Dual-State Reinforcement-Learning-Based Dynamic Ensemble Model for Pork Price Forecasting

GUOBAO LIU[1],XUDONG LIN[1],ZHIGUO DU[1],BIN WEN[1] , ZHIHUI WU[1]

[1]College of Mathematics and Information, South China Agricultural University, Guangzhou 510642, China
E-MAIL: liuguobao@stu.scau.edu.cn,hunanlxd@163.com ,Bluetooth@scau.edu.cn,wenzip@scau.edu.cn,wuzhihui@scau.edu.cn

Abstract:

Reasonable forecasting of pork prices is of great significance for stabilizing price fluctuations in the pig market and promoting the healthy and sustainable development of the pig industry.The study addresses the high volatility and nonlinear characteristics of pork prices by proposing a composite forecasting framework that integrates a dynamic state transition mechanism with a dual random forest model to predict the monthly average pork prices. The accuracy of price forecasting is significantly enhanced through the introduction of adaptive feature engineering, a reinforcement learning-driven state transition mechanism, and a trend correction module. The core of the algorithm comprises three innovative mechanisms: (i) a volatility state detection system based on dynamic time warping, which accurately categorizes market states through the synergistic analysis of annualized volatility matrices and price momentum;(ii) a parallel architecture of dual random forest models, designed with differentiated parameters for stable and volatile periods;(iii) a reinforcement learning-driven adaptive weight adjustment mechanism that achieves online optimization of model weights by integrating state transition probability matrices and dynamic threshold control. Empirical research using pork price data from China from 2002 to 2024 demonstrates that the proposed hybrid model achieves an MSE of 0.4691, MAE of 0.4892, RMSE of 0.6849, and MAPE of 2.4056%. The research findings offer a novel methodological framework for agricultural product price forecasting and hold practical value for stabilizing the supply chain of the pork market.

Keywords:

Machine Learning, Dynamic State Transition, Multi-Model Ensemble Forecasting, Reinforcement Learning, Pork Price Forecasting.

1. Introduction

According to the data released by the National Bureau of Statistics, the national pig output in 2024 was 702.56 million heads, a decrease of 24.06 million heads from the previous year, down by 3.3%. Despite the decline in pig output, the figure remains above 700 million heads, demonstrating the overall stability of China's pig farming industry. Against this backdrop, this paper focuses on analyzing the trend of pork prices themselves. By constructing a forecasting model, this study delves into the intrinsic patterns of price fluctuations, providing a scientific basis for the stable development of the pork industry.

In the realm of agricultural product price forecasting, scholars have conducted extensive research with diverse methods. Initially, single-prediction models were prevalent. For instance, Martin-Rodriguez proposed a restrictive evolutionary spline model to analyze the daily seasonal price changes of canary tomatoes in the UK market [1]. With the development of research, due to the non-linear characteristics of time-series data, machine-learning models gained wide application. Shi Bo et al. utilized an improved RBF neural network model to accurately predict China's soybean prices [2]. Subsequently, combined models became a research focus. Zhou et al. combined the Transformer with the seasonal trend decomposition method, which significantly reduced the prediction errors of FEDformer in time-series forecasting [3]. Liu Heb ing et al. constructed a CEEMDAN-PCA-CNN-LSTM integrated model, effectively addressing the residual white-noise issue and enhancing prediction accuracy [4]. Regarding specific agricultural products, Lu Chaofan et al. employed the CCA-PCA-LSTM model for apple price forecasting, effectively reducing prediction errors [5]. Wu Zhan and Wang Chunxiao's ARIMA-LSTM model achieved good results in cottonseed meal price prediction [6]. Zeng Yurong et al. proposed the DE-TFT model for corn futures price prediction. They optimized TFT parameters and used other techniques to improve accuracy [7]. Wang Jie et al. enhanced the Informer model and proposed the STL-Informer-ARIMA model for pork price forecasting, improving prediction accuracy and reliability [8].

