electronics

MDPI

Article

From Time-Series Prediction to System Modeling: A Dual-Attention Framework for Multi-Source Interaction in Soybean Futures Markets

Hongjiu Liu [1], Qingyang Liu [2,*] and Yanrong Hu [1,*]

[1] College of Mathematics and Computer Science, Zhejiang A & F University, Hangzhou 311300, China; joe_hunter@zafu.edu.cn
[2] GWDG and Institute of Informatics, Georg-August-Universität Göttingen, 37073 Göttingen, Germany
* Correspondence: qingyang.liu@stud.uni-goettingen.de (Q.L.); yanrong_hu@zafu.edu.cn (Y.H.)

Abstract

Futures price forecasting is often treated as a time-series prediction task. However, agricultural futures markets function as complex information systems in which prices emerge from the interaction of heterogeneous sources, including trading behavior and news-driven sentiment. Ignoring such cross-domain interactions limits the ability of traditional models to capture systemic price dynamics. This study reconceptualizes soybean futures forecasting as a multi-source information interaction problem and proposes a dual-attention LSTM framework to model cross-system coupling effects. A RoBERTa-based sentiment classifier is first developed to quantify market sentiment from news headlines. The extracted sentiment features are then integrated with historical trading variables and fed into an LSTM network equipped with temporal and feature-level attention mechanisms to capture dynamic evolution patterns and heterogeneous factor interactions. Empirical results show that the proposed system consistently outperforms conventional models. With a sliding window of 30 and a forecasting horizon of 7 days, the R^2 improves from 0.922 to 0.9797, demonstrating enhanced capability in modeling medium-term price dynamics. The findings highlight that futures forecasting should be approached as a system-level information integration task rather than a purely statistical extrapolation problem.

Keywords: complex systems; agricultural futures markets; multi-source heterogeneous data; information interaction modeling; dual-attention mechanism; system-level forecasting

Check for updates

Academic Editor: Domenico Ursino

Received: 4 April 2026
Revised: 1 May 2026
Accepted: 4 May 2026
Published: 8 May 2026

1. Introduction

Complex time series forecasting has long been a central topic in neural networks and learning systems, with important applications in financial engineering, energy systems, and commodity markets. Soybean futures prices represent a typical example of price sequences influenced by multiple heterogeneous factors, including macroeconomic conditions, climate variability, trade policies, and market expectations. The interactions among these factors generate nonlinear, nonstationary, and structurally evolving dynamics, making soybean futures an important testbed for advanced modeling approaches [1].

However, treating futures forecasting purely as a univariate or multivariate time-series prediction task overlooks a fundamental characteristic of modern financial markets: prices emerge from the interaction of multiple information subsystems. Empirical studies have increasingly operationalised this conceptualisation using econometric and information-theoretic methods. For instance, Granger causality tests, transfer entropy, and spillover indices have been extensively employed to quantify the information flow between public news, market sentiment, and asset prices [2–4]. In agricultural futures markets, trading behavior reflects endogenous market dynamics, while textual news, policy announcements, and external signals represent exogenous information diffusion processes. These heterogeneous components form an interconnected information system in which cross-domain interactions drive price evolution. From this perspective, forecasting is not merely statistical extrapolation but a system-level modeling problem.

Traditional econometric approaches, such as autoregressive integrated moving average (ARIMA) and vector autoregression (VAR), are generally constructed under linear or weakly nonlinear assumptions and rely primarily on structured historical trading data [5]. Although these models offer interpretability and solid theoretical foundations, they often fail to capture high-order nonlinear dependencies and cross-system interactions, particularly when unstructured information sources influence market behavior. As financial markets become increasingly information-driven, models based solely on numerical inputs are insufficient to represent the full structure of the underlying agricultural financial system.

Recent advances in deep learning have significantly enhanced nonlinear sequence modeling capability. Long short-term memory (LSTM) networks effectively capture long-range temporal dependencies through gated mechanisms and have been widely adopted in financial forecasting tasks [6]. Nevertheless, conventional LSTM architectures typically treat input variables and time steps with uniform importance, limiting their ability to selectively emphasize informative components in high-dimensional or heterogeneous settings. Attention mechanisms address this limitation by learning adaptive importance weights, thereby improving representation learning in complex environments.

Recent literature has made significant strides in integrating textual sentiment and structured trading variables within Chinese financial markets. A notable precedent is the work by Wu et al. [7], who developed an integrated Chinese-language framework combining financial indicators and text sentiment for real-estate early warning. However, while their work establishes a foundational precedent for macroscopic risk classification, applying such multimodal integration to commodity futures requires fundamentally different mechanisms beyond a mere change in asset class. Unlike the real estate market, which is policy-driven, relatively illiquid, and evaluated for long-term discrete risk thresholds (early warning), the soybean futures market is highly liquid and continuously driven by high-frequency global information shocks (e.g., weather, international trade). Consequently, our study differs structurally: rather than performing risk classification, we focus on continuous time-series price forecasting. This necessitates our proposed DA-LSTM architecture, which utilizes a dual-attention mechanism to dynamically arbitrate the high-frequency, shifting dominance between unstructured news sentiment and structural trading momentum at a granular daily level.

To address these challenges, this paper conceptualizes soybean futures forecasting as a multi-source heterogeneous information interaction problem and proposes a dual-attention LSTM framework to model cross-system coupling effects. Structured trading data are integrated with quantified sentiment features extracted from news text, forming a unified representation of the agricultural financial system. The model employs feature-level attention to capture cross-factor importance and temporal attention to model dynamic dependencies, enabling the simultaneous learning of structural interactions and temporal evolution patterns.

The main contributions of this work are threefold: (1) We propose a multimodal Dual-Attention LSTM (DA-LSTM) forecasting architecture that explicitly models the futures market as an interconnected information system. By building upon and advancing the foundational DA-RNN model [8], our framework incrementally adapts the dual-stage attention mechanism specifically for the joint integration of unstructured text sentiment and structured trading variables. (2) We operationalize the feature-attention mechanism not merely as a generic numerical selector but as a dynamic modality-weighting tool. It adaptively arbitrates the time-varying influence between news-derived sentiment shocks and technical market momentum at each time step, which is jointly optimized with temporal dependency learning to capture market regime shifts. (3) We provide empirical validation on Chinese soybean futures, demonstrating that this system-level integration under the DA-LSTM framework offers observable improvements in predictive performance, thereby offering new insights into how human sentiment and market prices interact in agricultural commodity forecasting.

The remainder of this paper is organized as follows: Section 2 reviews the related work on financial time-series forecasting, sentiment analysis, and the development of attention mechanisms, specifically focusing on the lineage from DA-RNN to current multimodal architectures. Section 3 details the data and methodology, including the data collection process, news topic modeling using BERTopic, sentiment quantification via the RoBERTa-based classifier, and the formal mathematical definition of the proposed DA-LSTM framework. Section 4 presents the experimental results and discussion, providing a comprehensive performance evaluation across different prediction horizons and a comparison with various baseline models. Finally, Section 5 concludes the paper by summarizing the key findings and outlining potential directions for future research.

2. Related Work

From a system perspective, agricultural futures markets can be viewed as multi-layer information interaction systems, where price dynamics emerge from the coupling between information diffusion processes and endogenous market responses. However, existing research often treats these two processes in isolation, failing to capture the dynamic feedback loops inherent in the system. Accordingly, related studies can be categorized into two streams, though we identify significant gaps in their current integration.

2.1. Market Information Extraction as an Information Diffusion Subsystem

Unstructured textual data constitute a critical information diffusion subsystem. While the typical pipeline involves preprocessing and feature extraction [9], traditional methods frequently fail to account for the domain-specific semantic density of agricultural trade.

For *topic modeling*, early matrix factorization (NMF, LSA) and probabilistic models (LDA, HDP, STM) provided foundations for thematic analysis [10]. However, a critical limitation of these models is their reliance on predefined topic numbers and their inability to capture long-range semantic dependencies in fast-moving news cycles. While neural models like BERTopic [11,12] have improved coherence, they often function as “black boxes” that overlook the temporal evolution of policy-driven shocks—a vital factor in soybean markets.

For *sentiment analysis*, the shift from handcrafted features [13] to Transformer-based models (BERT, RoBERTa) has significantly improved accuracy [14,15]. Nevertheless, current literature often ignores the “sentiment inversion” problem in commodity markets, where a “price surge” may be positive for producers but negative for industrial consumers. While the most recent studies in 2025 and 2026 have begun exploring Large Language Models (LLMs) for financial reasoning, their high computational latency often precludes them from high-frequency system integration [16–18]. Consequently, this study utilizes a fine-tuned RoBERTa model to balance deep contextual representation with the operational efficiency required for real-time market systems.

https://doi.org/10.3390/electronics15101988

2.2. Market Response Modeling as a Dynamic System Component

While textual signals reflect exogenous diffusion, price series represent endogenous responses. The evolution of modeling this response reveals a trend toward complexity, yet robustness and cross-modal coupling remain underdeveloped.

Traditional econometric models (ARIMA, GARCH) are limited by linear assumptions and struggle to represent structural breaks caused by sudden geopolitical or climatic information shocks [19–21]. Machine learning approaches (SVM, Random Forests) offer non-linear mapping [22] but lack the recursive memory needed to model the long-term "echoes" of supply chain disruptions.

