

Article

Short-Term Agricultural Commodity Price Forecasting: A Metaheuristic-Optimized LSTM–Attention Framework

Chang Su, Yuanping Zhang, Xiang Li, Songsong Hou, Yukai Chang and Yan Guo *

College of Information Engineering, Sichuan Agricultural University, Ya'an 625000, China; 2024219008@stu.sicau.edu.cn (C.S.); 18723357532@163.com (Y.Z.); 13550148182@163.com (X.L.); hss@stu.sicau.edu.cn (S.H.); m15670757378@163.com (Y.C.)

* Correspondence: 14403@sicau.edu.cn

Abstract

Accurate short-term agricultural commodity price forecasting is important for market monitoring, procurement planning, inventory management, and price-risk awareness. This study develops a data-driven forecasting framework, namely CMACS–ARO–LSTM–Attention (CALA), which combines an LSTM–Attention predictor with an improved Artificial Rabbits Optimization algorithm. The original reported CALA results are retained unchanged. This minimal revision adds persistence/random-walk and ARIMA references on the archived comparison blocks. For wheat, the reported CALA result is $R^2 = 0.995$, RMSE = 0.008, MAE = 0.004, and MAPE = 0.021%; the newly added persistence/random-walk and ARIMA references have RMSE values of 0.102 and 0.092, respectively. The soybean application is presented as supplementary mixed evidence rather than proof of broad generalization. The high goodness of fit is interpreted as short-horizon price-level approximation in a persistent series, not as proof of directional trading profitability. Repeated CALA runs, formal predictive-accuracy tests, equal-budget ablations, and multi-step interval forecasts remain outside this minimal revision.

Keywords: agricultural commodity price forecasting; agricultural market risk; LSTM–Attention; metaheuristic optimization; capacity calibration; wheat futures; soybean prices

1. Introduction

Agricultural commodity prices directly affect farm income, planting decisions, procurement costs, inventory management, public reserve arrangements, and the stability of food supply chains [1,2]. In grain markets, short-term price fluctuations are not only financial signals for traders, but also operational signals for agribusiness firms, processors, storage enterprises, and policymakers [3,4]. Reliable next-day or short-horizon forecasts can therefore support market monitoring, procurement scheduling, inventory adjustment, and early warning of abnormal price movements [5]. With increasing market integration, domestic agricultural prices are influenced by regional spot-market conditions, related futures contracts [3], and cross-market price transmission and spillover effects [4]. These factors create nonlinear temporal dependence and cross-market interactions that are difficult to capture with a single fixed-form model. Recent agricultural commodity forecasting studies have therefore increasingly adopted attention-based recurrent models, decomposition–LSTM hybrids, and image- or alternative-data-driven deep-learning approaches for agricultural and wheat futures price prediction [6–8].



Academic Editor: Giuseppe Timpanaro

Received: 17 June 2026

Revised: 29 July 2026

Accepted: 30 July 2026

Published: 3 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Existing forecasting research has advanced along three broad lines. The first emphasizes statistical structure and relies on econometric models [9,10] or decomposition-based models [11] to characterize trend, seasonality, and volatility dynamics; broader forecasting comparisons further show that unbiased benchmark design remains essential when comparing statistical and machine-learning methods [12,13]. The second centers on nonlinear function approximation through machine-learning models, including random forest [14] and support vector regression [15,16], as well as deep-learning models such as LSTM [17,18], GRU [19], and other recurrent or hybrid networks for agricultural and commodity price forecasting [20–23]. Recent surveys also show that recurrent, attention-based, and other deep-learning architectures have become important tools for one-step and multi-horizon time-series forecasting, including financial and commodity price applications [24–28]. The third introduces metaheuristic optimization, including differential-evolution-based LSTM configuration [29], PSO-based neural-network optimization [30], and PSO–CS forecast combination [2], to improve model configuration and mitigate the instability associated with manual hyperparameter selection. Taken together, these studies suggest that forecast gains may arise from structural representation, expressive nonlinear learning, or more effective search over model configurations, although the relative contribution of each channel is rarely identified cleanly.

Beyond one-step-ahead point forecasting, related studies also suggest that forecast-horizon extension and uncertainty characterization are important directions for time-series forecasting. Multi-step-ahead forecasting requires dedicated strategies such as direct, recursive, or multiple-output prediction [31], while probabilistic forecasting provides a formal framework for characterizing predictive uncertainty beyond single-point estimates [32]. In agricultural commodity and food-supply-chain contexts, recent news-event-driven forecasting research further shows that textual event signals extracted from news articles can improve commodity price forecasting when historical prices alone are insufficient [33].

Three issues remain important for agricultural commodity price forecasting. First, many studies emphasize forecast accuracy but pay insufficient attention to whether the forecasting framework is useful for agricultural market monitoring and operational decisions. Second, optimization-enhanced deep-learning models are often evaluated within a model-contest framework, so reported gains should be interpreted with respect to the benchmark set and experimental protocol actually used. Third, extremely high goodness-of-fit values in persistent price-level series may overstate practical usefulness if directional performance, robustness across commodities, and leakage-safe evaluation protocols are not examined carefully [34].

To address these issues, this study proposes a CALA framework for short-term agricultural commodity price forecasting. The inner layer learns a nonlinear conditional mapping from recent multi-source market observations to future prices through relevance-weighted sequence modeling, whereas the outer layer performs capacity calibration by searching over the predictor's hyperparameter configuration. Specifically, we enhance the ARO algorithm with Cauchy mutation and adaptive crossover to construct the CMACS–ARO optimizer, and then integrate this optimizer with an LSTM–Attention predictor so that key hyperparameters are selected through validation-driven search rather than ad hoc manual tuning.

The study makes three main contributions. First, it develops an integrated forecasting framework that combines relevance-weighted sequence modeling with validation-based capacity calibration, allowing the model to use recent spot–futures information more effectively while reducing dependence on manual hyperparameter selection. Second, it evaluates the framework on Chinese wheat futures data using regional spot prices and related futures variables, thereby linking the forecasting task to a practical agricultural market set-

ting. Third, it assesses comparative performance against a selected set of machine-learning, recurrent, attention-based, and metaheuristic-optimized baselines and further examines the same modeling pipeline using soybean price data. The results are interpreted conservatively as evidence of improved short-horizon price-level forecasting for agricultural market monitoring rather than as proof of universal superiority or trading profitability.

2. Materials and Methods

This section formalizes the proposed framework as an integrated forecasting system rather than a loose sequence of algorithmic steps. It begins by defining next-trading-day agricultural price forecasting as conditional-expectation approximation under a short information set, then specifies CALA as the outer-loop search mechanism, introduces the recurrent-attention predictor together with its hyperparameter space, and finally sets out the implementation routine and full workflow. Figure 1 provides an overview of these stages.

