

Article

A Consistency-Aware Multimodal Sensing Framework for Short-Term Cross-Border Prediction Under Dynamic Trade

Yongyi Wan ^{1,2,†}, Xutong Wang ^{1,2,†}, Xinyue Zeng ^{1,†}, Shangshan Chen ^{1,2}, Beier Luo ^{1,2}, Weijing Yu ²
and Manzhou Li ^{1,2,*}

¹ China Agricultural University, Beijing 100083, China; 2025505610120@cau.edu.cn (Y.W.); wangxutong@cau.edu.cn (X.W.); 2500935059@stu.pku.edu.cn (X.Z.); chenshangshan@stu.pku.edu.cn (S.C.); luobeier@cau.edu.cn (B.L.)

² Peking University, Beijing 100871, China; 2600935009@stu.pku.edu.cn

* Correspondence: limanzhou_pku@163.com

† These authors contributed equally to this work.

Abstract

Short-term cross-border prediction is of significant importance for global supply-chain risk management, international trade decision-making, and stability. However, existing approaches mainly rely on individual price sequences or limited structured variables, making it difficult to effectively perceive complex external factors, including cross-border trade flows, logistics state variations, exchange-rate fluctuations, and policy-driven shocks. To address these challenges, an artificial intelligence-driven sensing-oriented reliability- and consistency-aware multimodal framework for short-term cross-border prediction is proposed. Market conditions, exchange rates, cross-border trade activities, logistics operations, and news-policy events are jointly modeled as multisource intelligent sensing signals. In the proposed framework, the reliability and cross-modal consistency perception module is first developed to dynamically evaluate the credibility of different data sources and suppress the interference caused by missing, delayed, and conflicting information. Subsequently, the asynchronous cross-border temporal interaction module is introduced to capture the time-dependent propagation relationships among trade, logistics, exchange rates, and states. Furthermore, the dynamic trade and policy event perception module is constructed to identify short-term disturbances induced by tariff adjustments, trade restrictions, port disruptions, and major international events, thereby enabling intelligent prediction under complex cross-border environments. Based on a multisource cross-border sensing dataset constructed from January 2022 to December 2025, the performance of the proposed framework is systematically evaluated through three tasks, including short-term direction prediction, volatility forecasting, and risk level prediction. Experimental results demonstrate that the proposed method achieves an accuracy of 0.842, precision of 0.836, recall of 0.829, macro-F1 of 0.832, and AUC of 0.913 in the short-term direction prediction task, significantly outperforming ARIMA, XGBoost, LSTM, TCN, Transformer, PatchTST, and existing multimodal fusion models. Ablation studies further verify the critical contributions of reliability modeling, consistency constraints, asynchronous temporal interaction, and dynamic event perception modules to improving prediction performance.

Keywords: artificial intelligence-driven sensing; multimodal sensing framework; cross-border prediction; reliability-aware data fusion; cross-modal consistency learning; asynchronous temporal interaction



Academic Editor: Jorn Mehnert

Received: 12 August 2026

Revised: 6 September 2026

Accepted: 11 September 2026

Published: 14 September 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.

1. Introduction

With the rapid development of digitalized global trade, cross-border e-commerce, international logistics, and commodity prices, exchange rates, and futures prices have become increasingly interconnected across countries and regions [1]. Dynamics are therefore jointly affected by trade flows, logistics conditions, exchange-rate fluctuations, policies, news, and unexpected events rather than by historical prices alone [2]. For example, port congestion or transportation disruptions in major exporting countries can alter supply expectations and subsequently propagate to commodity futures, exchange rates, and related asset prices [3]. Consequently, short-term cross-border prediction should be viewed as a multimodal sensing task that integrates states with trade, logistics, exchange-rate, and external semantic information, providing support for risk monitoring, price forecasting, and supply-chain coordination [4,5]. Traditional prediction methods mainly rely on statistical models such as ARIMA and VAR to capture temporal autocorrelation and linear dependencies among historical variables [6]. Although interpretable, they are limited in modeling nonlinear relationships and external shocks [7]. Machine learning methods such as Random Forest and XGBoost improve nonlinear modeling capability but still depend heavily on engineered features and have difficulty capturing the temporal propagation of trade, logistics, and policy events [8]. More importantly, existing approaches remain dominated by internal variables, while external signals reflecting real supply–demand conditions are insufficiently exploited, leading to degraded responsiveness and prediction performance during abrupt policy changes, logistics disruptions, and severe fluctuations [9].

Recent years have witnessed the widespread application of deep learning models, including long short-term memory (LSTM), gated recurrent units (GRU), temporal convolutional networks (TCNs), Transformer, and PatchTST, to commodity forecasting [10]. Although these models effectively capture temporal dependencies, most studies still formulate prediction as a single-series forecasting problem and insufficiently exploit external signals from real environments [11]. To address this limitation, some studies have incorporated news, social media, and policy texts through sentiment analysis, pretrained language models, and cross-modal attention, while cross-border trade research has introduced customs records, trade orders, automatic identification system (AIS) trajectories, port operations, and policy information for supply-chain monitoring and risk identification [12]. However, these two research directions remain largely isolated, and existing methods rarely model the complete cross-border transmission process from trade and logistics changes to supply–demand adjustments and subsequent responses [13]. Meanwhile, heterogeneous multimodal fusion remains a fundamental challenge [14]. Trade statistics, logistics trajectories, and textual information differ substantially in temporal granularity, statistical characteristics, and semantic meaning [15,16]. Simple feature concatenation cannot effectively capture cross-modal dependencies and may introduce redundant information and spurious correlations [17]. Moreover, heterogeneous sources often contain missing observations, asynchronous updates, delayed responses, and conflicting signals [18]. For example, logistics activities or news reports may indicate supply risks earlier than prices, while trade statistics may lag behind actual conditions [19]. Therefore, reliability-aware multimodal fusion and asynchronous modeling of trade, logistics, exchange rates, policies, and dynamics remain key challenges for improving short-term cross-border prediction [20]. Recent studies have demonstrated the potential of multimodal semantic modeling for forecasting [21]. Representative frameworks integrate cross-lingual semantic alignment, temporal fusion, multimodal attention, or trend-aware feature fusion to enhance risk perception and prediction accuracy across multiple datasets [22]. Nevertheless, they mainly focus on semantic information fusion and still provide limited modeling of cross-border

trade propagation mechanisms, heterogeneous information reliability, and asynchronous cross-modal interactions.

