close to 1), and trading features such as volume and amount are also highly correlated (greater than 0.97), but there is almost no significant correlation with the target variable 30min_return, indicating that their short-term price changes are mainly driven by nonlinear factors.

In comparison, energy and mineral varieties such as Iron Ore (I) and Coking Coal (JM) have a certain negative correlation trend between open_interest and price variables (e.g., the correlation is about −0.26), and the negative correlation between open_interest and price in Stainless Steel (SS) is higher (reaching −0.44), indicating that position changes may imply trend reversal signals in some varieties. The correlation structures of alloy varieties such as Ferrosilicon (SF) and Silicomanganese (SM) are relatively scattered, and the linear correlation between variables is generally low, indicating that their sequence structures are complex, and the model needs to rely on stronger nonlinear mapping capabilities.

In addition, the linear correlation between the target variable 30min_return and other input features is extremely low in all varieties (generally lower than 0.05), further verifying the limited applicability of linear regression methods in this task and emphasizing the necessity of deep learning methods.

3.3. Stationarity test

Since most machine learning models assume that the input data has weak stationarity, this paper uses the Augmented Dickey-Fuller (ADF) unit root test method to test the stationarity of the closing price sequence of each variety. The results are shown in Table 3:

DOI: 10.54254/2755-2721/2026.30497

Table 3. ADF unit root test results for each variety

Variety	ADF Statistic	p-value	1% Critical Value	5% Critical Value	10% Critical Value	Conclusion
HC	-62.672	0.000	-3.4304	-2.8615	-2.5668	Reject the null hypothesis, the sequence is weakly stationary
I	-62.338	0.000	-3.4304	-2.8615	-2.5668	Reject the null hypothesis, the sequence is weakly stationary
JM	-61.892	0.000	-3.4304	-2.8615	-2.5668	Reject the null hypothesis, the sequence is weakly stationary
RB	-62.697	0.000	-3.4304	-2.8615	-2.5668	Reject the null hypothesis, the sequence is weakly stationary
SF	-50.310	0.000	-3.4304	-2.8615	-2.5668	Reject the null hypothesis, the sequence is weakly stationary
SM	-60.379	0.000	-3.4304	-2.8615	-2.5668	Reject the null hypothesis, the sequence is weakly stationary
SS	-53.174	0.000	-3.4304	-2.8615	-2.5668	Reject the null hypothesis, the sequence is weakly stationary

The test results show that the yield sequences of all varieties reject the unit root null hypothesis and meet the weak stationarity condition, laying a data foundation for subsequent feature construction and modeling.

3.4. Nonlinear structure analysis

Commodity futures market prices are often driven by a variety of nonlinear factors, and it is difficult to fully describe their dynamic characteristics with linear models. Therefore, we use two commonly used indicators, BDS test and sample lag Hurst exponent, to detect nonlinear structures and long-term dependence. The results are summarized as follows:

Table 4. Nonlinear structure test results for each variety

Variety	BDS Test (p-value)	Sample Entropy	Hurst Exponent	Comprehensive Conclusion
HC	0.7687 (not detected)	1.824 (> 0.5)	0.493 (≈0.5)	Significant nonlinear characteristics exist, and long-term dependence is not obvious
I	0.5431 (not detected)	1.733 (> 0.5)	0.509 (> 0.5)	Significant nonlinear characteristics exist, with long-term dependence
JM	0.4531 (not detected)	1.671 (> 0.5)	0.519 (> 0.5)	Significant nonlinear characteristics exist, with long-term dependence
RB	0.5018 (not detected)	1.688 (> 0.5)	0.500 (≈0.5)	Significant nonlinear characteristics exist, and long-term dependence is not obvious
SF	0.6909 (not detected)	1.862 (> 0.5)	0.473 (< 0.5)	Significant nonlinear characteristics exist, and long-term dependence is not obvious
SM	0.5942 (not detected)	1.880 (> 0.5)	0.491 (≈0.5)	Significant nonlinear characteristics exist, and long-term dependence is not obvious
SS	0.4722 (not detected)	1.876 (> 0.5)	0.477 (< 0.5)	Significant nonlinear characteristics exist, and long-term dependence is not obvious

DOI: 10.54254/2755-2721/2026.30497

From three dimensions, all varieties have significant nonlinear characteristics. Although the BDS test does not significantly detect deviations from the self-distribution, the sample entropy is generally greater than 0.5, indicating that the data structure is complex and has obvious nonlinear characteristics. Some varieties (such as JM, I, SM, RB) also show obvious long-term dependence, suggesting that nonlinear long-memory models (such as LSTM, GRU, etc.) can be considered for modeling in the follow-up.