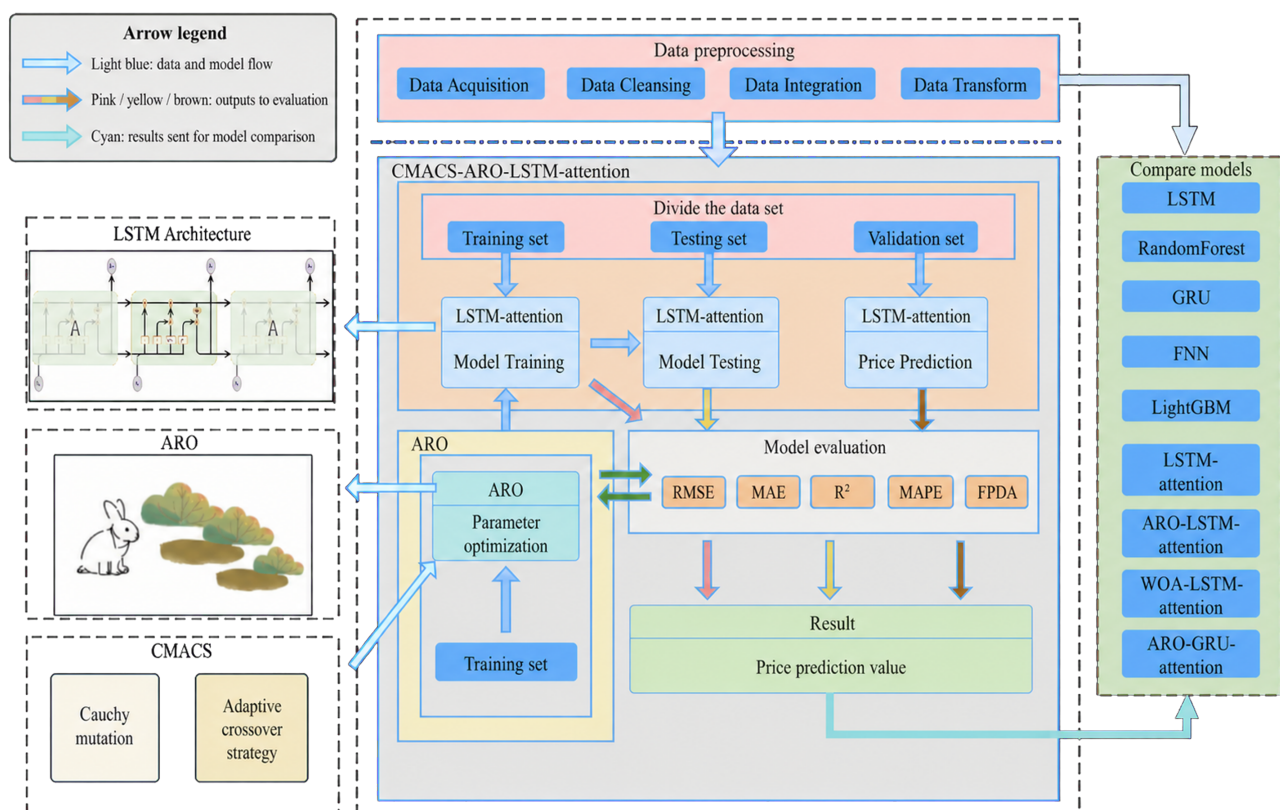


Figure 1. Overall architecture of the proposed CALA framework.

2.1. Problem Definition

Let $\{y_t\}_{t=1}^T$ denote the daily series of the target agricultural futures price, and let $\mathbf{x}_t \in \mathbb{R}^p$ denote the accompanying feature vector composed of regional spot prices and related futures prices. The task of interest is next-trading-day forecasting conditional on a short recent history rather than long-horizon structural prediction. Under a sliding-window scheme with length $k = 5$, the forecast of the price on day $t + 1$ is written as

$$\hat{y}_{t+1} = f_{\theta}(y_{t-k+1:t}, \mathbf{x}_{t-k+1:t}), \quad (1)$$

where $f_{\theta}(\cdot)$ denotes the nonlinear mapping constructed by the LSTM-Attention model and θ represents the trainable network parameters. The five-trading-day input window was fixed a priori to represent one trading week and to constrain model complexity rel-

ative to the available sample; the reported findings are conditional on this predefined look-back length.

From a first-principles perspective, the target of prediction can be viewed as the conditional expectation of the next-trading-day price given the available information set $\mathcal{F}_t$:

$$\hat{y}_{t+1} \approx \mathbb{E}(y_{t+1} \mid \mathcal{F}_t), \quad (2)$$

where $\mathcal{F}_t = \sigma\{y_{t-k+1:t}, \mathbf{x}_{t-k+1:t}\}$ summarizes the recent target-price history and the associated spot–futures market information. In this sense, the LSTM–Attention predictor approximates the conditional-expectation operator by compressing the information set into latent sequential states and then reweighting those states according to their marginal predictive relevance.

Hyperparameter selection is further recast as a bilevel optimization problem. Let λ denote the hyperparameter vector, including the number of hidden units, batch size, and training epochs. For a given λ , the network parameters are estimated through the inner problem

$$\theta^*(\lambda) = \arg \min_{\theta} \mathcal{L}_{\text{train}}(\theta; \lambda), \quad (3)$$

and the outer problem searches for the hyperparameter configuration with the strongest validation performance:

$$\lambda^* = \arg \min_{\lambda \in \Lambda} \mathcal{L}_{\text{val}}(\theta^*(\lambda)). \quad (4)$$

Here, Λ denotes the admissible hyperparameter search space and $\theta^*(\lambda)$ denotes the trained network-parameter vector obtained by minimizing training loss for a fixed configuration λ . The CMACS–ARO optimizer is used to obtain an approximate solution to this outer nonconvex search problem over a predefined search space. Once a best-performing admissible configuration is identified, the LSTM–Attention model is retrained accordingly, after which performance is reported for the archived comparison block and probed again on an additional agricultural commodity dataset.

2.2. Improved Artificial Rabbits Optimization Algorithm Based on CMACS

Artificial Rabbits Optimization [35] is a bio-inspired metaheuristic algorithm based on the survival behavior of wild rabbits, particularly detour foraging and random hiding strategies. In the detour-foraging phase, each search individual tends to update its position toward another randomly selected member of the population, while stochastic perturbations are introduced to improve search efficiency and avoid local optima. The corresponding update rule can be written as

$$\mathbf{v}_i(t+1) = \mathbf{x}_i(t) + R \cdot [\mathbf{x}_i(t) - \mathbf{x}_j(t)] + \text{round}[0.5(0.05 + r_1)] \cdot \mathbf{n}_1, \quad (5)$$

$$R = L \cdot c, \quad L = \left(e - e^{\left(\frac{t-1}{T}\right)^2} \right) \sin(2\pi r_2), \quad (6)$$

where $\mathbf{v}_i(t+1)$ is the candidate position of the i th rabbit at iteration $t+1$, $\mathbf{x}_i(t)$ is its current position, n is the population size, T is the maximum number of iterations, and r_1 and r_2 are random numbers in $(0, 1)$. Here, R is the detour-foraging step vector, L is the dimension-wise scaling vector, c is the scalar control coefficient, and $\mathbf{n}_1$ is a random perturbation vector.

To balance exploration and exploitation, ARO adopts an energy-decay mechanism. The energy factor is defined as

$$A(t) = 4 \left(1 - \frac{t}{T} \right) \ln \left(\frac{1}{r} \right), \quad r \sim U(0, 1), \quad (7)$$

which controls the transition between the two search strategies during the iterative optimization process.

Starting from the original ARO algorithm, this study introduces a Cauchy Mutation and Adaptive Crossover Strategy (CMACS) to construct the improved CMACS–ARO optimizer. In high-dimensional configuration spaces, standard ARO may converge prematurely or refine too weakly around promising regions, and recent ARO variants similarly emphasize the importance of balancing global exploration with local exploitation [36]. The CMACS–ARO optimizer is designed as a response to that problem by coupling heavy-tailed escape with fitness-conditioned recombination. The updated candidate solution includes a Cauchy-distributed perturbation term around the current global best solution:

$$\mathbf{x}_{\text{new}}^* = \mathbf{x}^* + \mathbf{x}^* \cdot \text{Cauchy}(0, 1), \quad (8)$$

where $\mathbf{x}^*$ denotes the current global best position. This mutation endows the update process with a heavy-tailed jump mechanism, increasing the probability of long-range moves and preserving accessibility to distant yet potentially superior regions of the hyperparameter space.