To address the aforementioned challenges, a reliability- and consistency-aware multi-modal sensing framework for short-term cross-border prediction is proposed. The proposed framework integrates international conditions, exchange rates, cross-border trade flows, logistics trajectories, port operation states, and multilingual news and policy information into unified cross-border sensing signals, thereby establishing an end-to-end deep learning prediction architecture. Specifically, modality-specific encoders are first employed to extract market, exchange-rate, trade, logistics, and event representations. Subsequently, a reliability- and cross-modal consistency-aware mechanism is constructed to dynamically adjust the credibility of different information sources according to missing observations, update intervals, local noise, source stability, and inter-modal correspondence relationships. Based on these representations, asynchronous propagation relationships among trade, logistics, exchange rates, and responses are captured through continuous-time encoding and multi-scale lag-aware cross-modal attention. Furthermore, a dynamic trade and policy event perception module is developed to identify short-term shocks induced by tariff adjustments, trade restrictions, port disruptions, and major international events. Finally, a context-aware fusion network is employed to dynamically regulate the contributions of heterogeneous signals, enabling joint prediction of direction, short-term volatility, and cross-border risk levels. The main contributions of this study can be summarized as follows.

1. **Reformulation of the prediction paradigm.** The conventional cross-border prediction problem is reformulated as a multimodal sensing and fusion problem jointly driven by conditions, exchange rates, trade activities, logistics operations, and policy events. This reformulation overcomes the limitations of traditional single-modal temporal modeling approaches that mainly rely on historical prices and technical indicators, enabling prediction to incorporate real-world cross-border commodity flows and external environmental changes.
2. **Development of an end-to-end multimodal framework.** A reliability- and consistency-aware multimodal cross-border prediction framework is developed to achieve unified encoding and deep interaction among dynamics, exchange rates, trade states, logistics conditions, and unstructured news and policy information. Through reliability estimation and cross-modal consistency modeling, the negative impacts of missing data, statistical delays, and information conflicts on prediction performance are effectively reduced.
3. **Targeted solutions for cross-border business challenges.** A series of specialized modules, including cross-modal signal reliability and consistency perception, asynchronous cross-border temporal interaction, dynamic event impact perception, and context-aware adaptive fusion, are systematically designed to address key challenges including heterogeneous temporal granularity, delayed information propagation, abrupt event shocks, and dynamically changing modality contributions. In particular, multi-scale lag modeling enables the characterization of cross-border transmission mechanisms among policies, logistics, trade activities, and responses with different propagation speeds.
4. **Comprehensive validation of model effectiveness.** The evaluation reported in this manuscript comprises short-term direction prediction, volatility forecasting, risk classification, and controlled component ablations. These experiments directly test the three prediction heads and the contributions of reliability modeling, consistency constraints, asynchronous interaction, event perception, and adaptive fusion. We do not claim separate cross-country, missing-modality, or event-window experiments that are not explicitly reported in the Section 4.

2. Related Work

2.1. Short-Term Commodity Prediction

The objective of short-term and commodity prediction is to learn the mapping from current conditions to future prices, returns, volatility, or movements by exploiting historical temporal dependencies [23]. The field has evolved from statistical models to machine learning methods and, more recently, deep learning architectures [24]. Early approaches, such as ARIMA and VAR, modeled linear temporal relationships among historical prices, trading volumes, and exchange rates, offering strong interpretability but limited capability for capturing complex nonlinear dynamics. Machine learning models, including Random Forest and XGBoost, further improved forecasting performance by exploiting nonlinear interactions among structured variables such as prices, technical indicators, order books, and exchange rates [25]. Recent advances in deep learning, including LSTM, GRU, TCN, Transformer, Informer, and PatchTST, have substantially enhanced the modeling of long-range temporal dependencies and nonlinear dynamics [26,27]. Nevertheless, most existing methods still formulate prediction as an isolated time-series forecasting task, relying primarily on historical variables while insufficiently exploiting external leading signals such as cross-border trade activities, logistics operations, and policy changes [28]. Consequently, when conditions are reshaped by tariff adjustments, logistics disruptions, or international policy events, these models often exhibit delayed responses and reduced prediction robustness [29].

2.2. Cross-Border Trade, Logistics, and Multimodal Risk Sensing

Cross-border trade and logistics sensing aims to characterize real trade activities by integrating heterogeneous information from transactions, commodity flows, transportation processes, and policy environments, thereby supporting trade condition monitoring, anomaly detection, and supply-chain-risk identification [30]. Because cross-border trade involves substantial data heterogeneity and regulatory fragmentation, individual customs declarations or transaction records cannot fully reflect actual commodity flows [31]. Accordingly, existing studies have expanded from structured trade records to multimodal sources such as customs and shipping records, AIS/GPS trajectories, port operations, logistics delays, commodity images, and policy information, which are increasingly modeled using machine learning, deep learning, and graph-based methods [32]. Recent research has further explored multimodal fusion to jointly model complementary sensing sources [33,34]. Trade records describe nominal commodity flows, logistics trajectories and port states capture actual transportation conditions, while news and policies provide external contextual information [35]. However, existing studies primarily use these signals for fraud detection, abnormal transaction identification, and supply-chain-risk assessment [36], with limited attention to how trade and logistics changes propagate through supply–demand expectations, exchange rates, and cross channels to affect short-term commodity [37]. Therefore, a key research direction is to extend trade, logistics, and policy information from conventional risk detection signals to cross-border sensing signals and integrate them into short-term prediction frameworks.