3.5. Principal component analysis and 1-minute-level feature dimensionality reduction strategy

To alleviate the multicollinearity problem among 1-minute-level feature variables, extract key influencing factors, and reduce dimensions, we conduct Principal Component Analysis (PCA) on the original feature data (opening price, highest price, lowest price, closing price, trading volume, turnover, open interest) of each commodity futures variety to construct modeling inputs with concentrated information and strong interpretability.

3.5.1. Data applicability test

Before principal component analysis, the KMO test and Bartlett's Test of Sphericity are used to judge whether the correlation basis between variables is established. The results show that the KMO values of the seven varieties are all greater than 0.7, and the significance p-values of Bartlett's test are all 0.000, indicating that there is a strong correlation between the original 1-minute-level feature variables, and the data is suitable for principal component extraction.

3.5.2. Principal component extraction and cumulative explained variance

According to the principle of eigenvalue greater than 1, the first two principal components are retained for each variety. Taking Hot-Rolled Coil (HC) as an example, the cumulative explained variance of its first two principal components reaches 76.043%, and those of Coking Coal (JM), Iron Ore (I), etc., also exceed 75%, indicating that most information can be represented by the first two principal components, and the dimensionality reduction effect is good.

Table 5. Cumulative explained variance of the first two principal components for each variety

Variety	Contribution Rate of Principal Component 1 (%)	Contribution Rate of Principal Component 2 (%)	Cumulative Explained Variance (%)
HC	53.283	22.760	76.043
I	52.451	24.440	76.891
JM	51.796	26.408	78.204
RB	51.088	24.667	75.755
SF	54.260	15.409	69.669
SM	50.553	19.098	69.651
SS	53.189	26.935	80.124

DOI: 10.54254/2755-2721/2026.30497

3.5.3. Principal component structure interpretation

Through component loading analysis, it can be seen that the loading of the first principal component on opening price, highest price, lowest price, and closing price is close to 0.98, which can be understood as the "price indicator factor"; the second principal component has a high loading on trading volume and turnover (generally greater than 0.90), which can be interpreted as the "market trading activity factor".

For example, the principal component structure of HC variety is as follows:

This structure is also significant in varieties such as Iron Ore (I), Coking Coal (JM), and Rebar (RB), indicating that the two main factors have good stability and explanatory power.

Table 6. Principal component loading matrix of HC variety

Variable	Principal Component 1 (Price Feature Factor)	Principal Component 2 (Trading Activity Factor)
Opening Price	0.978	-0.208
Highest Price	0.978	-0.207
Lowest Price	0.978	-0.209
Closing Price	0.978	-0.207
Open Interest	0.580	0.419
Trading Volume	0.538	0.810
Turnover	0.538	0.817
30-Minute Price Fluctuation Range	0.300	-0.0017

4. Model establishment and solution

4.1. Reasons for model selection and specific implementation of the model

Since the target to be predicted in this study is the commodity futures yield in the next 30 minutes, it is essentially a regression problem with significant temporal dependence and multivariate feature interaction. Under the high-frequency background with data volume exceeding 500,000 records, traditional linear modeling is difficult to capture nonlinear fluctuation rules, so we select a deep neural network combining convolution and recursive structures.

The CNN module is used to extract local patterns and macro trends, and the convolution kernel enhances the model's ability to identify short-term mutations and cyclical fluctuations through the sliding window mechanism;

The BiLSTM (Bidirectional Long Short-Term Memory) network takes into account the information modeling of the past and future, capturing the context-dependent characteristics of price changes;

The Attention mechanism dynamically assigns different weights to different time steps during the integration of sparse information, thereby improving the model's key attention ability in long sequences.

This structure performs well in capturing complex temporal nonlinear features and improving prediction robustness, and is particularly suitable for micro-modeling problems at the commodity futures level.

DOI: 10.54254/2755-2721/2026.30497

4.1.1. Basic model: LSTM network

LSTM (Long Short-Term Memory) is a classic temporal neural network structure, which effectively alleviates the gradient disappearance problem of ordinary RNN in long-sequence modeling by introducing memory cells and gating mechanisms.

In terms of model definition, let the input sequence be $X_t = \{x_{t-n}, \dots, x_t\}$, and the corresponding LSTM unit can be described by the following formulas:

$$f_t = \sigma(W_f\left[h_{t-1}, x_t\right] + b_f)$$

$$i_t = \sigma(W_f\left[h_{t-1}, x_t\right] + b_i)$$

$$\widetilde{C}_{\mathrm{t}} = tan_h(W_C\left[h_{t-1}, x_t\right] + b_C)$$

$$C_t = f_t \bigodot C_{t-1} + i_t \bigodot \widetilde{C}_{\mathrm{t}}$$

$$o_t = \sigma(W_o\left[h_{t-1}, x_t\right] + b_o)$$

$$h_t = o_t \bigodot tan_h\left(C_t\right)$$

where f_t, i_t, o_t represent the forget gate, input gate, and output gate respectively, and $\bigodot$ represents the Hadamard product. This structure enables the model to learn the complex interaction between long-term memory (through cell state $\mathrm{C_t}$) and short-term state (through hidden state ht), thereby adapting to complex time series patterns.