At the same time, the crossover probability is adjusted adaptively according to the fitness of each individual:

$$p_c(i) = p_{c,\min} + (p_{c,\max} - p_{c,\min}) \cdot \frac{f(i) - f_{\min}}{f_{\max} - f_{\min}} \cdot \frac{t}{T}, \quad (9)$$

where $f(i)$ is the fitness of the i th individual, $f_{\max}$ and $f_{\min}$ are the maximum and minimum fitness values of the population, and $p_{c,\max}$ and $p_{c,\min}$ denote the upper and lower bounds of the crossover probability, respectively.

2.3. LSTM–Attention Forecasting Model and Hyperparameter Setting

LSTM [17,37] is introduced as the inner sequential learner through which recent market information is compressed into an evolving latent state. By combining a cell state with input, forget, and output gates, LSTM alleviates the vanishing- and exploding-gradient problems that undermine long-sequence learning. The forward propagation process of the LSTM unit is given by

$$f_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f), \quad (10)$$

$$i_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i), \quad (11)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c), \quad (12)$$

$$\mathbf{C}_t = f_t \odot \mathbf{C}_{t-1} + i_t \odot \tilde{\mathbf{C}}_t, \quad (13)$$

$$o_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o), \quad (14)$$

$$\mathbf{h}_t = o_t \odot \tanh(\mathbf{C}_t), \quad (15)$$

where f_t , i_t , and o_t denote the forget gate, input gate, and output gate, respectively; $\tilde{\mathbf{C}}_t$ is the candidate cell state; $\mathbf{C}_t$ is the current cell state; $\mathbf{h}_t$ is the hidden state; $\mathbf{x}_t$ is the input at time step t ; $\mathbf{W}$ and $\mathbf{b}$ denote weight matrices and bias vectors; $\sigma(\cdot)$ denotes the sigmoid function; and $\odot$ denotes the Hadamard product.

Figure 2 visualizes the gate operations and state transitions specified in the preceding equations.

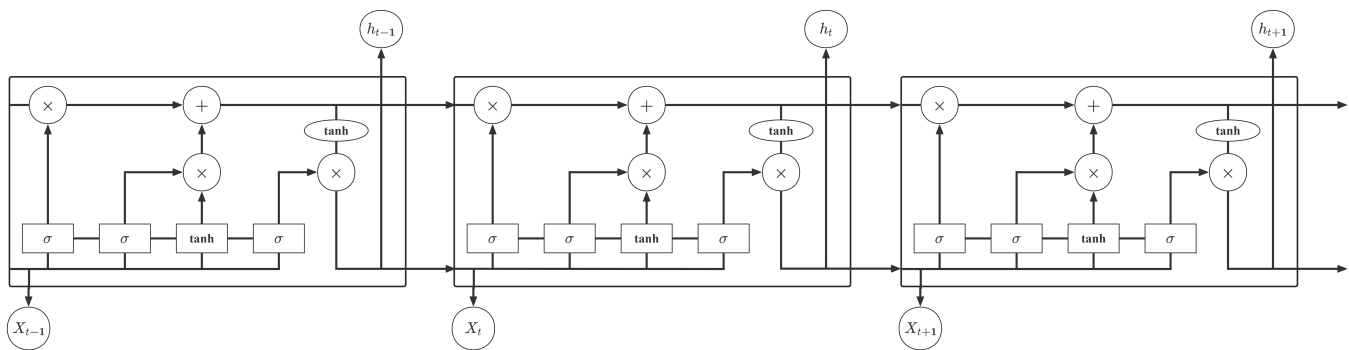


Figure 2. Structure of the LSTM unit.

To further improve forecasting performance, a temporal attention mechanism is introduced after the LSTM layer so that the model can reweight hidden states according to their marginal predictive relevance for next-day price formation rather than treating all lags symmetrically; this design follows the broader idea of attention-based relevance weighting in neural sequence modeling [38]. The attention weight of time step t is computed as

$$\alpha_t = \frac{\exp\{\mathbf{u}^\top \tanh(\mathbf{W}_a \mathbf{h}_t + \mathbf{b}_a)\}}{\sum_{\tau=1}^k \exp\{\mathbf{u}^\top \tanh(\mathbf{W}_a \mathbf{h}_\tau + \mathbf{b}_a)\}}, \quad (16)$$

and the attention-weighted hidden representation is

$$\mathbf{h}_{\text{att}} = \sum_{t=1}^k \alpha_t \mathbf{h}_t. \quad (17)$$

The resulting representation is then fed into a fully connected layer to generate the final price forecast. Here, α_t is the normalized temporal attention weight, satisfying $\sum_{t=1}^k \alpha_t = 1$; $\mathbf{W}_a$, $\mathbf{b}_a$, and $\mathbf{u}$ are learned attention parameters.

Within the LSTM–Attention model, the number of hidden units, batch size, and training epochs are the hyperparameters most directly implicated in model capacity, training stability, and generalization. Because hyperparameter tuning can materially affect the observed performance of machine-learning and deep-learning models, validation-driven search is widely recommended over purely manual configuration [39–42]. To balance expressive flexibility against overfitting risk, this study optimizes only these three quantities, while fixing other settings such as the optimizer, learning rate, and activation function at commonly used values.

2.4. Implementation Procedure of the Proposed CALA Framework

Embedding the CMACS–ARO optimizer into the LSTM–Attention model to optimize hidden units, batch size, and training epochs can be interpreted as an approximate solution procedure for the outer problem in Equation (4). The implementation consists of four steps. First, the objective function is defined with reference to the LSTM–Attention model, using root mean squared error (RMSE) on the validation set as the primary criterion. Second, an initial population is generated randomly within the predefined hyperparameter search space, with each individual representing one admissible LSTM–Attention configuration. Third, in each iteration, candidate positions are updated according to the hiding–searching behavioral modes of CMACS–ARO, with Cauchy mutation applied around the current global best solution and crossover probability adjusted adaptively according to individual fitness. Fourth, when the maximum number of iterations is reached or when the improvement in the global best fitness remains below a preset threshold for several consecutive

generations, the search is terminated, and the hyperparameter combination associated with the global best individual is returned.

Figure 3 visualizes the validation-based outer-loop procedure. The foregoing components are assembled into an integrated CALA framework for short-term agricultural price forecasting. Multi-source spot–futures observations are first converted into supervised sequential samples, the LSTM–Attention predictor then learns a relevance-weighted representation of those samples, and CMACS–ARO searches for the configuration under which that predictor generalizes most effectively. The original workflow uses chronological partitions, with scaler parameters estimated from the training partition and then applied to later partitions. CALA configuration selection is validation-driven. In this minimal revision, the original CALA aggregate values are retained as archived reported results, while persistence/random-walk and ARIMA values are calculated on the same archived comparison block. Sliding-window samples are constructed without using future observations beyond the prediction date.