To address the aforementioned challenges, this study proposes a dual-state reinforcement learning dynamic ensemble model. The innovative system consists of three core modules: First, a dynamic state perception system is constructed to achieve accurate real-time identification of market conditions by integrating multi-dimensional market features with intelligent temporal alignment algorithms. Second, a differentiated modeling architecture is designed, utilizing parallel random forest models to capture long-term trends during stable periods and short-term abrupt changes during volatile periods, and introducing a temporal weighting mechanism to enhance the representation of recent data. Finally, an intelligent optimization engine is developed, integrating a multi-level linkage mechanism of trend adaptive compensation, dynamic volatility constraints, and residual feedback correction, forming a closed-loop intelligent decision-making system.

2. Data Sources and Preprocessing

2.1. Data Sources

The data was integrated from a hog-farming enterprise in southern China and the price monitoring platform of the China Hog Farming Network. The southern Chinese hog-farming enterprise provided data from January 2002 to December 2016, while the China Hog Farming Network offered provincial pork price data from December 2016 to December 2024. A weighted average of the pork prices collected from the China Hog Farming Network was calculated and then spliced and integrated with the enterprise's historical data to form the current price series, with the unit being CNY per kilogram.

2.2. Data Preprocessing

2.2.1 Dynamic Window Interpolation

To address the issue of random missing values in the price series, this method proposes a volatility-driven adaptive interpolation strategy [9]. By computing the annualized historical price volatility in real time, the interpolation window size is dynamically adjusted: the window is reduced during periods of high volatility to preserve short-term abrupt features, and enlarged during stable markets to capture long-term trends. A bidirectional linear interpolation algorithm is used, combining the weighted averages of adjacent data points before and after the missing value, to fill in the gaps while minimizing distortion to the original trend. This method effectively resolves the problem of excessive smoothing during periods of high volatility that is common with traditional fixed-window interpolation. The window size and interpolation formula are given in formula (1) and formula (2).

$$W_t = \max\left(5, \left\lfloor \frac{30}{1+10\sigma_t} \right\rfloor\right) \tag{1}$$

$$\widehat{p_t} = \frac{1}{W_t} \sum_{i=1}^{W_t} \left[\alpha p_{t-i} + (1-\alpha)\, p_{t+i}\right], \alpha \in [0.3, 0.7] \tag{2}$$

In the equation above,$P = \{p_1, p_2, \ldots, p_T\}$represents the original price series, and σ_t represents the dynamically calculated annualized volatility.

2.2.2 Improved Anomaly Detection

To identify outliers caused by transaction anomalies, a dynamic threshold detection mechanism based on the median absolute deviation is designed. By calculating the deviation of prices from the median within a sliding window, and adaptively adjusting the anomaly determination threshold according to volatility: relaxing the threshold tolerance when market volatility increases to avoid misjudging short-term sharp fluctuations as anomalies. The detected anomalies are replaced with a local sliding median, which eliminates noise interference while preserving the continuity characteristics of the price series. The algorithm for testing is as follows:

(i) Calculate the median price within the window:

$$\widetilde{P_{W_t}} = \text{median}\left(\widehat{P_{t-W_t+1:t}}\right) \tag{3}$$

(ii) Compute the Median Absolute Deviation (MAD):

$$\text{MAD} = \text{median}\left(\left|\widehat{p_t} - \widetilde{P_{W_t}}\right|\right) \tag{4}$$

(iii) Improved Z - score calculation:

$$z_t^* = \frac{0.6745\left(\widehat{p_t} - \widetilde{P_{W_t}}\right)}{\text{MAD} + \epsilon}, \quad \epsilon = 10^{-6} \tag{5}$$

(iv) Set dynamic threshold:

$$\tau_t = 3.5 \times \left(1 + \frac{\sigma_t}{0.15}\right) \tag{6}$$

(v) Correction strategy: When $|z_t^*| > \tau_t$, replace the outlier with a 7 - day sliding median:

$$p_t^{\text{corrected}} = \text{median}\left(\widehat{P_{t-3:t+3}}\right) \tag{7}$$

2.2.3 Data Normalization

To eliminate the dimensional differences between features and prevent the impact of outliers on the model, the data set range is scaled between 0 and 1[10]. Normalization is implemented as follows:

$$Z' = \frac{Z - Z_{\min}}{Z_{\max} - Z_{\min}} \tag{8}$$

where $Z_{\max}$ and $Z_{\min}$ represent the maximum and minimum values of the sample, Z is the original value, and Z' is the normalized value. After normalization, the numerical characteristics are still preserved on an equal scale, which can effectively improve the training effect and generalization performance of the model.