In the implementation process, we introduce a one-dimensional convolution (Conv1D) as the feature pre-mask to extract local temporal patterns, and then connect a bidirectional LSTM layer to improve the time perception dimension.

4.1.2. Enhanced model: CNN-LSTM-attention network

Although LSTM has good sequence memory ability, it is difficult to automatically focus on the most important time segments when processing long time windows. To further improve the model's ability to distinguish key moments, we introduce the Attention Mechanism after the LSTM structure to construct the CNN-LSTM-Attention enhanced model.

In this structure, the model calculates the importance weight α_i for the output hi (i.e., the hidden state of LSTM) at each time step, thereby generating a weighted context vector $\mathrm{C_t}$ to represent the summary of historical information:

DOI: 10.54254/2755-2721/2026.30497

$$e_i = tanh(W_h h_i + b_h)$$

$$\alpha_i = \frac{exp(e_i)}{\sum_j exp(e_j)}$$

$$C_t = \sum_i \alpha_i h_i$$

where W_h , b_h are learnable parameters of the attention layer, and α_i is the attention score of the i-th time step, reflecting the model's attention to this moment.

Through the context vector C_t , the model can significantly improve the ability to extract information from key feature jump segments (such as opening, high volatility periods, etc.), thereby enhancing prediction performance.

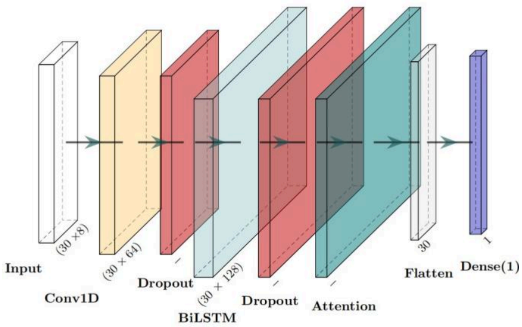

Figure 16. Structure diagram of the integrated deep neural network before optimization

4.1.3. Final model: RevIN-CNN-LSTM-attention network

Due to the huge differences in price levels and fluctuation scales of commodity futures among different varieties and trading periods, if the model is directly trained based on original values, it may face problems of distribution drift and normalization consistency, leading to a decline in generalization performance. Therefore, we introduce the Reversible Instance Normalization (RevIN) module into the model structure, and combine Min-Max Normalization to uniformly process the data of each variety, thus constructing the final RevIN-CNN-LSTM-Attention network.

In the input stage, RevIN normalizes each dimension feature of each sample instance in the following way:

$$\widehat{x}_t^{(i)} = \frac{x_t^{(i)} - \min\left(x^{(i)}\right)}{\max\left(x^{(i)}\right) - \min\left(x^{(i)}\right)} \tag{1}$$

where $x_t^{(i)}$ represents the original value of the i-th dimension at time t in the sample, and $\min\left(x^{(i)}\right)$, $\max\left(x^{(i)}\right)$ are the minimum and maximum values of this feature in the sample

DOI: 10.54254/2755-2721/2026.30497

respectively. This normalization process enables the model to train the main differences of different samples on different interval scales.

At the same time, the RevIN module also supports denormalization of the output results, that is, restoring the predicted values to the original coordinates according to the maximum and minimum values recorded in the input samples:

$$\hat{y}_t = \hat{y}_t^* \cdot \left(\max(y) - \min(y)\right) + \min(y) \tag{2}$$

This mechanism ensures that the model focuses on learning subsequence patterns during training, rather than generating errors affected by the numerical ranges of different varieties. In actual training, we further guide the feature preprocessing part to the multi-target factor and high-level feature extraction module, making RevIN a key bridge between the normalization strategy and the network structure.

To effectively model the sequential structure and multidimensional feature interaction information in high-frequency futures data, this paper constructs an integrated deep network model, whose structure is shown in Figure 17. The model first extracts stationary features within the local time window through one-dimensional convolution (Conv1D), and then models the forward and backward dependence of the time series through the bidirectional LSTM network.