2.3. Multimodal Fusion and Event-Driven Forecasting

Multimodal fusion improves prediction robustness by exploiting complementary information from heterogeneous sources and constructing unified representations through feature fusion, cross-modal attention, and multimodal Transformer architectures. In forecasting, textual information such as news and announcements has increasingly been integrated with time series to enhance event perception, while dynamic and reliability-aware fusion mechanisms further adapt modality contributions according to contextual condi-

tions and data quality. However, general-purpose multimodal fusion remains insufficient for cross-border prediction. First, trade data, AIS-based logistics, exchange rates, news, and prices are naturally sampled at different frequencies, making forced synchronization likely to distort their temporal relationships [38]. Second, heterogeneous sources frequently contain missing, delayed, noisy, or conflicting observations, requiring explicit evaluation of modality reliability and cross-source consistency before fusion [39]. More importantly, different signals affect markets through distinct propagation delays: policy and news events may trigger rapid responses, whereas logistics disruptions and trade-flow changes generally propagate through supply–demand expectations before influencing prices [40]. Existing methods still lack effective mechanisms for jointly modeling such asynchronous cross-modal interactions, information reliability, and event-driven propagation delays [20].

Therefore, beyond conventional multimodal fusion strategies, a reliability- and cross-modal consistency-aware representation mechanism, asynchronous temporal interaction module, and dynamic trade and policy event perception mechanism are introduced. A specialized cross-modal alignment, event impact modeling, and context-aware fusion architecture is constructed, enabling simultaneous integration, exchange-rate, trade, logistics, and news–policy information while evaluating source reliability, identifying cross-modal conflicts, and learning the temporal propagation pathways of cross-border events toward target. Consequently, a unified multimodal sensing framework tailored for short-term cross-border prediction is established.

3. Materials and Method

3.1. Data Collection

A multimodal heterogeneous sensing dataset tailored for short-term cross-border prediction was constructed by integrating multidimensional information sources, including international states, exchange-rate variations, cross-border trade activities, logistics operations, and news–policy events, as shown in Table 1. The data collection period covered January 2022 to December 2025, encompassing multiple major international trade and commodity categories, and diverse evolution periods. Quotation data were primarily collected from international data platforms, including major stock indices, commodity futures, and related asset transaction records. The collected variables included open price, high price, low price, close price, trading volume, return rate, historical volatility, bid–ask spread, and order-book imbalance indicators. These high-frequency trading characteristics were collected at minute-level intervals to ensure that short-term fluctuations and price movement patterns could be effectively captured. Exchange-rate and cross information were obtained from international foreign exchange databases and publicly available interfaces. The collected information mainly consisted of major currency exchange rates against the US dollar, currency return variations, exchange-rate volatility, overseas indices, and cross spreads, which were utilized to characterize the dynamic interactions across different countries and regions.

Cross-border trade data were collected from international trade statistical databases, public customs databases, and trade information service platforms. Import and export records from 2022 to 2025 were included, containing commodity categories, trade values, trade quantities, origin countries, destination countries, trade growth rates, and import–export imbalance indicators. Since trade statistics generally exhibit relatively low update frequencies, the original publication timestamps were preserved and directly incorporated into model training through continuous-time encoding rather than being artificially replicated into high-frequency intervals. Logistics and port sensing data were primarily collected from the AIS, global vessel trajectory databases, and port operation monitoring platforms. The collected information included vessel coordinates, navigation speed,

shipping trajectories, arrival time, departure time, estimated arrival time, port congestion level, container throughput, and transportation delays. AIS data were continuously obtained through vessel broadcasting signals, and port-level logistics states were further extracted through geographic region matching approaches to characterize dynamic variations in real commodity transportation processes. News and policy event data were collected from international news databases, government announcement platforms, and trade policy information websites. The collected events covered trade restrictions, tariff adjustments, economic sanctions, supply-chain disruptions, international conflicts, extreme weather events, and central bank policy announcements. Textual information was processed through multilingual semantic models for event identification, and high-level semantic attributes, including event type, associated countries, related commodities, event intensity, sentiment tendency, and uncertainty level, were further extracted.

Table 1. Statistics of the collected cross-border multimodal sensing dataset.

Data Source	Sensor/Data Platform	Data Quantity	Main Features
Market Data	Bloomberg/Exchange API	2,340,000 records	OHLC, volume, return, volatility, order-book
Foreign Exchange Data	International FX Database	1,280,000 records	Exchange rate, currency return, volatility, spread
Cross-Border Trade Data	Customs and Trade Database	186,500 records	Import/export value, quantity, commodity, country
AIS Vessel Data	Automatic Identification System (AIS)	6,420,000 records	Position, speed, trajectory, arrival, departure
Port Operation Data	Port Monitoring Platform	428,600 records	Throughput, congestion, delay, container flow
Logistics Tracking Data	GPS Logistics Tracking System	1,960,000 records	Route, transportation time, delay, activity index
News and Policy Events	International News Database	96,800 records	Event type, country, commodity, intensity, sentiment
Policy Announcement Data	Government and Trade Agencies	18,600 records	Tariff, restriction, sanction, regulation change

For reproducibility, every raw record is associated with a source identifier, instrument or entity identifier, native timestamp, time zone, access route, and preprocessing version. Market and foreign-exchange observations are deduplicated by instrument and exchange timestamp, corrected for daylight-saving transitions, and converted to UTC. Trade quantities and values retain their reported units and are converted only after commodity-code harmonization; AIS/GPS points with invalid coordinates or implausible speeds (e.g., coordinates exceeding geographic boundaries or vessel speeds > 40 knots) are removed; and duplicated news items are grouped by normalized URL, headline similarity, and a 12 h entity–event window. Specifically, the market quotation records cover major energy and agricultural futures (crude oil, natural gas, soybeans, corn, and copper) retrieved via Bloomberg Terminal API and Refinitiv Eikon, alongside spot FX rates (EUR/USD, USD/CNY, USD/JPY, and GBP/USD). Bilateral trade records are accessed through UN Comtrade and national customs monitoring interfaces, covering major trading corridors across China, the United States, the European Union, Brazil, and Southeast Asia under six-digit HS codes. Vessel trajectories and port states are collected from MarineTraffic AIS feeds and global port performance trackers, encompassing key trans-shipment hubs including Shanghai, Singapore, Rotterdam, Ningbo–Zhoushan, and Los Angeles. All external APIs

and historical bulk archives were retrieved between January 2022 and December 2025 and frozen, with specific endpoint identifiers and field mappings documented in the repository.