3 Research Methodology

3.1 Overall Process

A dynamic state adaptive method for predicting pork prices is proposed, the flowchart of which is shown in Figure 1. The specific process includes six core steps:

Step 1: Data Integration and Cleaning: Integrate pork price data, fill in missing values through temporal alignment and cubic spline interpolation, correct abnormal fluctuations, and construct a complete time series dataset.

Step 2: Feature Space Construction: Generate multi-scale trend baselines using sliding windows, combine dynamic volatility bands with momentum indicators to separate residual items, forming a three-dimensional feature matrix.

Step 3: Market State Identification: Determine market conditions based on dual threshold mechanisms of annualized volatility, and dynamically update the transition probabilities between stable and volatile periods through reinforcement learning state transition matrices.

Step 4: Hybrid Model Training: Build a dual-model random forest architecture: a shallow model for stable periods and a deep model for volatile periods, assigning 1.5 times the weight to the recent 30 days of data.

Step 5: Forecast Result Correction Implementation of Three-Level Optimization: RSI threshold triggers ±25% trend correction, dynamic volatility bands constrain the forecast range, and residual feedback integrates with entropy weight method to correct items.

Step 6: Iterative Optimization System: Establish a 30-day rolling window evaluation, update the state matrix parameters through reinforcement learning reward mechanisms, and optimize the feature engineering module with a learning rate of 0.1 in the reverse direction.

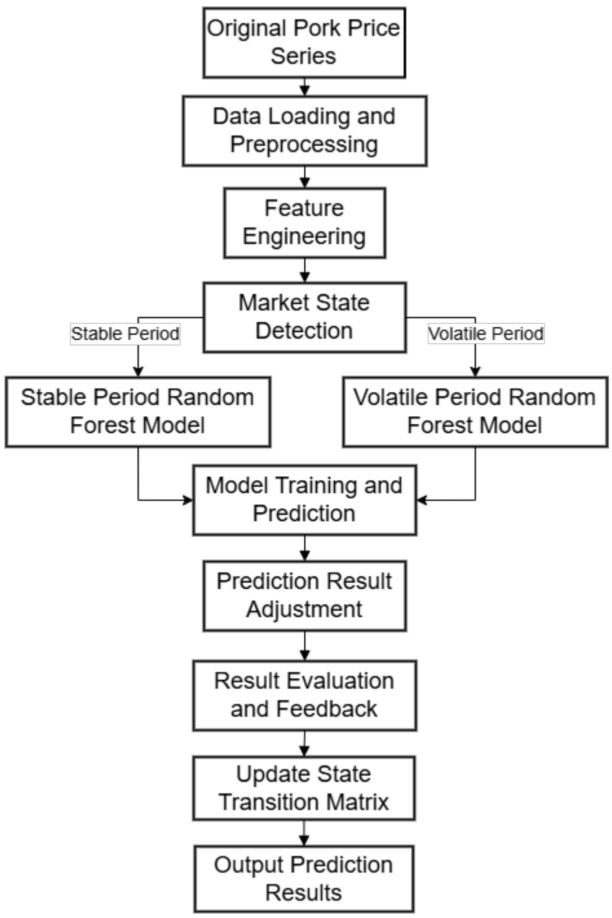

FIGURE 1. Pork Price Forecasting Process

3.2. Dynamic State Adaptive Forecasting Method

3.2.1 Multimodal Feature Space Construction

A three-variable representation system is constructed to fuse trend, cycle, and volatility features. The trend modality uses a sliding window to generate multi-scale baselines, and the window length is adjusted according to the real-time volatility. The cycle modality extracts the main frequency components of the monthly cycle through Fourier transform. The volatility modality detects anomalies based on the exponentially smoothed dynamic Bollinger Bands. A feature gating network is designed to generate dynamic weights through the hidden states of LSTM. During the stable period, more emphasis is placed on the trend modality (with a weight of 60% - 80%), while during the volatile period, the weights of the cycle and volatility modalities are increased (55% - 70%).