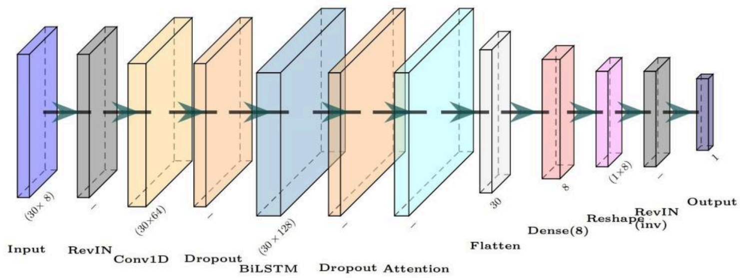

Figure 17. Structure diagram of the integrated deep neural network after optimization

The complete downstream connection structure of the final model is: input layer → RevIN normalization → convolution to extract local features (Conv1D) → bidirectional LSTM → attention mechanism → fully connected output layer → RevIN denormalization result recovery. This structure is the main modeling scheme finally selected in this study in terms of the expression and stability of high time-varying features, the modeling of temporal structures, and the generalization adaptability to different variety features.

4.2. Model training and verification process

After setting the model structure, to verify the robustness of its prediction effect on commodity futures price sequences, this study separately models and trains the samples of the seven varieties, and conducts a total of seven model experiments. Each training is based on the same process and parameter settings, adopts a unified data division strategy and standardization process, and uses unified evaluation indicators for comparison. The following takes the last training process (taking Rebar (RB) variety as an example) to introduce in detail the basic steps and implementation process of model training and verification.

DOI: 10.54254/2755-2721/2026.30497

4.2.1. Data division and sequence construction

The training data comes from the cleaned minute-level main contract dataset. The model input is a time series window of the recent 30 minutes, and the corresponding target is the yield (price fluctuation range) 30 minutes after the current moment. To ensure the temporal continuity of modeling and no future data leakage, each output sample is constructed by a sliding window method, and the dimension of its feature matrix is (BID_TIME_STEPS = 30), and the target variable is the last column in each sample, consistent with the predicted price fluctuation range.

In terms of dataset division, a fixed proportion division method is adopted: 90% is used as the training set, and the remaining 10% is the test set. To ensure the stability of the reference value of the training results, 10% of the data is also divided from the training set as the validation set through the setting of validation_split = 0.1, so as to obtain a better early stopping strategy and generalization ability.

4.2.2. Data normalization and denormalization processing

Due to the large differences in the price ranges of different varieties, to improve the efficiency and stability of model training, the input features are uniformly normalized before training. The normalization method adopts Min-Max Normalization, and the specific formula is as follows:

$$\mathrm{x}_{\mathrm{norm}}^{(\mathrm{i})}=\frac{\mathrm{x}^{(\mathrm{i})}-\min\left(\mathrm{x}_{\mathrm{train}}^{(\mathrm{i})}\right)}{\max\left(\mathrm{x}_{\mathrm{train}}^{(\mathrm{i})}\right)-\min\left(\mathrm{x}_{\mathrm{train}}^{(\mathrm{i})}\right)} \tag{3}$$

The normalization process ensures that the training set and test set are modeled under the same normalization interval, effectively improving the problem of numerical imbalance. When restoring the model's predicted value $\hat{\mathrm{y}}_{\mathrm{norm}}$ to the actual value, the following denormalization method is adopted:

$$\hat{y}=\hat{y}_{norm}\cdot\left(\max\left(y_{train}\right)-\min\left(y_{train}\right)\right)+\min\left(y_{train}\right) \tag{4}$$

The normalization and denormalization processes are encapsulated in the unified normalize_data and denormalize functions for implementation.

4.2.3. Model training configuration and process

The model training process is implemented using the Keras framework and runs based on the TensorFlow backend. The unified configuration in all trainings is as follows:

Loss function: Mean Squared Error (MSE), measuring the squared error between the predicted value and the actual observed value;

Optimizer: Adam, an adaptive first-order and second-order momentum estimation optimization algorithm, with the initial learning rate using the default value;

Evaluation indicators: coefficient of determination R^2 (r2_keras) and Mean Absolute Error (MAE);

Training epochs: the maximum number of epochs is 50 (epochs=50), and training is stopped if convergence is achieved in advance;

Batch size: 64 (batch_size=64);

DOI: 10.54254/2755-2721/2026.30497

Early stopping mechanism: set the EarlyStopping callback function to monitor the validation set loss val_loss, and stop training early if there is no improvement for 10 consecutive epochs (patience=10) to prevent overfitting;

Model saving: use the ModelCheckpoint callback to only save the model weights with the best performance on the validation set (path such as "JM.h5").

4.2.4. Model verification and result output

After the model training is completed, this paper predicts and evaluates the performance of the test set samples to verify the generalization ability and fitting level of the model on unseen data. The evaluation indicators cover seven commodity varieties, namely Hot-Rolled Coil (HC), Iron Ore (I), Coking Coal (JM), Rebar (RB), Ferrosilicon (SF), Silicomanganese (SM), and Stainless Steel (SS), and summarize the training and validation results and test performance of each variety.