Deep learning architectures, particularly RNN variants like LSTM and GRU, have set new benchmarks for temporal dependency modeling [23]. The addition of attention mechanisms has further enabled adaptive weighting [24,25]. The seminal DA-RNN [8] serves as the direct architectural antecedent to our framework. However, while DA-RNN was designed for homogeneous numerical sequences, our proposed DA-LSTM provides a critical incremental contribution: it adapts this dual-stage mechanism specifically for multimodal, heterogeneous fusion. By repurposing feature-attention into a dynamic modality-weighting tool, DA-LSTM explicitly arbitrates between unstructured text-sentiment shocks and structured trading momentum, overcoming the limitations of traditional 'shallow coupling' (simple concatenation).

In contrast to these limited approaches, this study adopts a system-level perspective. By integrating a RoBERTa-based engine with a Dual-Attention LSTM architecture, we explicitly model the interaction task rather than simple feature stacking. This addresses the "unevenness" of prior studies by ensuring the sentiment model (RoBERTa [26]) and the response model (LSTM) are coupled through a joint attention mechanism, thereby capturing the systemic complexity of the global soybean financial system.

2.3. Joint Multimodal Fusion Architectures for Financial Forecasting

While the aforementioned streams often treat text mining and time-series modeling as isolated or loosely coupled tasks, recent advancements have driven the emergence of a third paradigm: joint multimodal fusion architectures. These frameworks aim to systematically integrate unstructured textual sentiment with structured trading data to capture complex market dynamics [27].

Several representative works have established foundational architectures for this multimodal integration. For instance, studies have successfully combined sentiment analysis with standard LSTM or GRU backbones to improve equity and commodity forecasting [28,29]. More recently, Transformer-based architectures and self-attention mechanisms have been employed to align news sentiment streams with numerical price sequences, allowing models to selectively focus on specific news events that trigger market volatility [2,30].

However, despite these advancements, most existing fusion models employ a "shallow coupling" strategy—typically early concatenation or late fusion of feature vectors—without dynamically modeling the shifting equilibrium between exogenous shocks and endogenous momentum. Furthermore, they are often designed for highly efficient equity markets rather than the specific supply-demand dynamics of agricultural commodities. In contrast to these approaches, our proposed DA-LSTM framework utilizes a specifically tailored dual-attention mechanism. It goes beyond simple concatenation by dynamically arbitrating the relative dominance of text-derived sentiment versus structural trading variables at each distinct time step, thereby providing a deeper, more adaptive systemic coupling suited for soybean futures forecasting.

https://doi.org/10.3390/electronics15101988

2.4. Emergence of Time-Series Foundation Models

Most recently, the forecasting landscape has been significantly disrupted by the introduction of Time-Series Foundation Models (TSFMs). Models such as Chronos [31], TimesFM [32], Time-MoE [33], Moirai [34], and Lag-Llama [35] have demonstrated remarkable zero-shot forecasting capabilities across diverse domains, including agricultural commodity prices. Pre-trained on massive datasets of univariate and multivariate sequences, these models excel at capturing general temporal patterns and structural breaks without the need for domain-specific fine-tuning.

However, despite their state-of-the-art performance in purely numerical extrapolation, current TSFMs are predominantly unimodal. They generally treat forecasting as an isolated numerical task and lack native architectural mechanisms to dynamically integrate unstructured exogenous shocks, such as text-derived market sentiment. As our study posits, the soybean futures market is an interconnected information system where prices emerge from the interaction between structural trading behavior and real-time news diffusion. While TSFMs represent the pinnacle of numerical time-series modeling, our DA-LSTM framework specifically addresses the multimodal fusion gap. By utilizing a dual-attention mechanism, the proposed framework dynamically arbitrates the shifting influence between text sentiment and numerical momentum, providing a system-level coupling that current unimodal foundation models cannot natively perform.

3. Data and Methodology

3.1. Data Sources and System-Level Data Construction

From a system modeling perspective, agricultural futures prices emerge from the interaction between multiple information subsystems. To construct a unified representation of the soybean futures market, this study integrates structured trading variables and unstructured textual information, corresponding to the market response subsystem and the information diffusion subsystem, respectively.

3.1.1. Information Diffusion Subsystem (Textual Data)

The textual corpus was collected from multiple authoritative financial and industry platforms to comprehensively capture external information signals affecting soybean futures markets:

- **Jintou Futures:** Provides futures price trends, exchange quotations, and market updates. News articles related to the soybean futures sector were collected from this platform.
- **Hexun:** A leading financial portal with a dedicated futures channel. Reports concerning soybeans within the agricultural futures section were systematically crawled.
- **Choice Financial Terminal (East Money):** Offers multi-asset financial data and industry information. Relevant news on soybeans, soybean meal, and soybean oil was obtained from its agricultural futures module.
- **China Soybean Industry Association (CSIA):** The official industry website providing policy, market, and technological updates. News reports were crawled to capture authoritative industry developments.

A total of 18,310 news entries were collected, including headlines, publication dates, authors, sources, full text content, and URLs. The temporal coverage aligns with the numerical dataset to ensure synchronous system-level analysis. These textual data serve as measurable representations of information diffusion, policy signals, and collective market expectations.

3.1.2. Market Response Subsystem (Numerical Data)

The endogenous market dynamics were represented using structured trading data. Historical daily trading data for the *Soybean No. 1* continuous contract listed on the Dalian Commodity Exchange (DCE) were collected via Sina Finance. The dataset includes Open, High, Low, and Close (OHLC) prices, open interest, and trading volume.

The study period spans from 13 April 2011 to 29 December 2023, comprising 4646 daily observations stored in CSV format. These numerical variables characterize the observable market response to internal trading behavior and external information shocks.

3.1.3. Data Acquisition and System Integration

Data acquisition was implemented in Python 3.8 using the `Requests` library. To enhance efficiency and stability, a multi-threaded web crawler based on the Producer–Consumer architecture was developed.

After collection, textual and numerical data were temporally aligned at the daily frequency, enabling the construction of a synchronized multi-source dataset. This integration establishes a unified data foundation for modeling cross-system interactions within the agricultural financial system. Because the proposed framework employs a sliding window approach to forecast future prices (i.e., using historical data up to day T to predict prices at day $T + n$, where $n \geq 1$), we implemented strict temporal alignment protocols to prevent data leakage. First, the daytime trading session of the *Dalian Commodity Exchange* (DCE) officially closes at 15:00 (*Beijing* Time); therefore, news published after 15:00 on day T is exclusively used to predict market responses on subsequent trading days (assigned to day $T + 1$). Second, the futures market is closed during weekends and statutory public holidays, and news published during these non-trading periods is aggregated and assigned to the first subsequent active trading day. Consequently, no future information is leaked into the input window, ensuring the temporal integrity of the forecasting system.

3.2. Methodology

From a system modeling perspective, the proposed framework consists of three interconnected components: (1) an information diffusion modeling module, (2) a sentiment quantification mechanism, and (3) a market response modeling module with cross-system coupling. These components jointly construct a unified representation of the agricultural financial system.

3.2.1. Information Diffusion Modeling via BERTopic

In agricultural futures markets, textual news serves as a key carrier of external information diffusion. To structurally extract latent thematic signals embedded in large-scale news corpora, this study adopts BERTopic for topic modeling. The workflow is structured as follows:

- **Data Preprocessing:** Text is cleaned by removing stopwords, punctuation, and numerals, followed by stemming and lemmatization to standardize vocabulary. We first removed duplicate news titles and irrelevant prefixes. Regular expressions were utilized to retain only Chinese and English characters. The cleaned text was then segmented using the `jieba` library with a custom domain dictionary, followed by the removal of standard stopwords (using the Baidu Stopwords list) via `CountVectorizer`.
- **Text Embedding:** High-dimensional vectors capturing semantic information are generated using Transformer-based models. In this study, the `acge_text_embedding` pre-trained model is employed for processing Chinese text.
- **Dimensionality Reduction:** Uniform Manifold Approximation and Projection (UMAP) is applied to reduce computational complexity while preserving local seman-

tic structures. UMAP was configured with *n_neighbors = 15*, *n_components = 5*, *min_dist = 0.0*, *metric = 'cosine'*, and *random_state = 42* (for deterministic behavior).

- **Clustering:** Potential topics are identified by clustering the reduced vectors using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). We applied HDBSCAN to discover dense clusters with *min_cluster_size = 50 d*, *metric = 'euclidean'*, and *cluster_selection_method = 'eom'*.
- **Topic Keyword Extraction:** The Class-based TF-IDF (c-TF-IDF) method extracts representative keywords for each cluster, facilitating topic labeling and interpretation. We configured the CountVectorizer to extract the top 10 representative words per topic (*top_n_words = 10*).
- **Topic Refinement:** Similar topics are merged or inconsistent topics are split as needed to enhance coherence and quality [12].

Implementation is conducted using the BERTopic library. The `acge_text_embedding` model hosted on HuggingFace is used for embeddings, UMAP for dimensionality reduction, and HDBSCAN automatically regulates the number of topics (Table 1).

Table 1. Parameter settings for the BERTopic pipeline.