3.2.2 Dynamic Market State Identification

To dynamically identify market states, we define the market state space $S = \{s_0, s_1\}$, where s_0 represents a stable state and s_1 represents a volatile state. The identification rules for stable and volatile periods are given by the following formulas:

Stable Period Identification Rules:

$$\begin{cases} \sigma_{30}^{a} < \theta_t \\ |TS_t| \leq 0.15 \\ C_{MACD} \in (-0.5, 0.5) \end{cases} \tag{9}$$

Volatile Period Identification Rules:

$$\begin{cases} \sigma_{7}^{a} > 1.5\theta_t \\ |TS_t| > 0.25 \\ \Delta P_{3d} \geq Q_{90}(|\Delta P|) \end{cases} \tag{10}$$

θ_t is the dynamic threshold defined by equation (11):

$$\theta_t = \theta_0 \cdot \left[1 + \alpha \cdot \tanh\left(\frac{P_t - \mu_P}{3\sigma_P}\right)\right] \tag{11}$$

TS_t is the trend strength index defined by equation (12):

$$TS_t = \frac{1}{3}\sum_{m=1}^{3}\left[\frac{P_t - P_{t-5m}}{5m} \times \operatorname{sign}(ACC_{3m})\right] \tag{12}$$

σ_{30}^{a} represents the 30-day annualized volatility,σ_{7}^{a} represents the 7-day annualized volatility, ΔP_{3d} is the 3-day price absolute change, and Q_{90} is the 90th percentile of historical price changes.

Additionally, $\alpha = 0.3$ is the price sensitivity coefficient, μ_P is the annual price median, and σ_P is the annual price standard deviation. $m \in \{1, 2, 3\}$ represents the time window multiplier (corresponding to 5-day, 10-day, 15-day periods), and $ACC_{3m} = \frac{1}{3}\sum_{k=1}^{3}(P_{t-k} - P_{t-k-1})$ is the 3-period price acceleration, used to measure the rate of trend change.

3.2.3 Dual-Model Random Forest Architecture

Random forests can interpret the impact of several independent variables $(X_1, X_2, \ldots, X_k)$ on the dependent variable Y. If the dependent variable Y has n observations and is associated with k independent variables, when constructing classification trees, the random forest randomly selects n observations from the original data, with some being chosen multiple times and others not at all. This is the method of Bootstrap resampling. Additionally, the random forest randomly selects a subset of variables from the k independent variables for determining the nodes of the classification tree. Each constructed classification tree may differ. Generally, the random forest randomly generates hundreds to thousands of classification trees, and then selects the most consistent tree as the final result[11], as shown in Figure 2.

Constructing a differentiated forecasting model ensemble: 1. Stable period model: Configure 180 decision trees with a maximum depth of 5 layers, a minimum number of samples per leaf of 2, and the number of split features set to $\sqrt{p}$ (where p is the total number of features). 2. Volatile period model: Enhance to 200 decision trees, extend the depth to 6 layers, and reduce the minimum number of split samples to 3. The dynamic weight allocation mechanism assigns 1.5 times the sample weight to the recent 30 days of data, and feature importance is calculated using both Gini impurity and permutation importance for dual verification.