The textual event-processing pipeline is implemented based on a multilingual pre-trained language model architecture. Specifically, we employ `xlm-roberta-base` (accompanied by domain-specific fine-tuning on a curated multilingual financial news corpus, checkpoint revision `v1.0`) to process incoming news reports and regulatory releases in English, Chinese, and Spanish. Unstructured texts are tokenized using the SentencePiece tokenizer with a maximum sequence length truncated to 256 tokens. The model is fine-tuned using the AdamW optimizer with a learning rate of 2×10^{-5} across 5 epochs, projecting the final pooling layer into a shared 128-dimensional event embedding space ($D_e = 128$). Events are categorized into a 7-class taxonomy: tariff adjustments, trade restrictions, economic sanctions, shipping disruptions, port operations, macro monetary policies, and extreme weather shocks. Associated countries and commodities are extracted via named entity recognition (NER) aligned with the ISO 3166-1:2020 country standard [41], and HS commodity categories. Event intensity $I_t \in [0, 1]$ is quantified by normalized entity-co-occurrence density and vocabulary salience, while sentiment valence and semantic uncertainty $U_t \in [0, 1]$ are derived from the Softmax entropy of the model's multi-class classification head. Duplicate announcements and minor updates across agencies are resolved using a greedy clustering window of 12 h based on cosine similarity (>0.85) over event embeddings.

3.2. Prediction Targets and Causal Temporal Alignment

Let P_t denote the tradable mid-price (or the settlement price when a mid-price is unavailable) of the target instrument at prediction time t , and let $h \in \{15, 30, 60\}$ minutes. The horizon return is $r_{t,h} = \log(P_{t+h}/P_t)$. Direction prediction is a three-class task: downward if $r_{t,h} < -\epsilon$, stable if $|r_{t,h}| \leq \epsilon$, and upward if $r_{t,h} > \epsilon$. The value of ϵ must be fixed from the training period and reported with the target-instrument roster. Volatility is the realized root-sum-of-squares of one-minute log returns in $(t, t+h]$, $v_{t,h} = \sqrt{\sum_{j=1}^h r_{t+j,1}^2}$. Risk prediction uses low, medium, and high labels determined by the 50th and 90th percentiles of $v_{t,h}$ estimated on the training period only; the same frozen thresholds are applied to validation and test data. These definitions ensure that all three labels refer to the same forecast horizon and that no validation/test statistic enters label construction.

Four timestamps are retained for every observation x : occurrence time $t_o(x)$, publication time $t_p(x)$, ingestion time $t_i(x)$, and model-availability time $t_a(x)$. We define $t_a(x) = \max\{t_p(x), t_i(x), t_q(x)\}$, where $t_q(x)$ is completion of any deterministic quality-control step. The formal prediction-time information set is $\mathcal{I}_t = \{x : t_a(x) \leq t\}$. Inputs for a sample ending at t are constructed exclusively from $\mathcal{I}_t$; the most recent available observation is joined without backward revision, and its age is supplied to the model. When a statistical release is revised, the revision is stored as a new vintage with its own publication and availability times and is never backfilled into earlier samples. News is available no earlier than its first publication/ingestion time, AIS and port states no earlier than the completed aggregation interval, and customs/trade records no earlier than their public release. All timestamps are converted to UTC before the chronological split. This rule is applied before window construction and augmentation.

3.3. Data Augmentation

Cross-border prediction involves heterogeneous data sources, including quotations, exchange rates, trade activities, logistics conditions, and news-policy information, which exhibit substantial differences in sampling frequency, missingness, noise intensity, and event distribution. If a model is trained exclusively on the original samples, excessive de-

pendence may be placed on particular high-frequency or high-quality modalities, resulting in increased sensitivity to missing observations, temporal misalignment, and cross-source conflicts in real-world environments. Therefore, while preserving the principal state and prediction labels of each sample, diversified training instances with different levels of information completeness and cross-modal consistency are constructed through random masking, noise perturbation, temporal shifting, and cross-modal recombination. In this manner, model robustness against missing modalities, asynchronous sampling, local noise, and unexpected events can be enhanced. Let the multimodal input within a temporal window be denoted as $X = X_{m=1}^M$, where M represents the number of modalities and X^m denotes the m -th cross-border sensing signal. The augmented sample can then be uniformly expressed as

$$\tilde{X} = \mathcal{A}(X; \theta), \quad (1)$$

where $\mathcal{A}(\cdot)$ denotes the stochastic data augmentation function, and θ represents augmentation parameters such as the masking ratio, noise intensity, and temporal shift magnitude. To simulate the temporary unavailability of individual information sources frequently encountered in real-world cross-border environments, random modality masking is employed to randomly suppress complete modalities, thereby preventing long-term dependence on a single information source. For the m -th modality, the augmented representation is defined as

$$\tilde{X}^m = b_m X^m, \quad b_m \sim \text{Bernoulli}(1 - p_m), \quad (2)$$

where p_m denotes the masking probability of the corresponding modality, and $b_m = 0$ indicates that the modality is temporarily removed from the current training sample. Meanwhile, temporal masking is further introduced to randomly mask consecutive temporal segments, thereby simulating interruptions in missing AIS trajectories, or local gaps in exchange-rate observations. Given a temporal mask M_t , the masked sequence can be expressed as

$$\tilde{X}_t^m = M_t X_t^m \quad M_t \in \{0, 1\}. \quad (3)$$

This strategy encourages the model to infer the current state from complementary temporal positions and alternative modalities rather than memorizing isolated local patterns. For continuous numerical signals, including prices, exchange rates, port operation states, and logistics indicators, Gaussian noise is further introduced to simulate sensing errors, statistical revisions, and high-frequency disturbances:

$$\tilde{X}_t^m = X_t^m + \epsilon \quad \epsilon \sim \mathcal{N}(0, \sigma_m^2), \quad (4)$$

where σ_m is configured according to the fluctuation scale of each modality. Considering the intrinsic asynchronous propagation relationships among trade, logistics, news, and responses, time shifting is simultaneously applied to randomly modify the temporal positions of non-target modalities:

$$\tilde{X}_t^m = X_{t+\Delta_m}^m \quad \Delta_m \sim \mathcal{U}(-\delta_m, 0), \quad (5)$$