Model performance evaluation includes coefficient of determination R^2, Root Mean Square Error (RMSE), and Mean Absolute Percentage Error. Among them, R^2 represents the model fitting degree, and RMSE reflects the average error between prediction and actual error. The combination of the two can evaluate the model's pros and cons from different angles.

The following table summarizes the main performance indicators of each variety on the test set:

Table 8. Summary of prediction performance metrics on test set for different commodity varieties ($R^2$and RMSE)

Variety	R²	RMSE
SM	0.7018	0.2029
RB	0.7346	0.1702
HC	0.7267	0.1619
SS	0.6337	0.1803
SF	0.6146	0.2400
JM	0.5774	0.2628
I	0.4628	0.2873

It can be seen from Table 8 that the R^2 of the model exceeds 0.6 in most varieties, indicating that it has good linear fitting ability. Among them, Silicomanganese (SM) and Rebar (RB) perform the best, with R^2 as high as more than 0.70, indicating that their price fluctuation patterns are relatively easy to be modeled; Rebar (RB) and Hot-Rolled Coil (HC) have good consistency.

From the perspective of error, RMSE is controlled within a reasonable range in most varieties, among which SM has the lowest RMSE (0.1323), showing the strongest prediction ability. This can be further optimized through the normalization mechanism.

4.3. Analysis of model prediction effect

This section aims to systematically analyze the prediction results of the optimized RevIN-CNN-LSTM-Attention model on the seven commodity varieties, and compare them with the pre-optimization model, so as to reveal the actual effects of each structural component (such as attention mechanism, RevIN normalization) and the combination potential of the model.

To improve the comprehensiveness and comparability of the evaluation, this section adopts three regression evaluation indicators: coefficient of determination R^2 , Root Mean Square Error

(RMSE), and Mean Absolute Percentage Error. Table 9 summarizes the prediction results of the pre-optimization and post-optimization models on the test sets of the seven varieties, which can clearly reflect the improvement of the prediction performance of each variety.

Table 9. Comparison of model performance metrics for each variety ($R^2$and RMSE before and after optimization)

Evaluation Indicator	Before Optimization		Before Optimization	
	R^2	RMSE	R^2	RMSE
HC	0.5893	0.1985	0.7267	0.1619
I	0.4289	0.3825	0.5614	0.2873
JM	0.5397	0.4628	0.5774	0.4289
RB	0.6311	0.1702	0.7346	0.1985
SF	0.3290	0.3167	0.6146	0.2400
SM	0.7057	0.2357	0.7818	0.2029
SS	0.5960	0.1519	0.6337	0.1323

It can be seen from Table 9 that after introducing RevIN normalization and attention mechanism, the coefficient of determination R^2 of the model has been significantly improved in all varieties, indicating that the model's ability to capture complex structural features has been enhanced. Among them, the variety with the most significant improvement in prediction performance is Ferrosilicon (SF), whose R^2 increased from 0.329 before optimization to 0.6146, and RMSE decreased to 0.240, while the volatility of prediction error changed little.

In addition, the RMSE indicator also shows a general downward trend. For example, the RMSE of Silicomanganese (SM) and Rebar (RB) decreased from 0.2357 and 0.1702 to 0.2029 and 0.1985 respectively, indicating that the overall error becomes smaller and the error distribution is more concentrated. On the one hand, this helps to improve the accuracy of subsequent risk assessment, and also reflects the effectiveness of the model structure optimization strategy. It is worth noting that although the R^2 and RMSE indicators have been significantly improved in most varieties, the model still has insufficient robust response to volatility in some extreme market conditions. This may be due to the enhanced response of the model to sharp price changes after structural optimization, and there may be a certain overfitting tendency when tracking maximum or minimum values, leading to a slight decline in prediction accuracy in some periods.

For example, although the R^2 indicator of SM variety has been greatly improved and the trend fitting effect has been enhanced, the error distribution in local extreme regions is still unbalanced, suggesting that the stability of the model in handling extreme cases needs to be further improved.

From the perspective of comprehensive indicator performance, the model has the best fitting effect on varieties such as Silicomanganese (SM), Hot-Rolled Coil (HC), and Rebar (RB), taking into account prediction error and generalization ability. Among them, SM has the highest R^2 (0.7818) among all varieties, indicating that there are obvious fittable rules in its historical sequence, which can improve the modeling efficiency and learning effect of the model; while RB has excellent performance in both R^2 and RMSE, showing high stability and practical value.

For commodities with high volatility such as Iron Ore (I) and Coking Coal (JM), although the increase in R^2 is limited, RMSE has decreased significantly, but it still shows a certain learning ability. This verifies that the RevIN module still has normalization advantages in adapting to different fluctuation structures.