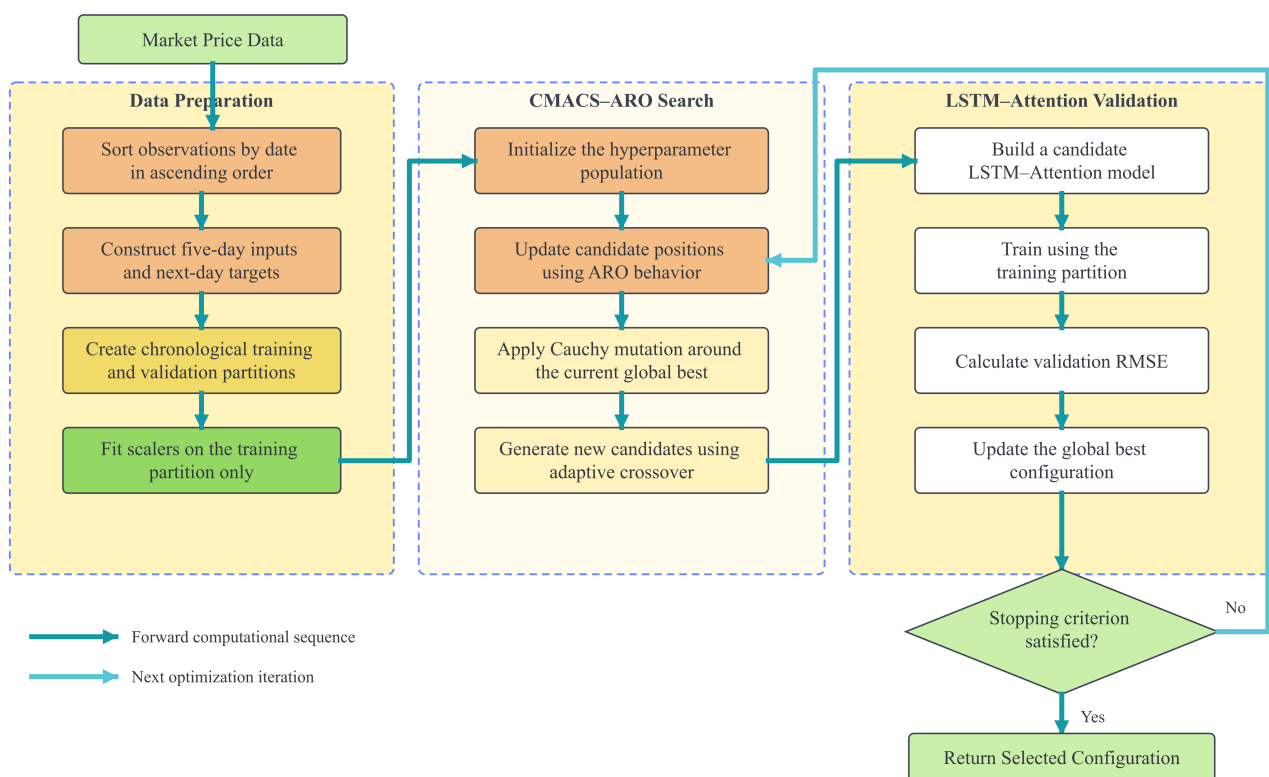


Figure 3. Workflow of the CALA-based hyperparameter optimization procedure. Hyperparameter updates use validation RMSE; final reporting follows configuration selection.

3. Case Study

3.1. Data Description

Wheat is chosen as the main empirical setting because it offers a tractable yet economically meaningful environment in which national futures prices, regional spot prices, and related futures contracts can be analyzed jointly. The data are collected from the Brick Agricultural Data Platform and public sources, including the Dalian Commodity Exchange and the Zhengzhou Commodity Exchange, and cover the period from December 2012 to September 2023. Twelve candidate predictors are retained: the national average wheat price; regional wheat spot prices for Fuzhou, Handan, Heihe, Jining, Shijiazhuang, Tianjin, Yuncheng, and Zibo; and the futures prices of corn, bran, and wheat middling. The target

variable is the wheat futures price. Descriptive statistics are reported in Table 1, while pairwise correlations are visualized in Figure 4.

Table 1. Descriptive statistics of the wheat dataset.

Variable	Observations	Maximum (Yuan/t)	Minimum (Yuan/t)	Mean (Yuan/t)	Std. Deviation
National average wheat spot price	2680	3305.56	2183.57	2560.40	245.14
Wheat futures price (target)	2680	3199.00	1948.00	2513.70	249.98
Regional wheat spot prices (8 regions)	2680	3410.00	1860.00	2319.55–2637.84	228.74–303.02
Related commodity futures prices	2680	3009.33	808.57	1673.88–2201.32	360.42–423.92

Because futures prices are observed only on trading days, the raw wheat futures series contains calendar gaps that would otherwise complicate sequential sample construction. To preserve a consistent trading-day index, non-trading dates with unavailable observations, including weekends and market holidays, are omitted, and all data are ordered chronologically. After preprocessing, the final wheat dataset contains 2680 valid observations. The dataset is then divided chronologically into training, validation, and test subsets at a ratio of 6:2:2. Before model estimation, all variables are normalized to the interval [0, 1]:

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (18)$$

where X denotes the original value and $X_{\min}$ and $X_{\max}$ denote the minimum and maximum values of the corresponding variable, respectively.

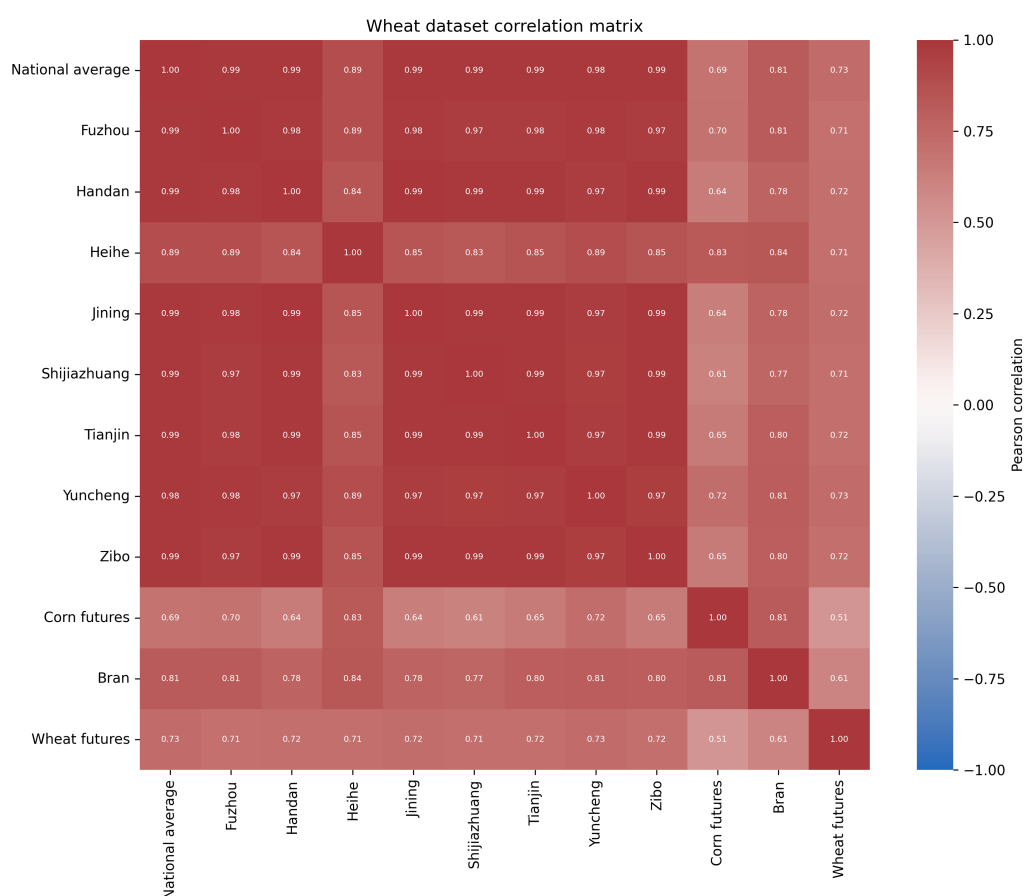


Figure 4. Correlation heatmap of the wheat dataset, shown with a color-blind-friendly scale and a correlation bar.