Here $\Delta_m \leq 0$ is a backward-only offset and δ_m is its maximum magnitude. Restricting the perturbation to earlier, already available inputs prevents temporal augmentation from importing future observations. For data with pronounced domain-specific characteristics, logistics trajectory segment masking and news/event masking are further employed. The former randomly removes consecutive AIS/GPS trajectory segments to simulate vessel-positioning loss, missing signals within port regions, and incomplete transportation trajectories. The latter randomly removes portions of news, policy announcements, or

event representations, enabling predictions to remain dependent on trade, and logistics states when event information is incomplete. For a sequence with length T , a randomly selected interval $[t_s, t_e]$ is treated as a missing region:

$$\tilde{X}_t^m = 0, \quad t_s \leq t \leq t_e. \quad (6)$$

In addition to improving robustness against missing information, cross-modal shuffle is specifically introduced to strengthen cross-modal consistency discrimination. The central idea is to replace one modality with information originating from another training sample at a predefined probability, thereby explicitly constructing inconsistent “trade–logistics”, or “trade–market” pairs. Assuming that X_i^a and X_i^b originally belong to the same sample, the perturbed representation is formulated as

$$\tilde{X}_i = X_i^a, X_{\pi(i)}^b, \quad \pi(i) \neq i, \quad (7)$$

where $\pi(\cdot)$ denotes a random permutation function. Through this strategy, differences between genuine cross-modal correspondence and artificially induced conflicts can be learned, thereby providing more explicit supervisory information for subsequent reliability and cross-modal consistency modeling. Finally, because tariff adjustments, sanctions, port disruptions, and major international news events account for only a limited proportion of the complete temporal sequence, event-window sampling is adopted to increase the occurrence probability of strong-event windows during training. Let the event time be denoted as t_e ; the interval $[t_e - \tau_1, t_e + \tau_2]$ surrounding the event is therefore treated as a prioritized sampling window. This strategy prevents the learning process from being dominated by abundant normal observations and enables the model to sufficiently characterize the dynamic process of “event occurrence–cross-border state variation response”. Through the combined augmentation mechanisms described above, the training data cover complex conditions including modality absence, local noise, asynchronous temporal off-sets, cross-source conflicts, and rare events, thereby providing a more robust learning basis for reliability perception, cross-modal consistency estimation, asynchronous temporal interaction, and dynamic event perception.

Augmentation is applied strictly to the training folds after the causal information set $\mathcal{I}_t$ has been constructed; validation and test samples remain untouched to prevent data contamination. Specifically, random modality masking is configured with modality-specific dropout probabilities of 0.05, 0.10, 0.20, 0.20, and 0.15 for market quotes, foreign exchange, trade flows, logistics indicators, and policy–news events, respectively, reflecting their empirical observation completeness in operational environments. Consecutive temporal masks are applied with segment lengths drawn uniformly from $\mathcal{U}(5, 15)$ native steps for high-frequency (minute-level) series and $\mathcal{U}(1, 3)$ steps for lower-frequency records. For continuous numerical features, Gaussian perturbations $\epsilon \sim \mathcal{N}(0, \sigma_m^2)$ are injected following robust standardization via training-set median and interquartile range (IQR), where standard deviations are fixed at $\sigma_m = 0.01$ for market and FX series and $\sigma_m = 0.02$ for trade and logistics series. To provide explicit supervisory contrast for cross-modal consistency discrimination, cross-modal shuffling is activated with a probability of $p_{\text{shuffle}} = 0.10$, drawing replacement candidates strictly from within the same chronological training split. Furthermore, to mitigate extreme class and regime imbalances during major international disruptions, event-centered sampling windows are constructed over $[t_e - 60 \text{ min}, t_e + 120 \text{ min}]$ surrounding identified trade and policy events, with a prioritized sampling factor of $3 \times$.

Random time shifting is strictly causal and input-only. The target modality and all labels remain fixed. For a non-target modality, the shift is sampled as $\Delta_m \sim \mathcal{U}(-\delta_m, 0)$

rather than from a two-sided interval; δ_m is 5 min for FX, 15 min for events, 30 min for AIS/port/logistics aggregates, and one native publication interval for trade statistics. A shifted value is accepted only if its recorded model-availability time remains no later than t ; otherwise it is masked. Consequently, neither an observation from $(t, t + h]$ nor a later data vintage can enter the input at t , and the label $y_{t,h}$ is never shifted or recomputed.