Component	Configuration
Embedding	`SentenceTransformer` ("acge_text_embedding")
UMAP	`n_neighbors` = 15, `n_components` = 5, `min_dist` = 0.0, `metric` = 'cosine', `random_state` = 42
HDBSCAN	`min_cluster_size` = 50, `metric` = 'euclidean', `cluster_selection_method` = 'eom'
Representation	`CountVectorizer` (`top_n_words` = 10)

As demonstrated in Table 2, varying the `min_cluster_size` parameter significantly impacts the granularity of the extracted topics. To balance topic coherence and mitigate feature sparsity, we selected the configuration yielding 10 distinct clusters (which achieved the highest c_v and u_{mass} coherence scores). In BERTopic, one cluster is inherently designated as the outlier/noise bucket (Topic -1). By discarding this noise cluster, we successfully recovered 9 highly coherent and semantically meaningful latent topics from the soybean futures news corpus.

Table 2. Comparison of topic coherence scores under different BERTopic parameters.

Topic Model	Number of Topics	Coherence (c_v)	Coherence (u_{mass})
BERTopic50	45	0.776786	−0.609076
BERTopic100	18	0.781242	−0.433905
BERTopic150	10	0.801291	−0.320123

We explicitly state that BERTopic in this study is not merely exploratory; it serves as the exact feature-generation mechanism. The topic-to-feature mapping is strictly one-to-one: the 9 recovered topics act as 9 distinct "information channels". For each trading day, news items are routed into their respective topic channels, and their sentiment scores are aggregated (as detailed in Equation (1)). Consequently, these 9 topic-specific daily sentiment series directly constitute the 9 additional text-derived features inputted into the DA-LSTM framework.

3.2.2. Sentiment Quantification as a System-Level Impact Factor

While topic modeling captures structural themes, sentiment analysis quantifies the directional influence of information diffusion on market expectations. Let $S_{t,f}$ denote the aggregated sentiment score of influence factor f on day t, and $n_{t,f}$ represent the number of news items mentioning factor f on that day. Let $S_{t,f,i}$ be the sentiment score of individual news item i. The daily sentiment score is computed as:

$$S_{t,f} = \frac{1}{n_{t,f}} \sum_{i=1}^{n_{t,f}} S_{t,f,i} \tag{1}$$

A higher $S_{t,f}$ indicates a stronger potential influence on price appreciation. Individual sentiment scores $S_{t,f,i}$ are obtained using RoBERTa-based sentiment analysis on news headlines [36].

Specifically, the identity of the influence factors f corresponds exactly to the 9 distinct latent topics discovered by the BERTopic model (as visualized later in Section 4.3). To clarify the exact composition of our 15-dimensional input space, the initial 6 structural features are derived from the trading data (Open, High, Low, Close, Volume, and Open Interest). The remaining 9 features are the topic-specific sentiment scores.

The explicit formulation for these 9 added features is defined as follows:

Source: The textual news corpus collected from the platforms detailed in Section 3.1.1.

Aggregation Window: The temporal window is strictly daily, aligned with the 15:00 cut-off and holiday mapping protocols defined in Section 3.1.3 to prevent any look-ahead bias.

Construction Rule: For a given trading day t and a specific topic $f \in \{1, 2, \ldots, 9\}$, we isolate all news items published within that daily window that are classified under topic f. Their individual RoBERTa sentiment polarities are then averaged using Equation (1) to produce the topic-specific daily sentiment score. If no news for topic f occurs on day t, the score is carried forward or zero-padded based on the temporal setting. This rigorous construction explicitly expands the system's input feature dimension from 6 to 15.

The aggregated sentiment indicator serves as a measurable representation of the information diffusion subsystem's daily impact intensity.

3.2.3. Market Response Modeling with LSTM

Within the agricultural financial system, price dynamics represent endogenous market response behavior to both internal trading activities and external information shocks. To model such dynamic response processes, Long Short-Term Memory (LSTM) networks (Figure 1) are employed due to their capability to capture long-term dependencies [37].

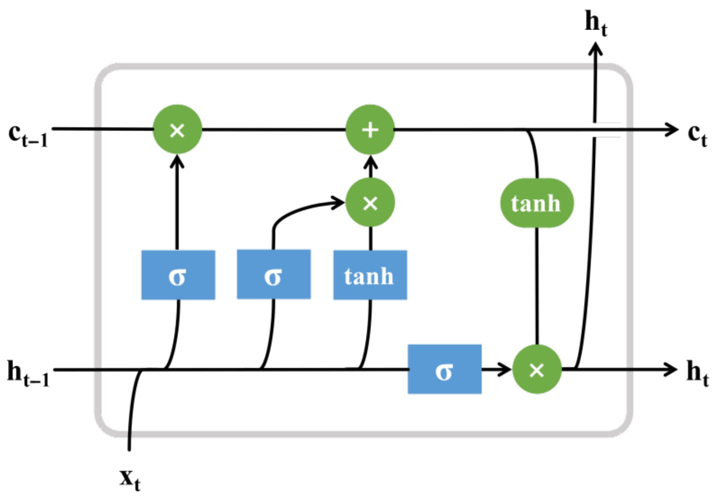


Figure 1. Internal structure of an LSTM unit, illustrating the interaction between the cell state and gating mechanisms.

The LSTM computation proceeds in three phases. First, the *forget gate* determines which information in the cell state should be retained or discarded:

$$f_t = \sigma\Big(W_f \cdot [h_{t-1}, x_t] + b_f\Big) \tag{2}$$

Next, the *input gate* i_t and candidate cell state $\tilde{C}_t$ are computed to update the cell state C_t:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = (C_{t-1} \odot f_t) + (i_t \odot \tilde{C}_t) \tag{5}$$

Finally, the hidden state h_t is computed using the *output gate* o_t, filtering the updated cell state through the hyperbolic tangent activation:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t \odot \tanh(C_t) \tag{7}$$

The activation function tanh is defined as:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{8}$$

From a system perspective, the cell state can be interpreted as an internal memory mechanism that accumulates and filters historical market information over time.

3.2.4. Cross-System Coupling via Dual-Attention LSTM

To explicitly model the interaction between the information diffusion subsystem and the market response subsystem, this study proposes a Dual-Attention LSTM (DA-LSTM) framework. The architecture, illustrated in Figure 2, integrates feature-level (spatial) and temporal-level attention mechanisms within an encoder–decoder structure to capture complex dependencies in multivariate time series.

(1) Feature Attention in the Encoder Given the input sequence $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)^\top \in \mathbb{R}^{T\times n}$, where n is the number of exogenous features. To capture the dynamic correlation between heterogeneous features, the encoder calculates an attention weight α_t^k for the k-th feature at time step t as follows:

$$e_t^k = \mathbf{v}_f^\top \tanh\Big(\mathbf{W}_f[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}] + \mathbf{U}_f \mathbf{x}^k + \mathbf{b}_f\Big) \tag{9}$$

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{j=1}^n \exp(e_t^j)} \tag{10}$$

where $\mathbf{h}_{t-1} \in \mathbb{R}^m$ and $\mathbf{s}_{t-1} \in \mathbb{R}^m$ represent the hidden state and cell state of the encoder LSTM, respectively. The learnable parameters include $\mathbf{v}_f \in \mathbb{R}^m$, $\mathbf{W}_f \in \mathbb{R}^{m\times 2m}$, $\mathbf{U}_f \in \mathbb{R}^{m\times T}$, and $\mathbf{b}_f \in \mathbb{R}^m$. The adaptively weighted input vector $\tilde{\mathbf{x}}_t$ is then constructed as:

$$\tilde{\mathbf{x}}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \dots, \alpha_t^n x_t^n)^\top \tag{11}$$

The encoder hidden state is updated by the LSTM unit:

$$\mathbf{h}_t = f_{LSTM}(\mathbf{h}_{t-1}, \tilde{\mathbf{x}}_t) \tag{12}$$

(2) Temporal Attention in the Decoder To address the performance degradation in long-term dependencies, a temporal attention mechanism is integrated into the decoder to select critical encoder hidden states. For the i-th encoder hidden state $\mathbf{h}_i$, the attention score l_t^i at decoding step t is:

$$l_t^i = \mathbf{v}_d^\top \tanh\left(\mathbf{W}_d[\mathbf{d}_{t-1}; \mathbf{s}'_{t-1}] + \mathbf{U}_d\mathbf{h}_i + \mathbf{b}_d\right) \tag{13}$$

$$\gamma_t^i = \frac{\exp(l_t^i)}{\sum_{j=1}^T \exp(l_t^j)} \tag{14}$$

where $\mathbf{d}_{t-1}$ and $\mathbf{s}'_{t-1} \in \mathbb{R}^p$ are the decoder's hidden and cell states. The context vector $\mathbf{c}_t$, which summarizes the relevant historical information, is computed as the weighted sum:

$$\mathbf{c}_t = \sum_{i=1}^T \gamma_t^i \mathbf{h}_i \tag{15}$$

The decoder hidden state is updated by concatenating the context vector with the target feature y_{t-1}:

$$\mathbf{d}_t = f_{LSTM}(\mathbf{d}_{t-1}, [\hat{y}_{t-1}; \mathbf{c}_t]) \tag{16}$$

Finally, the predicted value $\hat{y}_t$ is obtained through a linear transformation:

$$\hat{y}_t = \mathbf{w}_y^\top [\mathbf{d}_t; \mathbf{c}_t] + b_y \tag{17}$$

By coupling feature-level attention to filter cross-factor noise and temporal attention to highlight evolutionary patterns, the DA-LSTM framework enables the joint learning of structural interactions and temporal dependencies across the two subsystems.