Considering that price movements may be affected by temporal dependence, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are employed to examine the degree to which past wheat futures prices influence current prices:

$$\text{ACF}_k = \frac{\text{Cov}(Y_t, Y_{t-k})}{\text{Var}(Y_t)}, \quad (19)$$

$$\text{PACF}_k = \text{Corr}(Y_t, Y_{t-k} \mid Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}). \quad (20)$$

The ACF and PACF plots are presented in Figures 5 and 6. In the ACF plot, the autocorrelations remain relatively high from lag 0 to lag 20, indicating strong serial dependence in the wheat futures price series. In the PACF plot, significant correlations are mainly observed at lag 0 and lag 1.

Against this empirical backdrop, the study adopts a sliding-window approach to construct forecasting samples. Specifically, the target price and related features from five consecutive trading days are used as inputs, while the wheat futures price on the sixth trading day is taken as the output, thereby operationalizing a short-horizon next-day forecasting task.

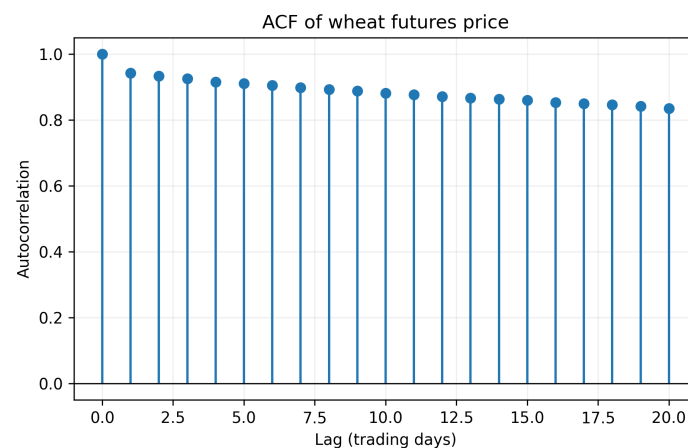


Figure 5. Autocorrelation of the wheat futures price series by lag (trading days); autocorrelation is dimensionless.

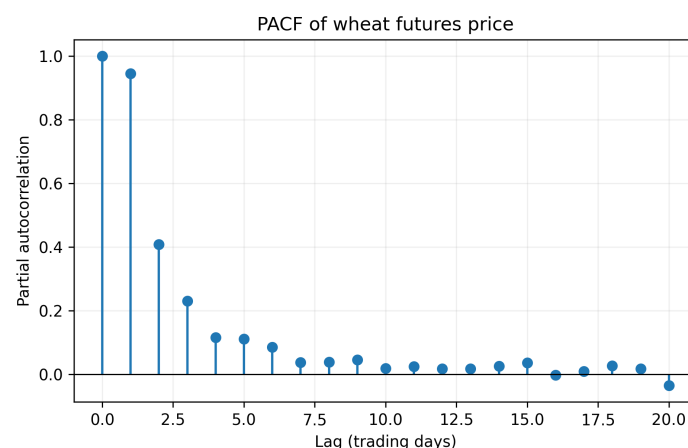


Figure 6. Partial autocorrelation of the wheat futures price series by lag (trading days); partial autocorrelation is dimensionless.

3.2. Benchmark Models

To provide a more discriminating evaluation, the proposed framework is compared with three groups of benchmark models. The first group contains conventional machine-learning models, namely Random Forest [14], LightGBM [43], and a feedforward neural network. The second group includes single deep-learning models, namely LSTM [17] and GRU [44]. The third group contains recurrent and hybrid deep-learning models, including LSTM–Attention, ARO–LSTM–Attention, WOA–LSTM–Attention [45], and ARO–GRU–Attention. These benchmarks are selected because they represent the main comparison objects in this study: static nonlinear learners, recurrent sequence learners, attention-enhanced recurrent models, and metaheuristic-optimized recurrent-attention models. Accordingly, the comparative claims in this paper are limited to the selected benchmark set rather than all possible econometric or forecasting alternatives. This revision additionally calculates persistence/random-walk and ARIMA references on the archived comparison blocks. ARIMA orders $p = 1, \dots, 5$ with one difference and no moving-average term were screened on the selected training portion; ARIMA(2,1,0) was retained for wheat and ARIMA(1,1,0) for soybean. The original CALA outputs are not rerun or modified. The processed inputs, executable script, and result files used for the added persistence/random-walk and ARIMA calculations are provided in Supplementary File S1.

3.3. Evaluation Metrics

Forecasting performance is evaluated using RMSE, mean absolute error (MAE), the coefficient of determination (R^2), and mean absolute percentage error (MAPE), which are standard point-forecast evaluation metrics but should be interpreted in relation to the target forecasting functional and loss function [46,47]. Formal predictive-accuracy comparisons are commonly conducted using the Diebold–Mariano test and its small-sample correction [48,49]; however, such formal tests were not conducted in this minimal revision. Unless otherwise stated, MAE and RMSE are computed on the min–max normalized target scale; therefore, the reported errors should be interpreted as normalized-scale forecast errors rather than direct yuan/t price deviations:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (21)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (22)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (23)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \quad (24)$$

where n denotes the number of evaluation samples, y_i is the observed price, $\hat{y}_i$ is the predicted price, and $\bar{y}$ is the sample mean of the observed prices. In addition, forecast-path directional agreement (FPDA) is reported as an auxiliary diagnostic:

$$\text{FPDA} = \frac{1}{n-1} \sum_{i=2}^n \mathbf{1}[(y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) > 0]. \quad (25)$$

FPDA compares successive changes along the observed and forecast paths; it is, therefore, not interpreted as a standard forecast-origin directional-accuracy measure or as evidence of trading profitability.

4. Results

4.1. Reported Performance of the Proposed Model

After CALA-based hyperparameter optimization, the selected LSTM–Attention configuration consists of 79 hidden units, a batch size of 119, and 190 training epochs. Using the wheat dataset under the training-validation-test split described above, the CALA framework is used to generate next-trading-day forecasts. The resulting prediction path is shown in Figure 7, while the corresponding quantitative evaluation is reported in Table 2.

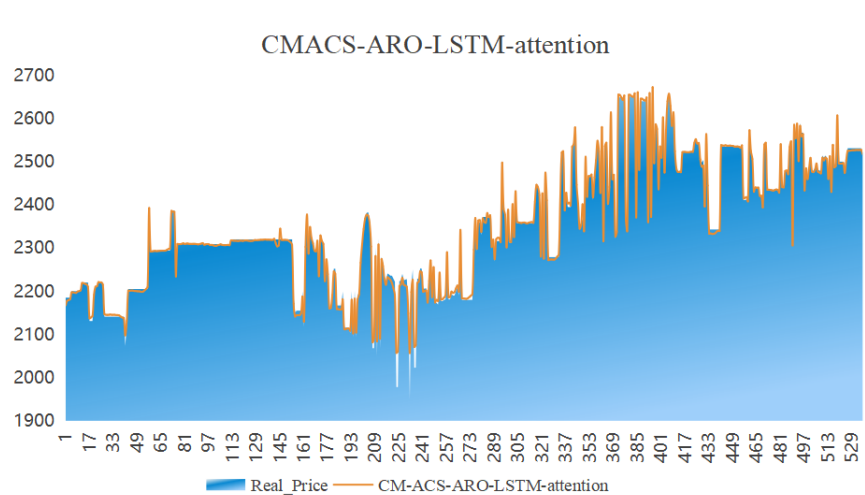


Figure 7. Reported prediction results of CALA on the wheat dataset.