3.4. Proposed Method

3.4.1. Overall

After preprocessing, quotations, exchange-rate and cross information, cross-border trade conditions, logistics operation states, and news-policy events are separately fed into modality-specific encoders to extract exchange rate, trade, logistics, and event representations within a unified latent feature space. Based on these representations, the reliability- and consistency-aware module is first constructed to dynamically estimate the credibility of each modality according to observation completeness, update delay, local fluctuation stability, and source availability. Meanwhile, cross-source consistency relationships among trade-logistics, news-trade, and news-pairs are learned through inter-modal representation interactions. A conflict representation is further generated to suppress abnormal, delayed, or contradictory information, thereby obtaining reliability-weighted multimodal representations. Subsequently, the modality representations are introduced into the asynchronous cross-border temporal interaction module, where continuous temporal information is incorporated to preserve the actual temporal distances among heterogeneous observations. Cross-modal attention mechanisms are employed to learn directional dynamic dependencies among trade activities, logistics conditions, and exchange rates. Furthermore, cross-border propagation characteristics are extracted from instantaneous, short-term, and long-term lag perspectives, and different temporal-scale representations are adaptively integrated through a gating mechanism according to current conditions. Consequently, a temporal representation containing cross-border transmission pathways and propagation delays is obtained. The resulting representation is then jointly combined with news and policy event features and delivered into the dynamic event perception module. The effective impact of external events is estimated according to event intensity, associated countries, related assets, and current trade, logistics, and states. Meanwhile, an event residual correction mechanism is introduced on top of the baseline trend prediction, enabling the model to deviate from normal evolution patterns when tariff adjustments, trade restrictions, port disruptions, or major international events occur. Finally, reliability-aware features, asynchronous propagation representations, event-impact features, and current contexts are jointly input into the context-aware adaptive fusion network. The contribution of heterogeneous information sources is dynamically adjusted according to states, and a unified cross-border representation is generated. This representation is subsequently delivered to three prediction heads for direction, volatility, and risk estimation, enabling joint forecasting of future short-term movements, volatility levels, and risk categories. Therefore, an end-to-end framework integrating multimodal sensing, reliability assessment, asynchronous propagation modeling, event correction, and multi-task prediction is established.

3.4.2. Module I: Reliability- and Consistency-Aware Cross-Border Sensing Representation

To enhance the stability of cross-border multimodal representations under missing observations, delayed updates, and cross-source conflicts, five types of input signals, including market, exchange-rate, trade, logistics, and event information, are denoted as X_t^m , where $m \in \{\text{mar, fx, tra, log, evt}\}$.

As shown in Figure 1, each modality is processed by an independent three-stage encoder to perform feature mapping. The input feature width, temporal length, and initial channel dimension are denoted as W_m , H_m , and C_m , respectively. The output channels of the three encoding layers are represented as C_1 , C_2 , and C_3 , and all modality representations are further projected into a shared latent dimension D through linear transformation. For sequential signals, including exchange-rate and logistics information, one-dimensional convolution and gated residual units are employed to capture local variation patterns. For trade and event representations, two-layer feed-forward projections combined with self-attention mechanisms are utilized for semantic compression. The unified encoding process is formulated as

$$Z_t^m = \phi_m(X_t^m) = W_m^{(3)} \sigma \left(W_m^{(2)} \sigma \left(W_m^{(1)} X_t^m + b_m^{(1)} \right) + b_m^{(2)} \right) + b_m^{(3)}, \quad (8)$$

where $\sigma(\cdot)$ denotes the nonlinear activation function. Subsequently, a reliability estimator is constructed by jointly incorporating modality representations with missing masks M_t^m , observation intervals Δ_t^m , local noise levels N_t^m , historical stability indicators S_t^m , and source availability measurements A_t^m . A two-layer perception network is employed to estimate the dynamic reliability score:

$$r_t^m = \text{Sigmoid} \left(W_r^{(2)} \sigma \left(W_r^{(1)} [Z_t^m, M_t^m, \Delta_t^m, N_t^m, S_t^m, A_t^m] + b_r^{(1)} \right) + b_r^{(2)} \right). \quad (9)$$