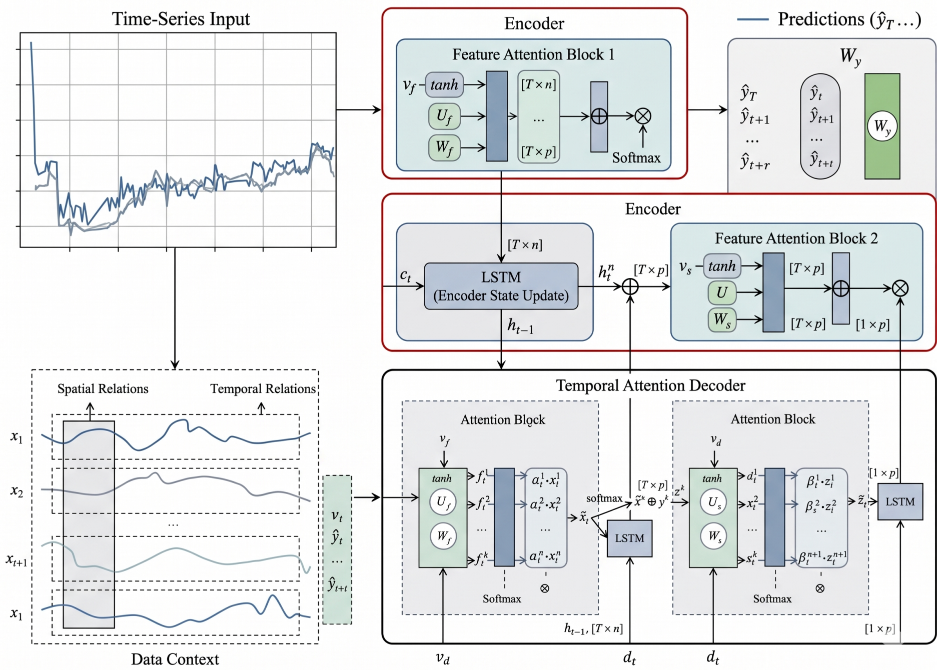


Figure 2. The proposed DA-LSTM architecture. The encoder utilizes a feature attention mechanism to adaptively select relevant exogenous inputs, while the decoder employs temporal attention to capture long-range dependencies across historical hidden states.

4. Results and Discussion

4.1. Sentiment Classification of News Headlines Based on RoBERTa

4.1.1. Dataset Construction

To capture the information signals embedded in financial news, a sentiment classification dataset was constructed. A public Chinese financial sentiment dataset from HuggingFace was initially employed, comprising 2329 financial news headlines labeled into three sentiment categories: Negative (0), Neutral (1), and Positive (2). Representative examples are listed in Table 3.

To construct the final input sequence for the Dual-attention LSTM, we chronologically aligned the daily aggregated topic sentiment scores with the historical trading data of soybean futures (Open, High, Low, Close, Volume, Open Interest, and Settlement Price) using the trading date as the primary key. The market and sentiment arrays were merged into a unified daily feature vector, and we applied *MinMaxScaler* to normalize all features into a $[0, 1]$ range, which stabilizes gradient updates and accelerates model convergence.

Table 3. Examples from the Public Chinese Financial Sentiment Dataset.

News Headline (Translated)	Label
Rural revitalization strategy series policies to be launched; new agricultural entities gain development opportunities	2
Cryptocurrency affects US forestry business; perceived threat	0
Bitcoin price struggles below $10,000 today	0
Yen net short positions hit four-month low; USD/JPY consolidates at low levels	1
Jinchuang Group hits limit up the day after listing	2

Although the public dataset provides general financial sentiment information, the soybean futures market represents a specialized subsystem within the broader financial market. To improve the domain adaptability of the sentiment model, an additional 1000 news headlines were randomly sampled from the collected soybean futures news corpus and manually annotated by a three-member team. Discrepancies in annotation were resolved through consensus discussion to ensure consistency and reliability. Examples of the annotated soybean futures headlines are shown in Table 4.

Table 4. Examples of Manually Annotated Soybean Futures News Headlines.

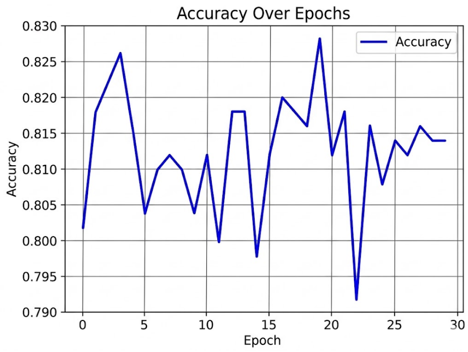
News Headline (Translated)	Label
Today's soybean futures price inquiry (8 September 2022)	1
Ministry of Agriculture: Increase investment and strengthen grain production capacity to secure national food safety	2
Soybean futures prices fell slightly today (6 November 2015)	0
Short covering pushes up bean futures	2
General Administration of Customs: January–October cumulative soybean imports hit 82.415 million tons, up 14.6% YoY	2

The final dataset contains 3329 labeled headlines, which were divided into an 80% training set and a 20% validation set for fine-tuning the RoBERTa model. This process allows the sentiment classifier to better capture domain-specific linguistic patterns and information signals associated with agricultural commodity markets.

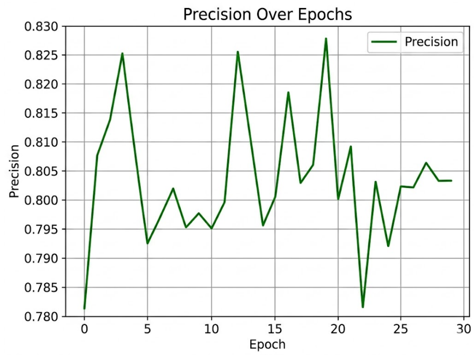
4.1.2. Impact of Sentiment Factors on Price Prediction

To assess the contribution of sentiment factors derived from news headlines, an ablation analysis was conducted by comparing models trained with and without sentiment-based features. The objective was to evaluate whether sentiment information extracted from news headlines provides additional predictive signals beyond traditional transaction data.

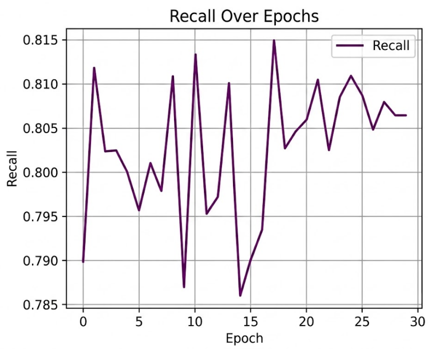
The training and validation performance of the fine-tuned RoBERTa sentiment classification model over the epochs is illustrated in Figure 3. As depicted by the learning curves, the model demonstrates stable and rapid convergence without signs of severe overfitting. Quantitatively, upon completing the training process, the model achieves a highly robust performance on the validation set, recording an Accuracy of 0.8111, a Precision of 0.7934, a Recall of 0.8081, and an F1 Score of 0.7988. These specific metrics indicate that the model is capable of reliably capturing sentiment polarity from complex financial news headlines, thereby providing a rigorous and robust basis for constructing the sentiment indicators used in the subsequent forecasting framework.



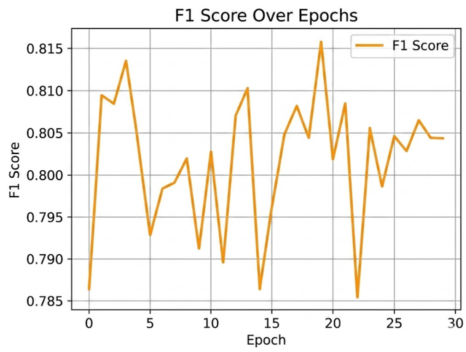
(**a**) Accuracy



(**b**) Precision



(**c**) Recall



(**d**) F1 Score

Figure 3. Training and validation trends of the fine-tuned RoBERTa sentiment classification model across epochs. Subfigures (**a**–**d**) illustrate the stable convergence of Accuracy, Precision, Recall, and F1 Score, respectively. The model ultimately achieves a high quantitative performance (e.g., an F1 Score of 0.7988), confirming the reliability of the extracted sentiment features for downstream tasks.

To provide a more comprehensive evaluation of the model's internal behavior, a confusion matrix was generated based on the dataset (Figure 4). The matrix reveals that the model demonstrates exceptional reliability in identifying Neutral and Negative news. Interestingly, the primary source of misclassification occurs between Positive and Neutral samples. This phenomenon is highly characteristic of financial texts, as favorable market developments are frequently articulated using objective and factual language, which the model occasionally interprets as neutral. Most importantly, extreme misclassifications (i.e., predicting a completely opposite polarity, such as Negative to Positive) are exceedingly rare. This confirms that the RoBERTa model effectively and safely captures the directional polarity of market sentiment without introducing systemic directional bias.