Table 2. Reported forecasting performance of the proposed CALA model.

R^2	MAE	RMSE	MAPE (%)	FPDA/DA
0.995	0.004	0.008	0.021	0.484

The proposed model obtains a close approximation to the observed wheat futures price level under the selected rolling-window design. However, the very high R^2 should be interpreted with caution because agricultural futures prices are highly persistent and level-based metrics can be strong even when directional forecasting remains difficult. The FPDA value of 0.484 is an auxiliary forecast-path statistic and does not establish standard direction-of-change predictability or trading profitability.

4.2. Comparison with Benchmark Models

The comparative results are reported in Table 3, while the overall metric profile is visualized in Figure 8. Three findings emerge. First, the recurrent models generally outperform the conventional machine-learning baselines, a pattern consistent with the importance of sequential dependence relative to static cross-sectional association alone. Second, adding attention improves on the standard LSTM benchmark, which is compatible with the view that predictive content is unevenly distributed across recent observations. Third, optimization appears to add a further layer of gain: CALA delivers the strongest overall performance among all competing models on the level-oriented error metrics reported here.

Table 3. Performance comparison across benchmark models on the archived wheat comparison block. The originally reported CALA values are retained unchanged; persistence/random-walk and ARIMA are newly added.

Model	R^2	MAE	RMSE	MAPE (%)	FPDA/DA
CALA	0.995	0.004	0.008	0.021	0.484
Persistence/Random walk	0.723	0.052	0.102	1.622	0.525
ARIMA(2,1,0)	0.774	0.054	0.092	1.695	0.445
LSTM-Attention	0.969	0.014	0.020	0.057	0.467
ARO-LSTM-Attention	0.981	0.011	0.016	0.045	0.475
WOA-LSTM-Attention	0.974	0.010	0.018	0.046	0.479
ARO-GRU-Attention	0.972	0.013	0.019	0.428	0.474
LSTM	0.941	0.017	0.028	0.407	0.471
GRU	0.902	0.028	0.036	0.534	0.477
Random Forest	0.921	0.033	0.054	0.089	—
LightGBM	0.877	0.048	0.068	0.118	—
FNN	0.899	0.043	0.062	0.107	—



Figure 8. Performance comparison across benchmark models on the wheat dataset.

The newly added statistical references have RMSE values of 0.102 (persistence/random walk) and 0.092 [ARIMA(2,1,0)], compared with the unchanged reported CALA RMSE of 0.008. Within this archived comparison, the reported CALA value remains the lowest level-forecast error, while FPDA is retained only as an auxiliary diagnostic. The pre-existing nested neural comparisons remain descriptive rather than a controlled ablation of attention, Cauchy mutation, or adaptive crossover (Figure 9).

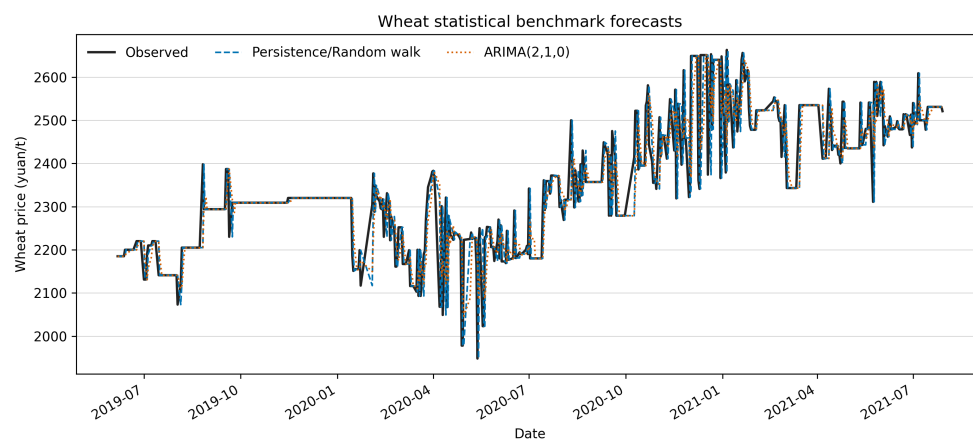


Figure 9. Observed wheat futures prices and newly added persistence/random-walk and ARIMA(2,1,0) forecasts on the 536-observation archived comparison block. The horizontal axis is date, and the vertical axis is wheat futures price (yuan/t).

4.3. Supplementary Application on Soybean Data

The supplementary application asks whether the proposed framework still behaves competitively when the commodity environment changes. The same modeling procedure is applied to soybean price data. The preprocessing steps and sample-construction strategy remain the same as in the wheat experiment. The soybean dataset includes multiple spot prices from major production regions together with diesel prices and the RMB–USD exchange rate, yielding a total of 16 candidate features. The correlation structure of these variables is illustrated in Figure 10; the forecasting results are shown in Figure 11; and the corresponding evaluation metrics are summarized in Table 4.

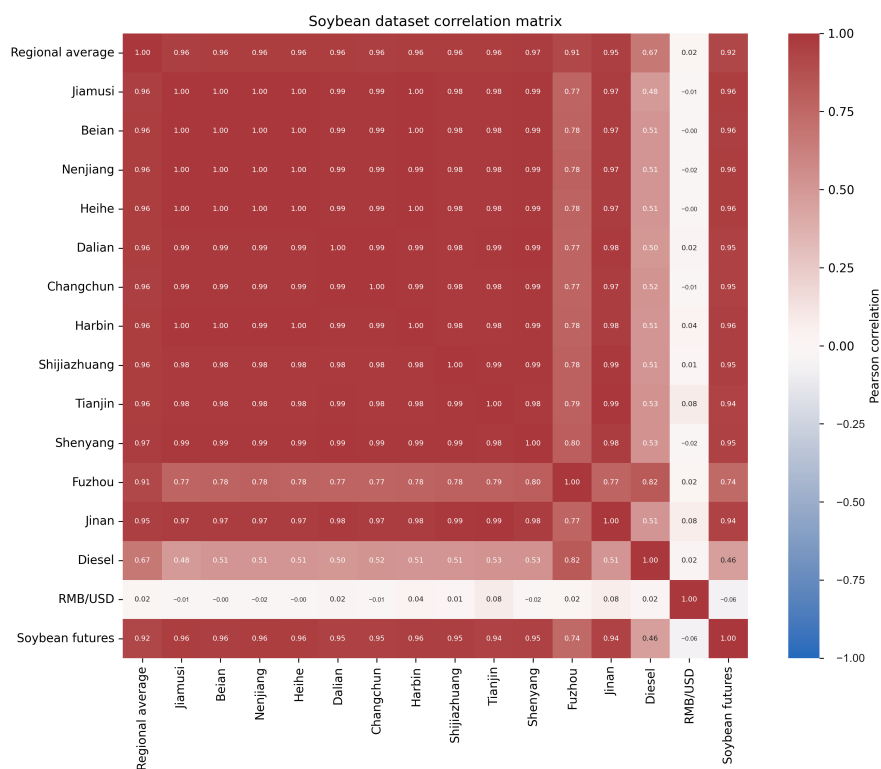


Figure 10. Correlation heatmap of the soybean dataset, with consistently rounded values and a readable correlation scale.