DOI: 10.54254/2755-2721/2026.30497

To more comprehensively evaluate the actual fitting effect of the model on each futures variety, this paper visually compares the prediction results of the pre-optimization (original CNN-BiLSTM-Attention model) and post-optimization (structure with RevIN introduced) on the test set.

Pre-optimization:

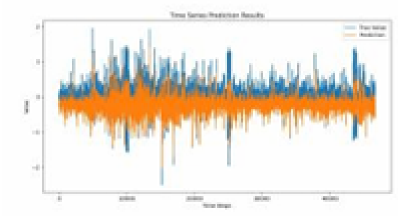

Figure 18. HC

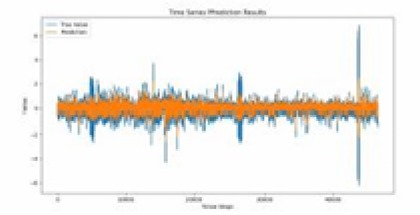

Figure 19. I

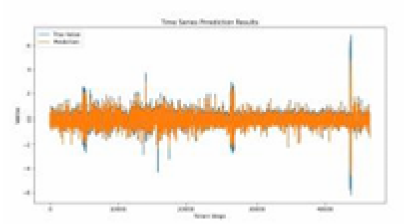

Figure 20. JM

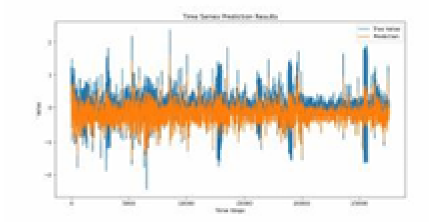

Figure 21. SF

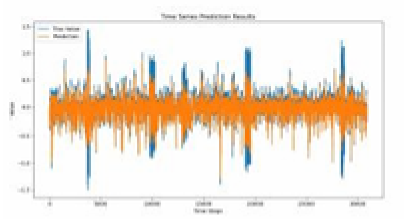

Figure 22. SS

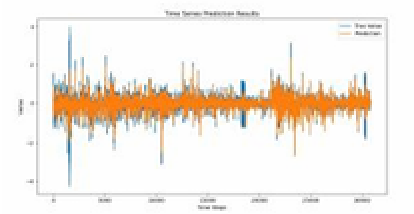

Figure 23. SM

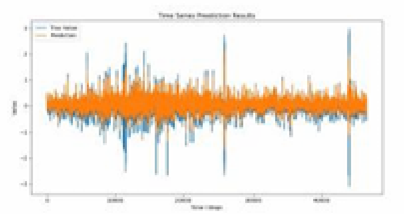

Figure 24. RB

Post-optimization:

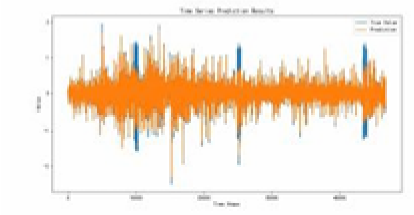

Figure 25. HC

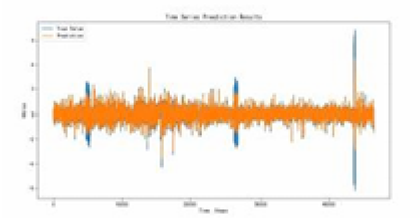

Figure 26. I

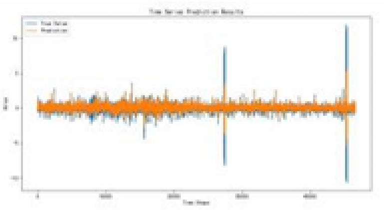

Figure 27. JM

DOI: 10.54254/2755-2721/2026.30497

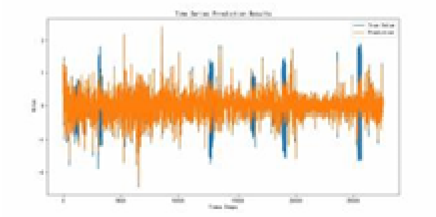

Figure 28. SF

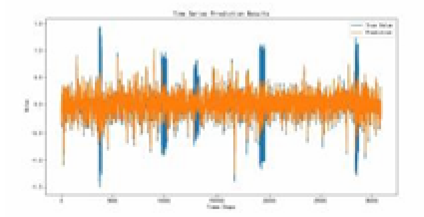

Figure 29. SS

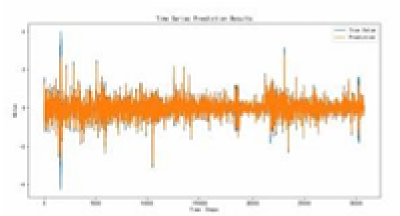

Figure 30. SM

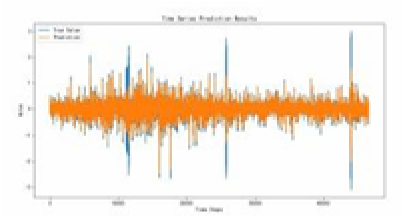

Figure 31. RB

In summary, this paper believes that through the introduction of attention mechanism and instance normalization mechanism in the model architecture, the model's ability to predict commodity price sequences has been significantly improved in multiple dimensions, especially in terms of real fitting and error convergence, which provides a basis for more optimized and refined modeling in the follow-up.