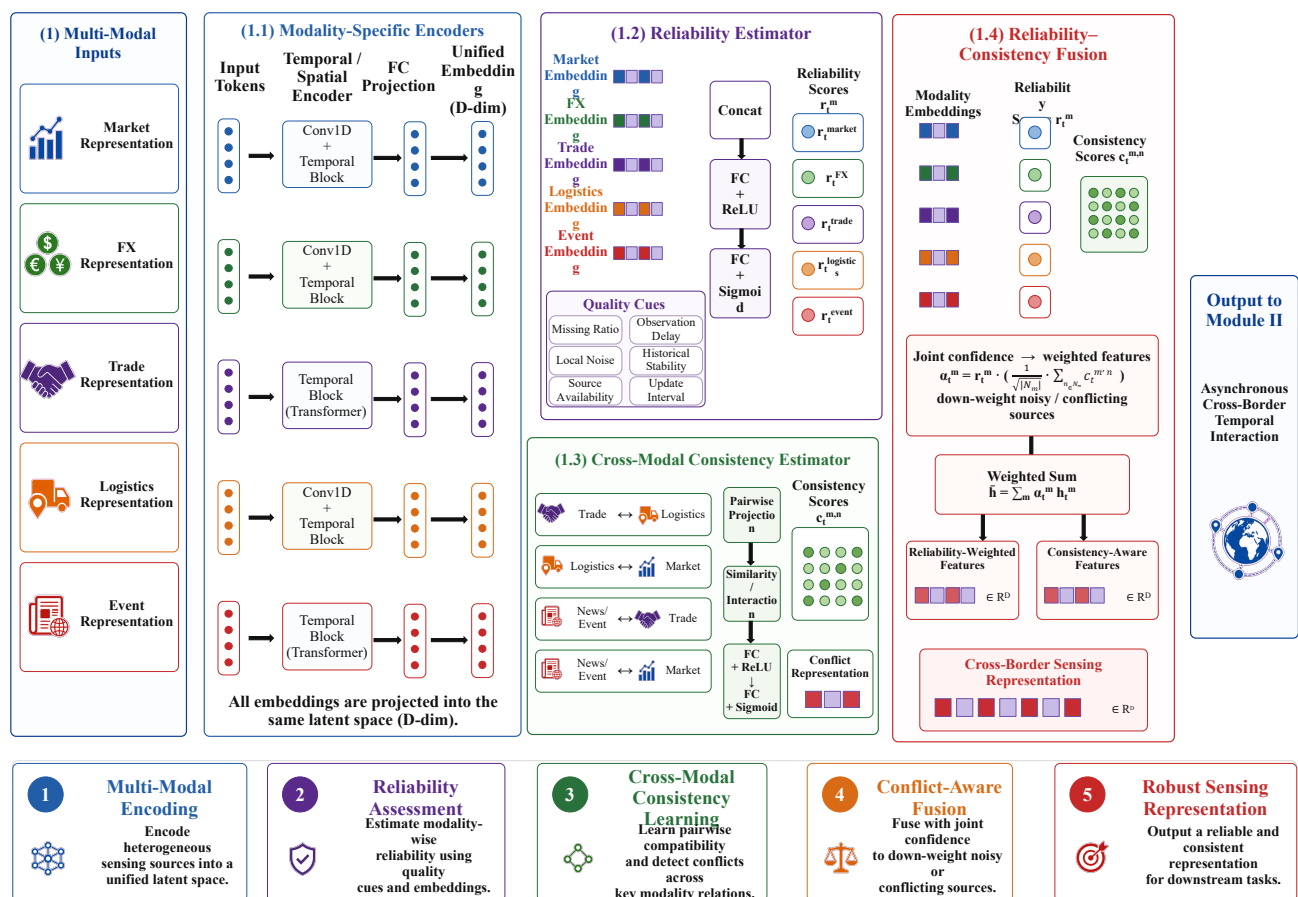


Figure 1. Reliability- and consistency-aware cross-border sensing representation module.

All reliability cues are computed from observations available at or before t . Specifically, M_t^m is the observed/missing indicator vector in the causal input window; $\Delta_t^m = \min\{(t - t_a(x_{\text{last}}^m))/d_m^{95}, 1\}$ is the clipped age of the last available observation normalized by the 95th

percentile d_m^{95} estimated on the training set; $N_t^m = \text{clip}(\text{MAD}(\Delta x^m) / \text{IQR}_{\text{train}}(x^m), 0, 1)$ measures local first-difference noise; $S_t^m = 1 - \text{clip}(\text{MAD}(x^m - \text{median}(x^m)) / \text{IQR}_{\text{train}}(x^m), 0, 1)$ measures historical stability; and A_t^m is the fraction of expected source updates successfully received during the preceding 30 calendar days. Local noise and stability use the most recent 60 native observations, or all available observations when fewer than 60 exist. Medians, interquartile ranges, clipping constants, and expected update schedules are fitted on the training period and then frozen. No cue uses x_{t+1} , a future release, or a retrospectively revised vintage. However, reliability estimation from individual modalities alone cannot determine whether information sources are mutually supportive. Therefore, a cross-modal consistency estimator is further developed. For each correlated modality pair (m, n) , both representations are first projected into a shared consistency space. The compatibility relationship is then calculated by jointly considering feature differences, Hadamard interactions, and conditional correlations:

$$c_t^{m,n} = \text{Sigmoid}(W_c[|P_m Z_t^m - P_n Z_t^n|, (P_m Z_t^m) \odot (P_n Z_t^n), P_m Z_t^m, P_n Z_t^n] + b_c). \quad (10)$$

Accordingly, the overall consistency of each modality is determined by jointly considering other economically related modalities:

$$\bar{c}_t^m = \frac{\sum_{n \neq m} a_{mn} c_t^{m,n}}{\sum_{n \neq m} a_{mn} + \epsilon}, \quad (11)$$

where a_{mn} represents predefined interaction relationships, including trade–logistics, news–trade, and news–associations. To prevent a single modality with relatively high intrinsic quality but significant inconsistency with the overall cross-border state from receiving excessive importance, reliability and consistency scores are jointly mapped into a unified credibility coefficient:

$$q_t^m = \frac{\exp(\eta_r r_t^m + \eta_c \bar{c}_t^m - \eta_f f_t^m)}{\sum_j \exp(\eta_r r_t^j + \eta_c \bar{c}_t^j - \eta_f f_t^j)}, \quad (12)$$

where the conflict measurement f_t^m is determined according to the deviation between a modality and its consistency neighborhood:

$$f_t^m = \sum_{n \neq m} a_{mn} (1 - c_t^{m,n}) \|P_m Z_t^m - P_n Z_t^n\|_2. \quad (13)$$

The final reliability-weighted representation is generated as

$$\tilde{Z}_t^m = q_t^m Z_t^m + (1 - q_t^m) G_m(\bar{Z}_t^{-m}), \quad (14)$$

where $G_m(\bar{Z}_t^{-m})$ denotes the compensation representation constructed from other reliable modalities. Since q_t^m is normalized through the Softmax function, $q_t^m \geq 0$ and $\sum_m q_t^m = 1$ are guaranteed. Therefore, the fused representation is constrained within the convex hull of multimodal representations, preventing unbounded amplification caused by individual abnormal modalities. Furthermore, assuming that the modality error is represented as ε_m and errors among different modalities are approximately independent, the weighted fusion error satisfies:

$$\text{Var}\left(\sum_m q_t^m \varepsilon_m\right) = \sum_m (q_t^m)^2 \text{Var}(\varepsilon_m). \quad (15)$$

When low-reliability and high-conflict modalities exhibit larger error variances, their corresponding q_t^m values are automatically reduced through the joint credibility mechanism.

Consequently, the upper bound of fusion uncertainty is further reduced compared with fixed-weight fusion strategies. This design enables abnormal information to be adaptively suppressed under conditions such as prolonged AIS signal absence, delayed trade statistics, uncertain news sources, or contradictory cross-border signals, while preserving valuable information when multiple sources provide mutually supportive evidence. Therefore, a stable and economically consistent multimodal representation is provided for subsequent asynchronous cross-border propagation modeling.

3.4.3. Module II: Asynchronous Cross-Border Temporal Interaction

After the reliability-calibrated multimodal representations $\tilde{Z}^m \in \mathbb{R}^{H_m \times W_m \times C_m}$ are obtained from the previous module, the asynchronous cross-border temporal interaction module avoids forcibly mapping heterogeneous sources into identical observation timestamps. Instead, the original temporal characteristics associated with each observation are preserved.