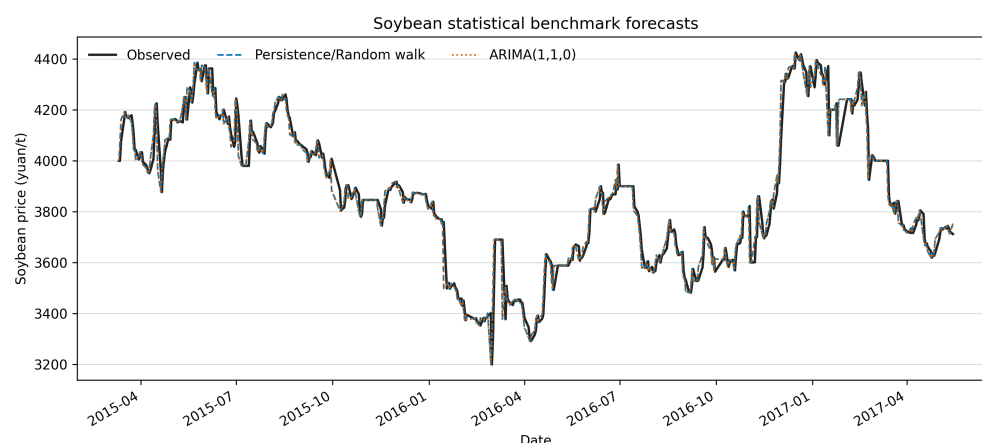


Figure 11. Observed soybean prices and newly added persistence/random-walk and ARIMA(1,1,0) forecasts on the archived soybean comparison block. The horizontal axis follows the supplied experiment order, and the vertical axis is soybean price (yuan/t).

Table 4. Comparison of CALA with newly added statistical benchmarks on the archived soybean comparison block. The originally reported CALA values are retained unchanged.

Model	R^2	MAE	RMSE	MAPE (%)	FPDA/DA
CALA	0.971	0.012	0.014	0.052	0.719
Persistence/Random walk	0.966	0.009	0.015	0.790	0.488
ARIMA(1,1,0)	0.966	0.009	0.015	0.802	0.492

On the soybean dataset, the originally reported CALA result has an R^2 of 0.971, an MAE of 0.012, an RMSE of 0.014, a MAPE of 0.052, and an FPDA/DA value of 0.719. The added persistence and ARIMA references have slightly smaller MAE values (0.009) but a slightly larger RMSE (0.015). Thus, the soybean comparison is mixed and should be interpreted as a supplementary application rather than conclusive proof of broad cross-commodity generality.

5. Discussion

The discussion is most coherent when it returns to the mechanism stated at the outset rather than treating the paper as a simple model contest. Once next-trading-day forecasting is framed as conditional-expectation approximation under a constrained spot-futures information set, performance depends on two coupled elements: the adequacy of the inner relevance-weighted sequence model and the quality of the outer-loop capacity calibration. The empirical gains of CALA are consistent with the possibility that both elements matter in the present application, although the design does not identify their marginal effects with experimental precision and therefore cannot rule out observationally equivalent explanations.

First, the comparison between CALA and the non-optimized LSTM–Attention benchmark is consistent with a role for capacity calibration, but the current design does not identify the separate contribution of attention, Cauchy mutation, or adaptive crossover. The observed error differences are therefore descriptive and should not be treated as a controlled ablation or causal mechanism test.

Second, the comparison between LSTM–Attention and the standard LSTM and GRU baselines is consistent with the proposition that price-relevant information is not distributed uniformly across recent observations. The attention layer appears to improve forecasting quality by reallocating representational mass toward historically informative states rather

than by imposing symmetric treatment across all lags. This interpretation is especially plausible in a linked spot–futures setting, where information arrival is temporally uneven and short-lived shocks may carry disproportionate predictive content.

Third, the soybean application provides mixed supplementary evidence rather than a generalization test. Although the reported CALA RMSE is slightly lower than the added statistical references, the references have slightly lower MAE. The case therefore does not establish robustness or broad external validity.

These findings also sharpen the boundary of the contribution. The present evidence does not imply that CALA universally dominates every plausible forecasting architecture, nor does it show that the observed gains would survive unchanged under materially different data regimes, forecast horizons, or decision objectives. The added statistical references are reported on the same archived comparison block as the original aggregate CALA results; repeated CALA runs, formal predictive-accuracy tests, and an equal-budget component ablation are not added in this minimal revision. The reported FPDA value is also not evidence of standard direction-of-change skill or trading value.

6. Conclusions

This paper develops a short-horizon agricultural price forecasting framework, namely CALA, that couples an LSTM–Attention predictor with the proposed CMACS–ARO outer-loop optimizer. Its central contribution is to formulate next-trading-day agricultural price forecasting as a bilevel learning problem: the inner layer performs relevance-weighted sequence modeling, whereas the outer layer searches over selected hyperparameter configurations. CMACS–ARO is designed to augment ARO with Cauchy perturbation and fitness-adaptive crossover; the current results do not independently verify the contribution of these mechanisms.

The empirical evidence supports three qualified conclusions. First, the original reported CALA aggregate values remain lower than the newly added persistence/random-walk and ARIMA errors on the archived wheat comparison block. Second, existing nested neural comparisons are descriptive and do not demonstrate the independent benefit of attention, Cauchy mutation, or adaptive crossover. Third, the soybean comparison is mixed: reported CALA RMSE is slightly lower, whereas the statistical references have slightly lower MAE. The results should therefore be interpreted as one-step price-level evidence, not as proof of directional skill, trading profitability, or broad generalization.

More broadly, the results imply that predictive gains in short-horizon commodity forecasting need not arise from ever more elaborate architectures alone. Under a carefully designed validation framework, they may also emerge from tighter alignment among model capacity, search dynamics, and market information structure. Practically, this suggests that more reliable short-horizon forecasts can support production planning, procurement, inventory management, price–risk awareness, and real-time market monitoring across agricultural supply chains.

Several directions remain open for future research. The present study focuses on one-step-ahead point forecasting and a fixed five-day window. Repeated CALA runs, confidence intervals, formal predictive-accuracy tests, equal-budget component ablations, multi-step forecasts, prediction intervals, and broader commodity validation should be implemented before stronger reliability or generalization conclusions are drawn.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agriculture16151669/s1>, Supplementary File S1: Reproducibility package for the added persistence/random-walk and ARIMA reference calculations, including a README file, processed wheat and soybean inputs, the executable Python script, and the associated result files.

Author Contributions: Conceptualization, C.S. and Y.G.; methodology, C.S. and Y.G.; software, C.S. and Y.Z.; investigation, Y.C.; data curation, C.S., Y.Z. and S.H.; writing—original draft preparation, C.S. and S.H.; writing—review and editing, X.L. and Y.G.; visualization, Y.C.; supervision, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Natural Science Foundation of China (No. 72501197), Chengdu Philosophy and Social Sciences Planning Project (No. 25CS090) and Southwest Sichuan (Ya'an) Rainstorm Laboratory (No. CXNBYSYSWZD202501).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The processed inputs, executable script, and result files for the newly added persistence/random-walk and ARIMA calculations are provided as Supplementary Materials. The exact original CALA code, random seeds, and complete optimization configuration are not reconstructed in this minimal revision.