5. Conclusion

5.1. Attribution analysis of prediction differences among varieties

Although the overall prediction performance of the model is good, there are still significant differences in the prediction effects among different varieties. The fundamental reasons for such differences can be attributed to the fluctuation characteristics, noise level, and structural complexity of the data itself.

(1) Varieties with excellent fitting effects: SM, RB, HC

The common characteristics of these varieties are: there are relatively stable trend patterns or cyclical fluctuations in the sequence, and the model can effectively learn these structures through LSTM and Attention.

SM: The price sequence shows relatively stable periodic rising and falling characteristics in the historical stage, and its R^2 is as high as 0.7818, the highest among all varieties, indicating that the model has well captured its main characteristics.

RB: Significant seasonal and cyclical fluctuations, and industry data (such as inventory, demand) are strongly correlated with its price. The Attention module helps predict sharp nodes, showing excellent prediction effects (R^2 = 0.7346).

HC: Clear trend characteristics and relatively clustered main driving factors, so the model is easy to extract dominant signals, and the prediction performance is also high (R^2 = 0.7267).

(2) Variety with significant improvement after optimization: SF (Ferrosilicon)

Before optimization, the R^2 of SF was only 0.3290, the lowest, showing that the sequence stationarity was fuzzy and the basic model was difficult to learn effective rules. After the introduction of the RevIN module, the R^2 increased to 0.6146, indicating that the model played a

DOI: 10.54254/2755-2721/2026.30497

significant structural role in matching sequence distribution shifts, proving that the price structure adjustment is particularly effective for this problem.

(3) Varieties with high prediction difficulty: I, JM

I (Iron Ore), JM (Coking Coal): The prices of these two varieties are severely affected by external factors (such as policies, transportation, international demand, etc.), with irregular trends and significant impact of sudden fluctuations.

These varieties are often accompanied by low R^2 values (0.5614 and 0.5774 respectively) and high RMSE, indicating that their prediction upper limit is limited.

To further explore the source of differences in the model's prediction performance on different varieties, we use the SHapley Additive exPlanations (SHAP) value method to visually analyze the importance of input features of the final model. SHAP values can quantify the marginal contribution of each input variable to the model output, helping to understand the model decision-making process and the driving role of each feature.

The following figures show the SHAP value violin plots of the seven varieties (HC, I, JM, RB, SF, SM, SS), from which the importance ranking of each feature and its positive and negative impacts can be observed. For example, in all varieties, the return feature shows the highest importance in most cases, indicating that historical yield has the most reference value for predicting future price fluctuations; while the impact of variables such as volume and openinterest varies greatly among different varieties.

Overall, varieties with better fitting effects (such as SM, RB, HC) show a more concentrated and stable feature contribution pattern in the SHAP distribution, indicating that the model can clearly identify the main driving factors; while for varieties with high volatility such as I and JM, their SHAP distributions are relatively scattered, suggesting that the model's ability to capture their features is limited, further explaining the fundamental reasons for the differences in prediction performance.