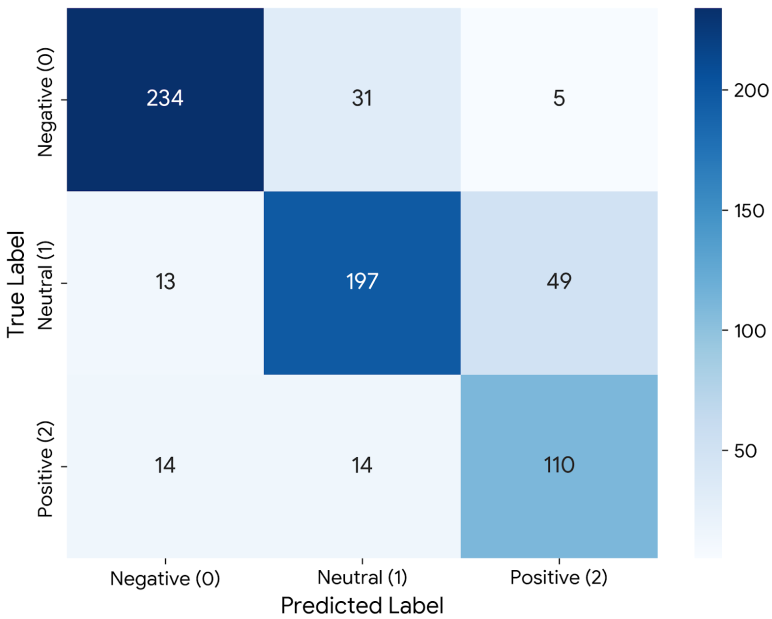


Figure 4. Confusion matrix of the RoBERTa-based sentiment classification model.

From a systems perspective, futures price dynamics should not be viewed as the result of an isolated numerical time series but rather as the outcome of a complex information system. Market prices evolve through the interaction of multiple information flows, including trading activities, macroeconomic signals, policy announcements, and media-reported events. Among these components, news sentiment acts as an important channel through which market expectations and external shocks are transmitted.

First, sentiment features extracted from news headlines provide timely signals related to policy developments, global trade conditions, and agricultural market dynamics. These signals often precede observable changes in trading data and therefore supply early information about potential market movements. Second, the impact of news information on market prices is typically subject to delayed absorption by market participants. As investors interpret and react to incoming information, the influence of news sentiment gradually propagates through the market, introducing additional temporal dependencies into the forecasting process.

Consequently, incorporating sentiment-based indicators into the forecasting framework enables the model to capture both the informational and temporal characteristics of market dynamics. This integration reflects a system-level perspective in which heterogeneous data sources jointly contribute to the formation of futures prices.

In addtion, to empirically validate this premise within our specific context, we conducted preliminary Granger causality tests on our integrated dataset. The results confirm that specific news-derived sentiment topics (e.g., Topic 2, Topic 4, and Topic 5) exhibit statistically significant Granger causality on soybean futures price returns ($p < 0.05$). While traditional formulations such as Granger causality and transfer entropy are powerful explanatory tools for quantifying pairwise or low-dimensional information flows, they generally rely on linear assumptions or fixed temporal windows, and struggle to serve as robust high-dimensional predictive architectures. The proposed DA-LSTM advances beyond these classical formulations by leveraging a dual-attention mechanism. Instead of merely measuring statistical causality, DA-LSTM handles highly non-linear dynamics and dynamically learns the shifting importance of different information subsystems (via feature attention) and critical time steps (via temporal attention). This enables the model to effectively translate complex, systemic information interactions into an adaptive and robust predictive framework.

4.2. Soybean Futures Price Level Prediction Based on News Headlines and Transaction Data

Financial markets are complex adaptive systems in which multiple heterogeneous factors interact dynamically. In this study, the `pivot_table` function from the Pandas library was used to aggregate sentiment scores derived from news headlines. These scores were then combined with transaction indicators to construct a unified dataset integrating multi-source heterogeneous information.

The integrated dataset captures both endogenous market variables (e.g., price, trading volume, and open interest) and exogenous information signals (e.g., sentiment derived from news). The dataset was subsequently divided into training, validation, and testing subsets in a ratio of 7:1:2 randomly.

Prior to model training, hyperparameter optimization was conducted for the Dual-Attention LSTM model. To ensure optimal and stable convergence of the proposed DA-LSTM framework, a systematic hyperparameter tuning process was conducted using a grid search strategy based on the validation set performance. The optimization strategy employed the Adam optimizer to iteratively minimize the Mean Squared Error (MSE) loss function. The search space for the primary hyperparameters was defined as follows: the learning rate was tuned over the set $\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$, the batch size was evaluated across $\{16, 32, 64, 128\}$, and the hidden state dimensions for both the encoder and decoder were tested within $\{32, 64, 128, 256\}$. Based on the minimum validation loss, the optimal hyperparameter configuration was established with a learning rate of 0.0001, a batch size of 64, and a hidden state dimension of 128 for both the encoder and decoder networks. Training was performed for up to 500 epochs with an early stopping mechanism to avoid overfitting. The optimal configuration—64 hidden neurons and a learning rate of 0.001—achieved MAE $= 0.0170$, RMSE $= 0.0216$, and $R^2 = 0.9950$.

To ensure a fair and standardized comparison, all trainable models (including MLP, CNN, LSTM, Transformer, and our DA-LSTM) were subjected to a rigorous hyperparameter optimization process. We utilized a grid search strategy over a predefined search space for key parameters—including the number of hidden units, layer depth, and dropout rates—to ensure that each model achieved its peak performance on the validation set. Furthermore, all experiments were executed under a unified framework using the Adam optimizer and a consistent early-stopping mechanism to eliminate bias arising from non-optimal settings or inconsistent training protocols. Note that the Naive Persistence baseline and the TimesFM foundation model required no training optimization, as they employ a deterministic heuristic and zero-shot inference, respectively. The search ranges and the resulting optimal configurations derived from these experiments are summarized in Table 5.

Using these configurations, the Dual-Attention LSTM model was trained on the dataset containing sentiment influence factors and compared against a comprehensively upgraded suite of baseline models. To rigorously evaluate the state-of-the-art validity of our framework, we expanded the comparator set to include a fundamental heuristic (Naive Persistence), a contemporary sequence architecture (Transformer), and a modern time-series foundation model evaluated in zero-shot mode (TimesFM). To further ensure that our model's superiority is statistically robust and not due to chance, we conducted 10 independent trials and performed formal statistical significance testing ($p < 0.05$ across all major forecasting horizons) using the Diebold-Mariano (DM) test with Heteroskedasticity and Autocorrelation Consistent (HAC) adjustments. The comparative results are summarized in Table 6.

Table 5. Hyperparameter search space and optimal configurations for all models.

Model	Hyperparameters	Search Range	Optimal Value
Naive	Training parameters	N/A (Deterministic)	N/A
TimesFM	Training parameters	N/A (Zero-shot)	N/A
Transformer	Hidden units (d_{model})	{16, 32, 64, 128}	32
	Number of heads	{2, 4, 8}	4
	Number of layers	{1, 2, 3}	2
MLP	Hidden units	{64, 128, 256, 512}	256
	Hidden layers	{1, 2, 3}	1
	Dropout rate	{0.1, 0.2, 0.3, 0.5}	0.1
CNN	Filters (Hidden size)	{16, 32, 64, 128}	64
	Kernel size	{3, 5, 7}	3
	Conv layers	{1, 2, 3}	2
LSTM	Hidden units	{32, 64, 128, 256}	128
	Number of layers	{1, 2, 3}	1
	Dropout rate	{0, 0.1, 0.2, 0.5}	0.1
DA-LSTM	Learning rate	$\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$	10^{-3}
	Batch size	{16, 32, 64, 128}	64
	Hidden units (m, p)	{32, 64, 128, 256}	64

The extensive comparative analysis reveals several critical insights regarding cross-domain information processing and model architecture. Firstly, the inclusion of the Naive Persistence baseline establishes the fundamentally autoregressive nature of short-term high-frequency data, where simple persistence inherently yields an R^2 above 0.99 at horizon $H = 1$. Against this rigorous baseline, only the DA-LSTM models demonstrate a definitive, statistically significant reduction in Mean Absolute Error (MAE), proving genuine predictive capabilities beyond mere lag-copying.

Secondly, the evaluation of modern sequence baselines strongly validates our core argument regarding heterogeneous noise. When sentiment features are introduced to conventional networks, they suffer significant performance degradation. For instance, at the sliding window of 30 and prediction step of 7, the R^2 of the MLP drops abruptly from 0.8783 to 0.6559 (a decline of $\sim$25%), and the LSTM drops from 0.8597 to 0.8070. Even more remarkably, for the state-of-the-art Transformer baseline, incorporating sentiment features ('w/ Sent') actually increased the prediction error compared to using price data alone ('w/o Sent'). For example, at Window 10, Step 1, the MAE rose from 0.1853 to 0.2072. This phenomenon perfectly illustrates a critical limitation of standard architectures: when confronted with highly noisy, multidimensional sentiment features, standard self-attention mechanisms suffer from "attention distraction," treating valuable textual sentiment as disruptive noise.

Furthermore, while time-series foundation models like TimesFM (Zero-Shot) exhibit powerful capabilities in extrapolating univariate numerical trends (achieving a short-term MAE of 0.0658 at $H = 1$), they possess a fundamental architectural limitation: they cannot natively ingest or dynamically align multidimensional, high-noise exogenous variables such as our 9-dimensional sentiment topics. Consequently, when sudden market volatility is driven by external news shocks rather than historical momentum, these foundation models fail drastically, as evidenced by TimesFM's error escalating to an MAE of 1.4330 at $H = 30$.

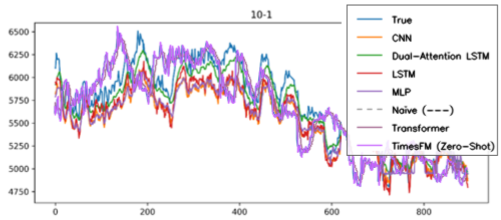
Table 6. Performance Comparison with Modern Baselines and Sentiment Features (MAE, RMSE, and R^2, reported as Mean $\pm$ Std).