Acknowledgments: The authors thank all individuals who provided technical or administrative support during the study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACF	Autocorrelation function
ARO	Artificial Rabbits Optimization
CALA	CMACS–ARO–LSTM–Attention
CMACS–ARO	CMACS-enhanced Artificial Rabbits Optimization
CMACS	Cauchy Mutation and Adaptive Crossover Strategy
FPDA	Forecast-path directional agreement
GRU	Gated recurrent unit
LSTM	Long Short-Term Memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
PACF	Partial autocorrelation function
RMSE	Root mean squared error
WOA	Whale Optimization Algorithm

References

1. Wang, J.; Wang, Z.; Li, X.; Zhou, H. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *Int. J. Forecast.* **2022**, *38*, 21–34. [\[CrossRef\]](#)
2. Zeng, L.; Ling, L.; Zhang, D.; Jiang, W. Optimal forecast combination based on PSO-CS approach for daily agricultural future prices forecasting. *Appl. Soft Comput.* **2023**, *132*, 109833. [\[CrossRef\]](#)
3. Ben Kaabia, M.; Gil, J.M. Forecasting agricultural commodity prices using asymmetric error correction models with generalized autoregressive heteroskedasticity. *J. Forecast.* **2007**, *26*, 87–111.
4. Guo, Y.; Hu, X.; Wang, Z.; Tang, W.; Liu, D.; Luo, Y.; Xu, H. The butterfly effect in the price of agricultural products: A multidimensional spatial-temporal association mining. *Agric. Econ.* **2021**, *67*, 489–502. [\[CrossRef\]](#)
5. Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl. Soft Comput.* **2020**, *86*, 105837. [\[CrossRef\]](#)
6. Gu, Y.H.; Jin, D.; Yin, H.; Zheng, R.; Piao, X.; Yoo, S.J. Forecasting Agricultural Commodity Prices Using Dual Input Attention LSTM. *Agriculture* **2022**, *12*, 256. [\[CrossRef\]](#)
7. Zhang, T.; Tang, Z. Agricultural commodity futures prices prediction based on a new hybrid forecasting model combining quadratic decomposition technology and LSTM model. *Front. Sustain. Food Syst.* **2024**, *8*, 1334098. [\[CrossRef\]](#)
8. Thaker, A.; Chan, L.H.; Sonner, D. Forecasting Agriculture Commodity Futures Prices with Convolutional Neural Networks with Application to Wheat Futures. *J. Risk Financ. Manag.* **2024**, *17*, 143. [\[CrossRef\]](#)

9. Terasvirta, T.; van Dijk, D.; Medeiros, M.C. Forecasting with smooth transition autoregressive models. *J. Forecast.* **2006**, *25*, 537–554.
10. Xu, Z.; Mohsin, M.; Ullah, K.; Ma, X. Using econometric and machine learning models to forecast crude oil prices: Insights from economic history. *Resour. Policy* **2023**, *83*, 103614. [\[CrossRef\]](#)
11. Zhu, H.; Xu, R.; Deng, H. A novel STL-based hybrid model for forecasting hog price in China. *Comput. Electron. Agric.* **2022**, *198*, 107068. [\[CrossRef\]](#)
12. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE* **2018**, *13*, e0194889. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.* **2020**, *36*, 54–74. [\[CrossRef\]](#)
14. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
15. Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Syst. Appl.* **2015**, *42*, 259–268. [\[CrossRef\]](#)
16. Liu, Y.; Duan, Q.; Wang, D.; Zhang, Z.; Liu, C. Prediction for hog prices based on similar sub-series search and support vector regression. *Comput. Electron. Agric.* **2019**, *157*, 581–588. [\[CrossRef\]](#)
17. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Fischer, T.; Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* **2018**, *270*, 654–669. [\[CrossRef\]](#)
19. Zhang, S.; Luo, J.; Wang, S.; Liu, F. Oil price forecasting: A hybrid GRU neural network based on decomposition-reconstruction methods. *Expert Syst. Appl.* **2023**, *218*, 119617. [\[CrossRef\]](#)
20. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175. [\[CrossRef\]](#)
21. Banerjee, T.; Sinha, S.; Choudhury, P. Long term and short term forecasting of horticultural produce based on the LSTM network model. *Appl. Intell.* **2022**, *52*, 16298–16328. [\[CrossRef\]](#)
22. Wu, C.; Wang, J.; Hao, Y. Deterministic and uncertainty crude oil price forecasting based on outlier detection and modified multi-objective optimization algorithm. *Resour. Policy* **2022**, *77*, 102780. [\[CrossRef\]](#)
23. Wu, J.; Hu, Y.; Wu, D.; Yang, Z. An aquatic product price forecast model using VMD-IBES-LSTM hybrid approach. *Agriculture* **2022**, *12*, 1185. [\[CrossRef\]](#)
24. Lim, B.; Zohren, S. Time-series forecasting with deep learning: A survey. *Philos. Trans. R. Soc. A* **2021**, *379*, 20200209. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Lim, B.; Arik, S.O.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [\[CrossRef\]](#)
26. Benidis, K.; Rangapuram, S.S.; Flunkert, V.; Wang, Y.; Maddix, D.; Turkmen, C.; Gasthaus, J.; Bohlke-Schneider, M.; Salinas, D.; Stella, L.; et al. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Comput. Surv.* **2022**, *55*, 1–36. [\[CrossRef\]](#)
27. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent neural networks for time series forecasting: Current status and future directions. *Int. J. Forecast.* **2021**, *37*, 388–427. [\[CrossRef\]](#)
28. Sezer, O.B.; Gudelek, M.U.; Ozbayoglu, A.M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Appl. Soft Comput.* **2020**, *90*, 106181. [\[CrossRef\]](#)
29. Peng, L.; Liu, S.; Liu, R.; Wang, L. Effective long short-term memory with differential evolution algorithm for electricity price prediction. *Energy* **2018**, *162*, 1301–1314. [\[CrossRef\]](#)
30. Ye, L.; Li, Y.; Liang, W.; Song, Q.; Liu, Y.; Qin, X. Vegetable price prediction based on PSO-BP neural network. In Proceedings of the 2015 8th International Conference on Intelligent Computation Technology and Automation, Nanchang, China, 14–15 June 2015; pp. 1093–1096.
31. Taieb, S.B.; Bontempi, G.; Atiya, A.F.; Sorjamaa, A. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Syst. Appl.* **2012**, *39*, 7067–7083. [\[CrossRef\]](#)
32. Gneiting, T.; Katzfuss, M. Probabilistic forecasting. *Annu. Rev. Stat. Appl.* **2014**, *1*, 125–151. [\[CrossRef\]](#)
33. Chakraborty, S.; Jagabathula, S.; Subramanian, L.; Venkataraman, A. Frontiers in Operations: News Event-Driven Forecasting of Commodity Prices. *Manuf. Serv. Oper. Manag.* **2024**, *26*, 1286–1305. [\[CrossRef\]](#)
34. Bergmeir, C.; Benitez, J.M. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **2012**, *191*, 192–213. [\[CrossRef\]](#)
35. Wang, L.; Cao, Q.; Zhang, Z.; Mirjalili, S.; Zhao, W. Artificial rabbits optimization: A new bio-inspired meta-heuristic algorithm for solving engineering optimization problems. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105082. [\[CrossRef\]](#)
36. Bakir, H. Dynamic fitness-distance balance-based artificial rabbits optimization algorithm to solve optimal power flow problem. *Expert Syst. Appl.* **2024**, *240*, 122460. [\[CrossRef\]](#)

37. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [[CrossRef](#)] [[PubMed](#)]
38. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
39. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
40. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian optimization of machine learning algorithms. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 2951–2959.
41. Li, L.; Jamieson, K.; DeSalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *J. Mach. Learn. Res.* **2018**, *18*, 1–52.
42. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [[CrossRef](#)]
43. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
44. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734.
45. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [[CrossRef](#)]
46. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [[CrossRef](#)]
47. Gneiting, T. Making and evaluating point forecasts. *J. Am. Stat. Assoc.* **2011**, *106*, 746–762. [[CrossRef](#)]
48. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [[CrossRef](#)]
49. Harvey, D.; Leybourne, S.; Newbold, P. Testing the equality of prediction mean squared errors. *Int. J. Forecast.* **1997**, *13*, 281–291. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.