3.2.4 Dynamic Optimization of Forecast Results

Establish a three-level correction system:

①. Trend Compensation: Based on the 14-day RSI indicator, trigger a ±25% amplitude correction, with the formula as follows:

$$\Delta P_{\text{RSI}} = \begin{cases} -0.25 \times \frac{RSI-70}{30} \times P_{\text{base}}, & \text{if } RSI > 70 \\ 0.25 \times \frac{30-RSI}{30} \times P_{\text{base}}, & \text{if } RSI < 30 \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

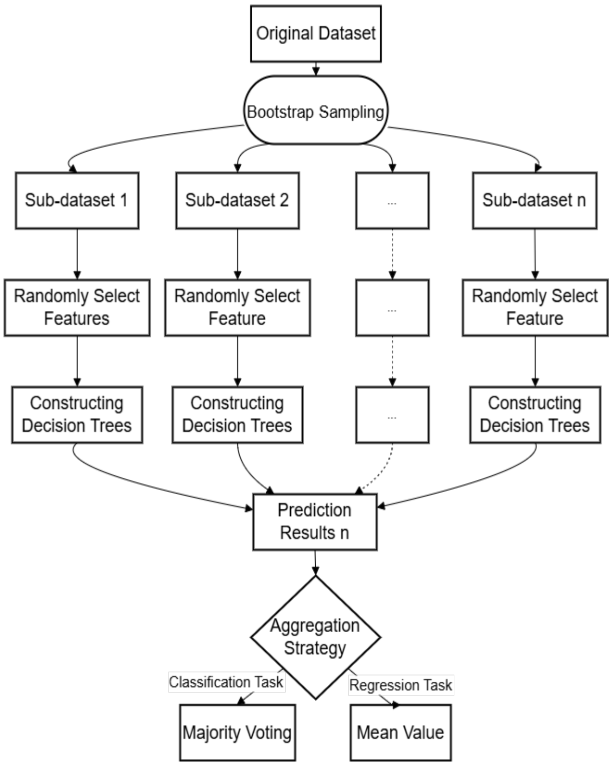

FIGURE 2. Random Forest Diagram

Where RSI is the 14-day relative strength indicator, P_{base} is the basic forecast value. This algorithm implements a gradual compensation for overbought and oversold states.

②.Volatility Constraint: Dynamic volatility band limits the forecast deviation range, with the maximum allowable deviation rate set to 15% of the base price, with the formula as follows:

$$B_t = 0.15 \times \left(1 + 0.5 \times \tanh\left(\frac{\sigma_t}{\theta_0} - 1\right)\right) \tag{14}$$

Where σ_t is the 30-day annualized volatility, $\theta_0 = 0.15$ is the base threshold.

③.Residual Feedback: Using an error propagation model with a decay coefficient of 0.3, the historical error window retains 6 periods. The final forecast value is integrated through entropy weight method, with the model output weight accounting for 60%, the current state information accounting for 25%, and the trend correction accounting for 15%. The formula is as follows:

$$\varepsilon_t = \sum_{k=0}^{5} (0.3)^k \cdot \varepsilon_{t-k-1} \tag{15}$$

$$W = \frac{1}{1 + \sum_{i=1}^{n} H_i} \begin{bmatrix} 0.6 \\ 0.25 \\ 0.15 \end{bmatrix} \tag{16}$$

Where ε_t is the historical error, H_i is the information entropy of each forecast source.

3.2.5 Reinforcement Learning-Driven Dynamic Optimization Mechanism

A reinforcement learning framework based on Q-Learning is introduced to construct a closed-loop optimization system of "state-action-reward." A two-dimensional state space (market state encoding, trend strength index) and a third-order action space (model weight adjustment, state transition threshold correction, exploration rate dynamic decay) are designed. A multi-dimensional reward function integrates prediction accuracy, state stability, and weight diversity indicators. A dual-channel deep Q-network is used to evaluate action values, where the target network maintains parameter stability through a soft update strategy ($\tau = 0.01$) . The training process uses a cyclic experience replay mechanism to store thousands of state transition records, combined with an adaptive ϵ-greedy strategy (initial exploration rate 0.3, exponential decay factor $\lambda = 0.001$) to balance exploration and exploitation. The Huber loss function is innovatively introduced to enhance network training stability, and a triple parameter linkage mechanism is achieved: dynamically adjusting the model weight allocation during stable/volatile periods (±10% amplitude) optimizing the annualized volatility threshold (±5% sensitivity) autonomously adjusting the state exploration decay rate (0.95-1.05 multiplier). The reward function is shown as follows:

$$R_t = 0.6 \left(1 - \frac{\left|P_t - \widehat{P}_t\right|}{P_t}\right) + 0.3e^{-2|\Delta S|} + \frac{0.1}{1 + \mathrm{Var}\,(W)} \tag{17}$$

where $\Delta S = \left\|s^{(t)} - s^{(t-1)}\right\|_2$ is the state transition amplitude, and $\mathrm{Var}\,(W)$ measures the model weight fluctuation.

3.3. Model Evaluation Metrics

This experiment employs five standards to assess the model's performance and forecasting accuracy: Mean

Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE). The formulas are as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2 \tag{18}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\widehat{y}_i - y_i| \tag{19}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2} \tag{20}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\widehat{y}_i - y_i}{y_i}\right| \tag{21}$$

where y_i is the actual value, $\widehat{y}_i$ is the predicted value, and n is the number of forecasted data points.

4. Experimental Results and Analysis

4.1. Model Evaluation Metrics

LSTM[12], GRU [13], SVR[14], and traditional Random Forest (RF) models[15] are selected, all of which have unique advantages and application scenarios in the field of time series. These models are compared with the forecasting model proposed in this paper to verify the predictive performance of the multi-model combination optimization method.We divided the dataset into training and testing sets in a 7:3 ratio. The experimental results show that our method (OURS) significantly outperforms the baseline models in four core metrics: MSE (0.4691) is 79.3% lower than LSTM (2.2763), MAE (0.4892) is 58.2% lower than GRU (0.8528), RMSE (0.6849) is only 40.4% of SVR (1.6945), and MAPE (2.4056%) is a 65.8% decrease compared to RF (7.046%).

TABLE 1. Comparative experiment of different models of monthly average pork price

Model	MSE	MAE	RMSE	MAPE
OURS	0.4691	0.4892	0.6849	2.4056%
LSTM	2.2763	1.1693	1.5087	8.086%
GRU	1.3468	0.8528	1.1605	5.603%
SVR	2.8712	1.0676	1.6945	6.649%
RF	2.5193	1.1342	1.7533	7.046%

4.2. Ablation Experiment Analysis

To verify the effectiveness of the model's core components, this study designed three sets of control experiments: (i) removing the dual model mechanism (only retaining the stable period model), (ii) disabling reinforcement learning optimization (fixing the state transition matrix), (iii) eliminating the trend adaptation module. As shown in Table 2, the complete model (MSE=0.4691, MAPE=2.41%) significantly outperforms all simplified versions. Among them, the dual model mechanism contributes the most, and its removal leads to an increase of 42.3% in MAE to 0.696; the absence of the reinforcement learning module increases the RMSE standard deviation by 2.03 times; after removing the trend correction module, the MAPE at price inflection points rises to 5.14%. The experiments prove that there is a significant synergistic effect among dynamic state perception, differentiated modeling, and trend compensation mechanisms.

5. Conclusions

This research developed a Dual-State Reinforcement Learning based Dynamic Ensemble Model for pork price forecasting, addressing their high volatility and non - linearity.

The dual discrimination mechanism, using annualized volatility matrices and trend strength indices, is crucial for differentiating market states. Ablation experiments show its removal increases prediction errors.

The differentiated dual-model random forest architecture, with 180-tree shallow and 200-tree deep models for stable and volatile periods respectively, outperforms single models like LSTM, GRU, SVR, and RF. For example, in MAPE, our model achieves 2.4056%, which is 65.8% lower than that of RF

The reinforcement learning-driven three-level correction system, combining trend compensation, volatility band constraints, and residual feedback, stabilizes forecasts. Its removal in ablation experiments notably increases RMSE standard deviation.

Practically, the model can offer 1-3-month price warning windows for pig farming enterprises, aiding decision-making in the pork market supply chain.