Sliding Window Size/Prediction Step (MAE)							
Model	**Features**	**10**			**30**		
		1	**7**	**30**	**1**	**7**	**30**
Naive	Persistence	0.0451 ± 0.0000	0.1453 ± 0.0000	0.3278 ± 0.0000	0.0449 ± 0.0000	0.1456 ± 0.0000	0.3283 ± 0.0000
TimesFM	Zero-Shot	0.0667 ± 0.0000	0.3412 ± 0.0000	1.4092 ± 0.0000	0.0658 ± 0.0000	0.3453 ± 0.0000	1.4330 ± 0.0000
Transformer	w/o Sent.	0.1853 ± 0.0031	0.3842 ± 0.0045	0.5891 ± 0.0048	0.2215 ± 0.0029	0.4856 ± 0.0036	0.6412 ± 0.0051
	w/ Sent.	0.2072 ± 0.0035	0.4071 ± 0.0041	0.6140 ± 0.0052	0.2459 ± 0.0028	0.5101 ± 0.0048	0.6736 ± 0.0061
MLP	w/o Sent.	0.0393 ± 0.0026	0.0668 ± 0.0011	0.1226 ± 0.0017	0.0497 ± 0.0016	0.0743 ± 0.0028	0.1239 ± 0.0027
	w/ Sent.	0.1139 ± 0.0032	0.1481 ± 0.0012	0.1356 ± 0.0021	0.1178 ± 0.0011	0.1310 ± 0.0015	0.1457 ± 0.0023
CNN	w/o Sent.	0.0394 ± 0.0011	0.0674 ± 0.0015	0.1225 ± 0.0026	0.0493 ± 0.0024	0.0755 ± 0.0016	0.1226 ± 0.0025
	w/ Sent.	0.1041 ± 0.0030	0.1093 ± 0.0010	0.1386 ± 0.0030	0.0904 ± 0.0027	0.1127 ± 0.0019	0.1316 ± 0.0014
LSTM	w/o Sent.	0.0331 ± 0.0034	0.0628 ± 0.0018	0.1249 ± 0.0012	0.0339 ± 0.0012	0.0785 ± 0.0031	0.1238 ± 0.0025
	w/ Sent.	0.1005 ± 0.0030	0.0943 ± 0.0028	0.1521 ± 0.0023	0.0465 ± 0.0034	0.0896 ± 0.0019	0.1488 ± 0.0024
DA-LSTM	w/o Sent.	0.0154 ± 0.0031	0.0648 ± 0.0025	0.1165 ± 0.0032	0.0256 ± 0.0024	0.0566 ± 0.0028	0.1146 ± 0.0011
	w/ Sent.	0.0200 ± 0.0016	0.0780 ± 0.0017	0.1157 ± 0.0012	0.0191 ± 0.0016	0.0608 ± 0.0013	0.1089 ± 0.0017
Sliding Window Size/Prediction Step (RMSE)							
Model	**Features**	**10**			**30**		
		1	**7**	**30**	**1**	**7**	**30**
Naive	Persistence	0.0684 ± 0.0000	0.1917 ± 0.0000	0.4026 ± 0.0000	0.0683 ± 0.0000	0.1920 ± 0.0000	0.4032 ± 0.0000
TimesFM	Zero-Shot	0.0894 ± 0.0000	0.4348 ± 0.0000	1.7805 ± 0.0000	0.0891 ± 0.0000	0.4352 ± 0.0000	1.8264 ± 0.0000
Transformer	w/o Sent.	0.2520 ± 0.0039	0.4715 ± 0.0043	0.7204 ± 0.0048	0.2988 ± 0.0035	0.5902 ± 0.0042	0.7631 ± 0.0056
	w/ Sent.	0.2813 ± 0.0042	0.4994 ± 0.0051	0.7551 ± 0.0055	0.3199 ± 0.0039	0.6147 ± 0.0045	0.7973 ± 0.0062
MLP	w/o Sent.	0.0523 ± 0.0026	0.0859 ± 0.0019	0.1478 ± 0.0019	0.0570 ± 0.0015	0.0923 ± 0.0017	0.1478 ± 0.0033
	w/ Sent.	0.1314 ± 0.0026	0.1714 ± 0.0025	0.1652 ± 0.0014	0.1346 ± 0.0028	0.1595 ± 0.0014	0.1843 ± 0.0019
CNN	w/o Sent.	0.0526 ± 0.0035	0.0871 ± 0.0026	0.1471 ± 0.0024	0.0629 ± 0.0027	0.0950 ± 0.0031	0.1449 ± 0.0029
	w/ Sent.	0.1226 ± 0.0016	0.1368 ± 0.0011	0.1616 ± 0.0018	0.1059 ± 0.0017	0.1393 ± 0.0015	0.1562 ± 0.0034
LSTM	w/o Sent.	0.0441 ± 0.0032	0.0812 ± 0.0018	0.1525 ± 0.0026	0.0445 ± 0.0020	0.0991 ± 0.0033	0.1537 ± 0.0021
	w/ Sent.	0.1165 ± 0.0017	0.1169 ± 0.0016	0.1800 ± 0.0024	0.0616 ± 0.0017	0.1163 ± 0.0025	0.1762 ± 0.0032
DA-LSTM	w/o Sent.	0.0215 ± 0.0020	0.0835 ± 0.0015	0.1402 ± 0.0035	0.0342 ± 0.0023	0.0738 ± 0.0012	0.1393 ± 0.0011
	w/ Sent.	0.0286 ± 0.0013	0.0994 ± 0.0026	0.1379 ± 0.0030	0.0264 ± 0.0021	0.0818 ± 0.0012	0.1338 ± 0.0020
Sliding Window Size/Prediction Step (R^2)							
Model	**Features**	**10**			**30**		
		1	**7**	**30**	**1**	**7**	**30**
Naive	Persistence	0.9924 ± 0.0000	0.9406 ± 0.0000	0.7391 ± 0.0000	0.9925 ± 0.0000	0.9406 ± 0.0000	0.7394 ± 0.0000
TimesFM	Zero-Shot	0.9871 ± 0.0000	0.6943 ± 0.0000	−4.1041 ± 0.0000	0.9872 ± 0.0000	0.6951 ± 0.0000	−4.3478 ± 0.0000
Transformer	w/o Sent.	0.8971 ± 0.0038	0.6406 ± 0.0042	0.1645 ± 0.0051	0.8559 ± 0.0041	0.4392 ± 0.0045	0.0664 ± 0.0055
	w/ Sent.	0.8718 ± 0.0051	0.5967 ± 0.0049	0.0819 ± 0.0055	0.8348 ± 0.0043	0.3916 ± 0.0047	−0.0192 ± 0.0058
MLP	w/o Sent.	0.9580 ± 0.0065	0.8886 ± 0.0044	0.6898 ± 0.0064	0.9528 ± 0.0059	0.8783 ± 0.0021	0.7042 ± 0.0052
	w/ Sent.	0.7355 ± 0.0051	0.6026 ± 0.0044	0.6124 ± 0.0032	0.7373 ± 0.0049	0.6559 ± 0.0025	0.5401 ± 0.0040
CNN	w/o Sent.	0.9575 ± 0.0040	0.8854 ± 0.0063	0.6929 ± 0.0059	0.9425 ± 0.0032	0.8712 ± 0.0043	0.7157 ± 0.0028
	w/ Sent.	0.7822 ± 0.0061	0.7343 ± 0.0059	0.6461 ± 0.0033	0.8281 ± 0.0049	0.7361 ± 0.0047	0.6533 ± 0.0027
LSTM	w/o Sent.	0.9701 ± 0.0054	0.9003 ± 0.0044	0.6696 ± 0.0055	0.9712 ± 0.0044	0.8597 ± 0.0020	0.6798 ± 0.0035
	w/ Sent.	0.8153 ± 0.0021	0.8049 ± 0.0062	0.5401 ± 0.0060	0.9448 ± 0.0057	0.8070 ± 0.0034	0.5795 ± 0.0023
DA-LSTM	w/o Sent.	0.9917 ± 0.0060	0.8978 ± 0.0063	0.7206 ± 0.0024	0.9830 ± 0.0042	0.9222 ± 0.0023	0.7372 ± 0.0054
	w/ Sent.	0.9896 ± 0.0054	0.8506 ± 0.0026	0.7434 ± 0.0041	0.9904 ± 0.0045	0.9797 ± 0.0032	0.7573 ± 0.0059

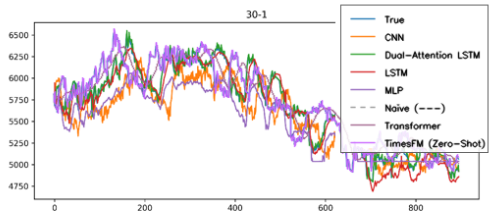
Note: The standard deviation for the Naive Persistence baseline and TimesFM zero-shot is exactly 0.0000, as they involve no stochastic training processes.

In stark contrast, our proposed DA-LSTM successfully digests these multi-source features. By elegantly coupling a Feature-Level Attention mechanism to filter text noise with a Temporal Attention mechanism, DA-LSTM prevents attention distraction. It effectively aligns the heterogeneous data, enabling the DA-LSTM to elevate its R^2 performance from an already strong 0.9222 to an exceptional 0.9797 in medium-term forecasting (Window 30, Step 7), while achieving an exceptional long-term MAE of 0.1089 where foundational models fail.