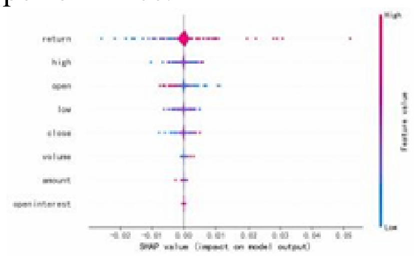

Figure 32. HC

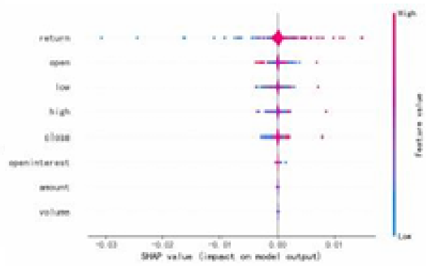

Figure 33. I

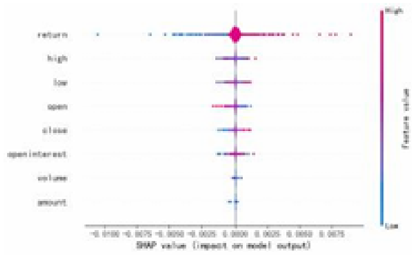

Figure 34. JM

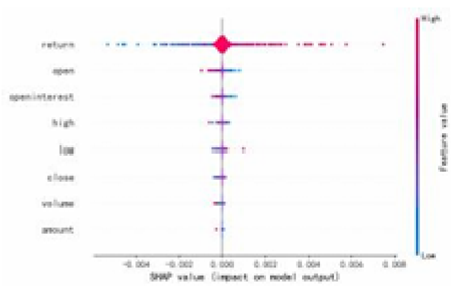

Figure 35. SF

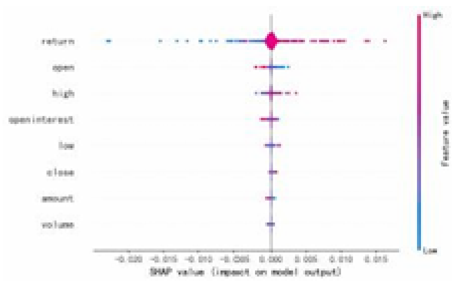

Figure 36. SS

DOI: 10.54254/2755-2721/2026.30497

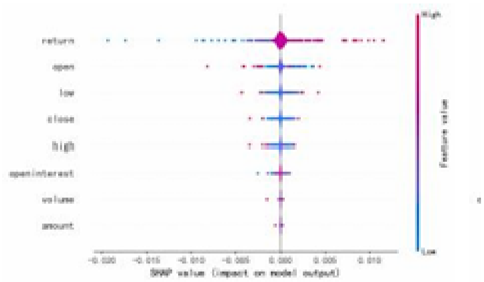

Figure 37. SM

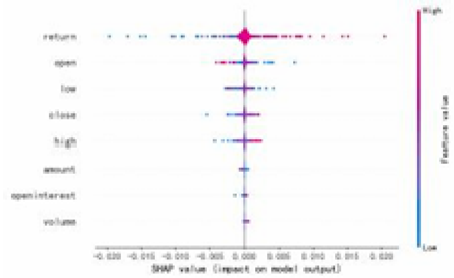

Figure 38. RB

5.2. Suggestions for subsequent optimization

In view of the above discovered applicable boundaries and limitations, this paper puts forward the following four dimensions of model optimization directions: model structure, feature engineering, training strategy, and model fusion. The specific suggestions are as follows:

(1) Model structure optimization:

Introduce Transformer structure mechanism and lightweight residual connection module;

It helps to strengthen the model's ability to model long-term dependency relationships and reduce the risk of the model in the fitting process of extreme points.

(2) Feature engineering expansion:

Integrate external macroeconomic data, periodic time factors, and sentiment indicators (such as news index, public opinion data);

It can further enhance the model's perception ability and explanatory ability for abnormal price fluctuations.

Future research can focus on the two core goals of "improving model generalization ability" and "relative error control ability", and further enhance the practicality and robustness of the model by combining the above optimization strategies.

References

[1] Kim, T., Kim, J., Tae, Y., Park, C., Choi, J.-H., & Choo, J. (2022). Reversible instance normalization for accurate time-series forecasting against distribution shift. In International Conference on Learning Representations (ICLR). Virtual.

[2] Tan, Y. (2021). Research on CNN-BiLSTM stock price prediction model and quantitative trading strategy based on attention mechanism [Doctoral dissertation]. Harbin Institute of Technology.

[3] Financial time series analysis. (2012). In Turing Mathematics & Statistics Series. Beijing: Posts & Telecom Press.

[4] Sun, R. (n.d.). Research on US stock index price trend prediction model based on LSTM neural network [Doctoral dissertation]. Capital University of Economics and Business.

[5] Chen, H. Y., & Ren, B. M. (n.d.). A multivariate time series prediction model based on time-frequency feature fusion. Journal of Huazhong University of Science and Technology (Natural Science Edition).

[6] Liu, Z. (2024). Research on financial performance evaluation of Company B based on factor analysis method [Doctoral dissertation]. Hebei University of Geosciences.

[7] Yang, Q., & Wang, C. (2019). Research on global stock index prediction based on deep learning LSTM neural network. Statistical Research, 36(3), 13.

[8] Nonlinear analysis and prediction of Chinese stock market. (2005). [Doctoral dissertation]. Systems Engineering.

[9] Zhou, Q. (2022). Research on the application of CNN and LSTM in short-term price rise and fall prediction of cyclical stocks [Master's thesis]. Zhejiang University.

[10] Cai, L. (2022). Research on lightweight of self-attention neural network model [Master's thesis]. South China University of Technology.