Future work will integrate feed price data to build a multi-source risk coupling early warning system and explore the model's transferability to poultry market price prediction. Our research provides a novel framework for agricultural product price forecasting.

TABLE 2. Comparison results of model ablation experiment

Dual Model Mechanism	RL State Transition	Trend Adaptation	MSE	MAE	RMSE	MAPE
—	✓	✓	1.735	1.105	1.598	6.1220%
✓	—	✓	0.981	0.696	1.087	4.7356%
✓	✓	—	1.267	0.785	1.125	5.1356%
✓	✓	✓	0.4691	0.4892	0.6849	2.4056%

References

[1] Martin-Rodriguez, G., Caceres-Hernandez, J. J., "Canary tomato export prices: comparison and relationships between daily seasonal patterns", Spanish journal of agricultural research, Vol 11, No. 4, pp. 882 - 893, 2013.

[2] Shi, B., Zhang, D. Q., Ma, K. P., et al., "Research on the Application of Improved RBF Neural Network in the Prediction of Soybean Prices in China", Soybean Science, Vol 35, No. 2, pp. 310 - 314, 2016.

[3] Zhou, T., Ma, Z., Wen, Q., "Fedformer: Frequency enhanced decomposed transformer for long - term series forecasting", International conference on machine learning, PMLR, pp. 27268 - 27286, 2022.

[4] Liu, H. B., Hua, M. D., Xi, L., et al., "Construction of a Prediction Model for White - Striped Pig Prices", Journal of Henan Agricultural University, Vol 58, No. 1, pp. 123 - 131, 2024.

[5] Lu, C. F., Shi, S. K., Wang, L., "Research on the Apple Price Prediction Model Based on Improved LSTM", Journal of Shandong Agricultural University (Natural Science Edition), Vol 53, No. 3, pp. 491 - 496, 2022.

[6] Wu, Z., Wang, C. X., "Application of ARIMA - LSTM model based on ensemble learning in cottonseed meal price prediction", Feed Research, Vol 2025, No. 2, pp. 227 - 231, 2025.

[7] Zeng, Y. R., Wu, B. R., Wang, L., et al., "Interpretable Corn Futures Price Forecasting with Multivariate Heterogeneous Data", Management Review, Vol 35, No. 12, pp. 41 - 52, 2023.

[8] Wang, J., Dong, G. A., Li, J. Q., "Pork Price Prediction Method Based on STL - Informer - ARIMA Combined Model", Journal of Shandong Agricultural University (Natural Science Edition), Vol 55, No. 3, pp. 367 - 375, 2024.

[9] Capobianco, J., Pallone, G., and Daudet, L., "Dynamic strategy for window splitting, parameters estimation and interpolation in spatial parametric audio coders", Proceeding of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, pp. 397 - 400, November 2012.

[10] Sinsomboonthong, Saichon, "Performance Comparison of New Adjusted Min - Max with Decimal Scaling and Statistical Column Normalization Methods for Artificial Neural Network Classification", International Journal of Mathematics and Mathematical Sciences, Vol 2022, No. 1, pp. 3584406, 2022.

[11] Fang Kuangnan, Wu Jianbin, Zhu Jianping, et al., "A Review of Random Forest Methods", Statistics & Information Forum, Vol. 26, No. 3, pp. 32 - 38, 2011.

[12] Yadav, H., & Thakkar, A., "NOA-LSTM: An efficient LSTM cell architecture for time series forecasting", Expert Systems with Applications, Vol 238, pp. 122333, 2024.

[13] Tu, B., Bai, K., Zhan, C., et al., "Real - time prediction of ROP based on GRU - Informer", Scientific reports, Vol 14, No. 1, pp. 2133, 2024.

[14] Ampountolas, A., "Enhancing forecasting accuracy in commodity and financial markets: Insights from garch and svr models", International Journal of Financial Studies, Vol 12, No. 3, pp. 59, 2024.

[15] Goulet Coulombe, P., "The macroeconomy as a random forest", Journal of Applied Econometrics, Vol 39, No. 3, pp. 401 - 421, 2024.