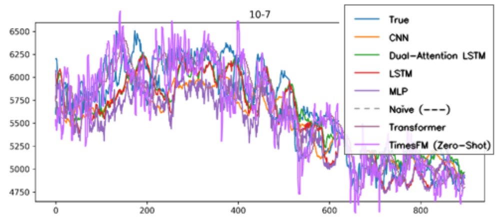
The prediction results on the test set are illustrated in Figure 5. In short-term forecasting scenarios, all models capture the general trend of soybean futures prices reasonably well. However, as the prediction horizon increases, baseline models gradually deviate from the observed price trajectory. In contrast, the Dual-Attention LSTM maintains a prediction curve closely aligned with the actual values, demonstrating stronger robustness and stability.



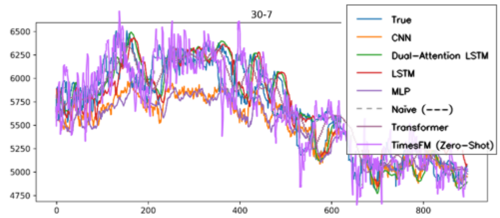
(**a**)



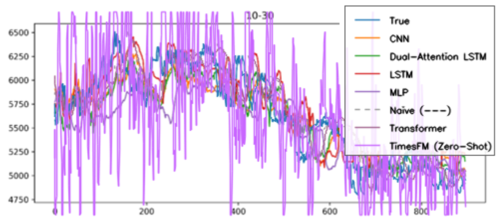
(**b**)



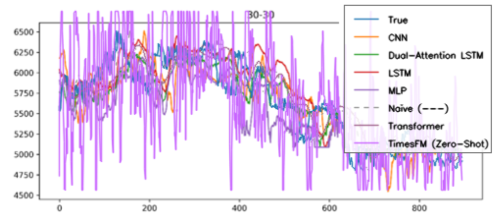
(**c**)



(**d**)



(**e**)



(**f**)

Figure 5. Comparison of prediction results on the test set with a sliding window size of 30 and a prediction step of 30.

From a systems perspective, the superior performance of the Dual-Attention LSTM can be attributed to its ability to model interactions among heterogeneous information sources. The temporal attention mechanism captures delayed dependencies between information signals and price fluctuations, while the feature attention mechanism dynamically identifies the most influential factors within the multi-source system.

Together, these mechanisms enable the model to better represent the information propagation process within the soybean futures market system, thereby improving forecasting accuracy under complex and volatile market conditions.

4.3. Model Interpretability: Attention Weight Analysis

While the performance metrics demonstrate the predictive superiority of the DA-LSTM, understanding the model's internal decision-making process is equally critical. By extracting the weights from the feature-level and temporal attention layers, we can interpret which variables and historical periods drive the market predictions. Figure 6 illustrates the distribution of these attention weights.

Feature-Level Attention: As shown in Figure 6a, the feature attention mechanism dynamically assigns varying degrees of importance to the heterogeneous inputs. Unsurprisingly, fundamental trading variables such as the closing price and highest price hold the most significant weights, reflecting the strong autoregressive nature of futures markets.

However, it is highly noteworthy that specific sentiment features—particularly **Topic_1** and **Topic_2**—also capture substantial attention weights, standing out among the multidimensional inputs. This visual evidence validates our core hypothesis: the DA-LSTM does not treat textual sentiment as mere noise; instead, it actively cross-references numerical trends with critical sentiment shifts to formulate its predictions.

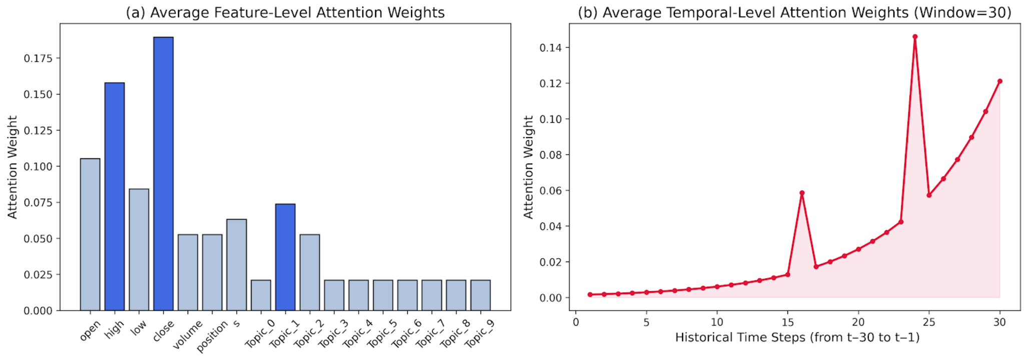


Figure 6. Average attention weight distributions of the DA-LSTM model. (**a**) Feature-level attention highlights the importance of core trading variables alongside key sentiment topics (Topic_1, Topic_2). (**b**) Temporal-level attention illustrates recency bias coupled with specific historical memory peaks.

Temporal-Level Attention: Figure 6b visualizes the temporal attention weights across the sliding window of 30 days. The distribution demonstrates a clear recency bias, where the most recent time steps ($t-1$, $t-2$) naturally exert the strongest influence on the short-term prediction. However, rather than decaying strictly linearly, the temporal weights exhibit distinct local peaks at further historical steps (e.g., around the 7-day lag). These peaks indicate that the temporal attention mechanism is capable of memorizing and heavily weighting specific historical periods characterized by weekly cyclical patterns or lingering news shocks, effectively bypassing temporal distance to extract relevant structural memory.

Together, these attention analyses confirm that the dual-attention architecture is not just a mathematical black box, but a highly interpretable framework capable of mimicking human-like analytical logic—weighing "what matters" (features) and "when it happened" (time) in a complex multi-source market environment.

5. Conclusions

This study investigates soybean futures price forecasting from a system-level perspective by integrating multi-source heterogeneous information, including historical trading data and news-derived textual information. Instead of treating futures price prediction as a purely numerical time-series problem, this work models the market as a complex information system in which price dynamics emerge from the interaction of multiple information flows such as trading activities, policy signals, and market sentiment.

In the sentiment feature extraction stage, a RoBERTa-based news headline sentiment classification model was developed to quantify market sentiment embedded in financial news. The model achieved an Accuracy of 0.8111, a Precision of 0.7934, a Recall of 0.8081, and a F1-Score of 0.7988 on the validation set. While these figures suggest the model's reasonable capability in capturing sentiment polarity to provide indicators for the forecasting framework, we acknowledge that relying on validation-set metrics may overstate generalisation, and evaluating the classifier on a strictly held-out test set remains an important objective for future refinement.

In the forecasting stage, a Dual-Attention LSTM model was proposed to capture both temporal dependencies and inter-feature relationships within the integrated dataset. Experimental results show that the proposed model consistently outperforms baseline methods across multiple prediction horizons. In particular, for the medium-term forecasting scenario with a sliding window of 30 and a prediction horizon of 7 steps, the R^2 value improved from 0.922 to 0.9797. This result indicates that incorporating sentiment-based information can effectively aid in capturing medium- and long-term market dynamics. The dual-attention mechanism further improves predictive performance by adaptively emphasizing important features and critical time steps within the complex information system.

Overall, the results demonstrate that integrating heterogeneous information sources can offer observable improvements in short-term forecasting accuracy for soybean futures compared to standard baseline models. By modeling the futures market as an interconnected information system rather than an isolated numerical process, the proposed framework provides a more comprehensive approach to understanding and predicting market dynamics. The proposed method can offer effective decision-support tools for participants in agricultural futures markets and contributes to the development of system-oriented forecasting methodologies in financial markets.

Future research can further extend this work by incorporating additional information sources, such as macroeconomic indicators, weather data, and international trade signals, to construct a more comprehensive agricultural market information system. In addition, future studies may explore more advanced deep learning architectures or hybrid mechanism–data-driven models to further improve the interpretability and generalization capability of futures price forecasting models.

Author Contributions: H.L.: Conceptualization, Methodology, Software, Formal analysis, Writing—Original draft. Q.L.: Conceptualization, Methodology, Software, Validation, Writing—Review & Editing, Supervision. Y.H.: Data curation, Investigation, Writing—Review & Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Humanity and Social Science Foundation of Ministry of Education of China (No. 21YJA630054), the Zhejiang Province Soft Science Research Program Project (No. 2025C35098), and the Open Access Publication Funds/transformative agreements of the Göttingen University.

Data Availability Statement: The datasets used in this study were collected from publicly available financial information platforms and commercial databases. News data related to soybean futures were crawled from Jintou Futures (https://m.cngold.org/futures.html (accessed on 1 March 2024)), Hexun Futures (https://m.hexun.com/futures.html(accessed on 1 March 2024)), the agricultural futures information module of the Choice Financial Terminal provided by Eastmoney (https://choice.eastmoney.com/(accessed on 1 March 2024)), the official website of the China Soybean Industry Association (http://www.chinasoybean.org.cn/(accessed on 1 March 2024)), and Sina Finance (https://finance.sina.com.cn/(accessed on 1 March 2024)). Soybean futures trading data were obtained from the soybean futures module of Sina Finance (https://finance.sina.com.cn/futuremarket/(accessed on 1 March 2024)).

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

References

1. An, W.Y.; Wang, L.; Zeng, Y.R. Social media-based multi-modal ensemble framework for forecasting soybean futures price. *Comput. Electron. Agric.* **2024**, *226* 109439. [CrossRef]
2. Khalil, R. AI driven sentiment analysis in financial markets: Using transformer base models and social media signals for stock market predictions. *J. Model. Manag.* **2026**, 1–25. [CrossRef]
3. Çınar, U.K.; Gölbaşı Şimşek, G. Analysis of Global Financial Connections and Information Flow Dynamics Using Transfer Entropy and Independent Component Analysis. *J. Risk Financ. Manag.* **2026**, *19*, 314. [CrossRef]
4. Agyapong, J. Macro-Financial Shocks and Media Coverage of Commodity Markets: A Time-Varying Spillover Analysis. *Stud. Nonlinear Dyn. Econom.* **2026**. [CrossRef]

5. Calheiros, R.N.; Masoumi, E.; Ranjan, R.; Buyya, R. Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications' QoS. *IEEE Trans. Cloud Comput.* **2015**, *3*, 449–458. [CrossRef]
6. Zhao, J.H.; Huang, Y.J.; Feng, J.; Xie, W.Y.; Jain, K. Fusion of KANO theory and Attention-BiLSTM models for user demand analysis and trend prediction. *Inf. Fusion* **2025**, *122*, 103210. [CrossRef]
7. Wu, C.F.; Lin, M.C.; Chao, T.W.; Chiou, C.C. Omnipresent AI and Big data for financial early warning: Integrating financial indicators and text sentiment analysis in Chinese real estate. *Enterp. Inf. Syst.* **2024**, *19*, 2434742. [CrossRef]
8. Qin, Y.; Song, D.; Chen, H.; Cheng, W.; Jiang, G.; Cottrell, G. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. *arXiv* **2017**, arXiv:1704.02971. [CrossRef]
9. Ahmed, A.; Huang, D.G.; Arafat, S.Y.; Rashid, K.I. Urdu Word Sense Disambiguation: Leveraging Contextual Stacked Embedding, Siamese Transformer Encoder 1DCNN-BiLSTM, and Gloss Data Augmentation. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2025**, *24*, 1–36. [CrossRef]
10. Geeganage, D.K.; Xu, Y.; Li, Y.F. A Semantics-enhanced Topic Modelling Technique: Semantic-LDA. *ACM Trans. Knowl. Discov. Data* **2024**, *18*, 1–27. [CrossRef]
11. Gardazi, N.M.; Daud, A.; Malik, M.K.; Bukhari, A.; Alsahfi, T.; Alshemaimri, B. BERT applications in natural language processing: A review. *Artif. Intell. Rev.* **2025**, *58*, 166. [CrossRef]
12. Hill, A.; Goo, K.; Agarwal, P. Recommending the right academic programs: An interest mining approach using BERTopic. *Data Min. Knowl. Discov.* **2025**, *39*, 20. [CrossRef]
13. Boughareb, D.; Boughareb, R.; Hallaci, S.; Boukherouba, M.R.E.; Seridi, H. Addressing the Complexity of Dialectal Arabic: An Enhanced Encoder-Decoder Ensemble Approach for Optimized Sentiment Analysis. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2025**, *24*, 1–21. [CrossRef]
14. Cai, L.; Li, H.B.; Zhang, N.; He, J. Bimodal Sentiment Analysis Based on a Pre-Trained Model and Masked Attention Fusion. *IEEE Trans. Audio Speech Lang. Process.* **2025**, *33*, 2139–2150. [CrossRef]
15. Xu, S.T. DEEP-CWS: Distilling Efficient pre-trained models with Early exit and Pruning for scalable Chinese Word Segmentation. *Inf. Sci.* **2025**, *719*, 122470. [CrossRef]
16. Mu, Z.; Wan, Y.; Zhuang, Y.; Tan, J.; Cheng, H.; Wang, Y. Exploring financial sentiment analysis via fine-tuning large language model and attributed graph neural network. *Neural Netw.* **2026**, *199*, 108620. [CrossRef] [PubMed]
17. Toksoz, S.B.; Isik, G. Parameter-efficient fine-tuning of LLaMA models for financial sentiment classification. *Clust. Comput.* **2025**, *29*, 41. [CrossRef]
18. Cheng, Y.; Liu, Y.; Zhou, H. Large Language Models and Futures Price Factors in China. *J. Futur. Mark.* **2025**, *46*, 262–282. [CrossRef]
19. Ausloos, M.; Zhang, Y.N.; Dhesi, G. Stock index futures trading impact on spot price volatility. The CSI 300 studied with a TGARCH model. *Expert Syst. Appl.* **2020**, *160*, 113688. [CrossRef]
20. Huang, K.K.; Cao, W.P.; Liu, Y.S.; Wu, D.H.; Yang, C.H.; Gui, W.H. A Weighted Deep Learning-Based Predictive Control for Multimode Nonlinear System With Industrial Applications. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 10814–10826. [CrossRef]
21. Cheng, Z.Y.; Tian, X.Y.; Sui, R.M.; Li, T.M.; Jiang, Y. An Adaptive Grasping Force Tracking Strategy for Nonlinear and Time-Varying Object Behaviors. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 17063–17076. [CrossRef]
22. Gu, Q.H.; Chang, Y.X.; Xiong, N.X.; Chen, L. Forecasting Nickel futures price based on the empirical wavelet transform and gradient boosting decision trees. *Appl. Soft Comput.* **2021**, *109*, 107472. [CrossRef]
23. Hajiabotorabi, Z.; Samavati, F.F.; Ghaini, F.M.M.; Shahmoradi, A. Multi-WRNN model for pricing the crude oil futures market. *Expert Syst. Appl.* **2021**, *182*, 115229. [CrossRef]
24. Hu, L.J.; Wang, X.H.; Liu, Y.X.; Liu, N.H.; Huai, M.D.; Sun, L.C.; Wang, D. Towards Stable and Explainable Attention Mechanisms. *IEEE Trans. Knowl. Data Eng.* **2025**, *37*, 3047–3061. [CrossRef]
25. Sun, S.G.; Liu, J.F.; Wang, J.Q.; Chen, F.; Wei, S.; Gao, H. Remaining Useful Life Prediction for AC Contactor Based on MMPE and LSTM with Dual Attention Mechanism. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 9004513. [CrossRef]
26. von Bloh, M.; Noia, R.D., Jr.; Wangerpohl, X.; Saltik, A.O.; Haller, V.; Kaiser, L.; Asseng, S. Machine learning for soybean yield forecasting in Brazil. *Agric. For. Meteorol.* **2023**, *341*, 109670. [CrossRef]
27. Mou, S.; Xue, Q.; Chen, J.; Takiguchi, T.; Ariki, Y. MM-iTransformer: A Multimodal Approach to Economic Time Series Forecasting with Textual Data. *Appl. Sci.* **2025**, *15*, 1241. [CrossRef]
28. Karadaş, F.; Eravcı, B.; Özbayoğlu, A. Multimodal Stock Price Prediction. In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence*; SCITEPRESS—Science and Technology Publications: Setúbal, Portugal , 2025; pp. 687–694. [CrossRef]
29. Nguyen, N.H.; Nguyen, T.T.; Ngo, Q.T. DASF-Net: A Multimodal Framework for Stock Price Forecasting with Diffusion-Based Graph Learning and Optimized Sentiment Fusion. *J. Risk Financ. Manag.* **2025**, *18*, 417. [CrossRef]
30. Kim, D.J.; Noh, E.; Choi, S.Y. A hybrid transformer framework integrating sentiment and dynamic market structure for stock price movement forecasting. *AIMS Math.* **2026**, *11*, 977–1020. [CrossRef]

31. Ansari, A.F.; Stella, L.; Turkmen, C.; Zhang, X.; Mercado, P.; Shen, H.; Shchur, O.; Rangapuram, S.S.; Arango, S.P.; Kapoor, S.; et al. Chronos: Learning the Language of Time Series. *arXiv* **2024**, arXiv:2403.07815. [CrossRef]
32. Das, A.; Kong, W.; Sen, R.; Zhou, Y. A decoder-only foundation model for time-series forecasting. *arXiv* **2023**, arXiv:2310.10688. [CrossRef]
33. Liu, J.; Xie, S.; Wang, J.; Wei, Y.; Ding, Y.; Zhang, L. Evaluating Language Models for Efficient Code Generation. *arXiv* **2024**, arXiv:2408.06450. [CrossRef]
34. Woo, G.; Liu, C.; Kumar, A.; Xiong, C.; Savarese, S.; Sahoo, D. Unified Training of Universal Time Series Forecasting Transformers. *arXiv* **2024**, arXiv:2402.02592. [CrossRef]
35. Rasul, K.; Ashok, A.; Williams, A.R.; Ghonia, H.; Bhagwatkar, R.; Khorasani, A.; Bayazi, M.J.D.; Adamopoulos, G.; Riachi, R.; Hassen, N.; et al. Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting. *arXiv* **2023**, arXiv:2310.08278. [CrossRef]
36. Wang, H.J.; Jiao, J.J. Sentiment Analysis of MOOC Reviews Based on Knowledge Dependency Tree. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2025**, *24*, 1–15. [CrossRef]
37. Wu, S.; Wang, W.; Song, Y.N.; Liu, S.Q. An EEMD-LSTM, SVR, and BP decomposition ensemble model for steel future prices forecasting. *Expert Syst.* **2024**, *41*, e13672. [CrossRef]