As shown in Figure 2, a continuous-time projection network with L_t layers is first employed to jointly encode feature representations, observation timestamps, adjacent sampling intervals, information publication delays, and modality types. The channel dimension is progressively transformed from C_m into the shared temporal representation dimension C_t . For the i -th observation of modality m , the continuous-time state is defined as

$$U_i^m = \text{LN}\left(\tilde{Z}_i^m + W_\tau^m \Psi(\tau_i^m, \Delta_i^m, d_i^m) + E^m\right), \quad (16)$$

where $\Psi(\cdot)$ represents a learnable continuous-time mapping function, and τ_i^m , Δ_i^m , and d_i^m denote the actual observation timestamp, adjacent observation interval, and publication delay, respectively. Subsequently, a directional cross-modal attention network consisting of L_a layers and K attention heads is constructed to model propagation relationships among trade-to-logistics, logistics-to-market, exchange-rate-to-market, policy-to-market, and overseas-market-to-target-market interactions. To prevent conventional attention mechanisms from ignoring temporal distances, a learnable temporal decay term is explicitly introduced between query and key representations:

$$e_{ij}^{m \rightarrow n} = \frac{(Q_n U_i^n)^T (K_m U_j^m)}{\sqrt{C_t}} - \lambda_{mn} \rho(\tau_i^n - \tau_j^m), \quad (17)$$

where $\rho(\cdot)$ denotes a non-negative temporal distance function, and λ_{mn} represents the temporal sensitivity coefficient associated with different cross-border propagation pathways. The asynchronous interaction representation is subsequently obtained as

$$A_i^{m \rightarrow n} = \sum_{j: \tau_j^m \leq \tau_i^n} \frac{\exp(e_{ij}^{m \rightarrow n})}{\sum_{k: \tau_k^m \leq \tau_i^n} \exp(e_{ik}^{m \rightarrow n})} V_m U_j^m, \quad (18)$$

where only observations occurring before the target timestamp are allowed to participate in the attention calculation, thereby preventing future information leakage. Furthermore, three symbolic lag branches are constructed to capture instantaneous, short-term, and long-term propagation patterns, whose receptive fields are denoted as Ω_{ins} , Ω_{sho} , and Ω_{lon} , respectively. Each branch consists of L_b lag-aware attention blocks and feed-forward networks with width C_b . For all three forecast horizons, the lag intervals are fixed in elapsed UTC time as $\Omega_{\text{ins}} = [0, 15]$ minutes, $\Omega_{\text{sho}} = (15, 60]$ minutes, and $\Omega_{\text{lon}} = (60, 240]$ minutes. Each branch uses two lag-aware attention blocks, four heads, and width 128; observations are assigned by their availability time rather than their occurrence time.

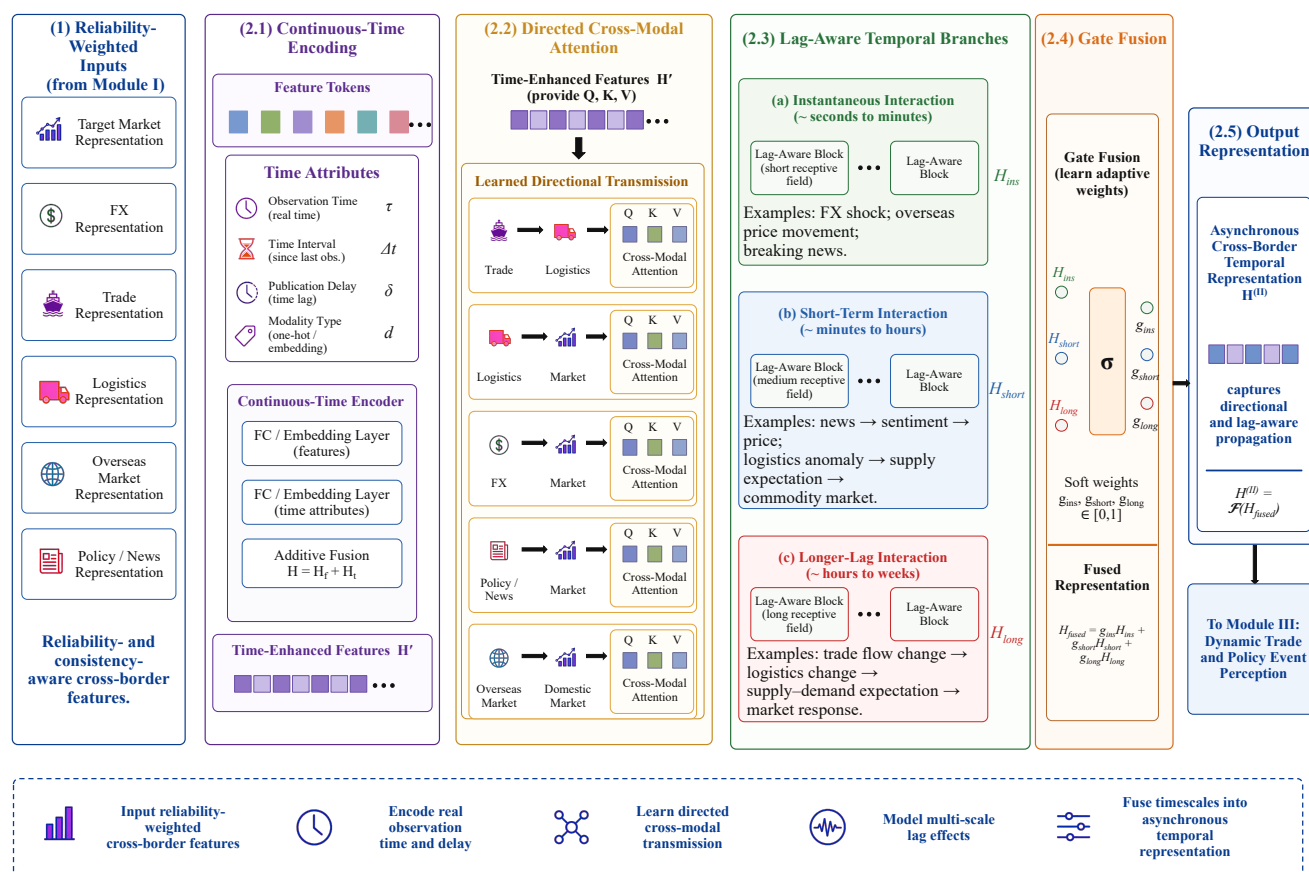


Figure 2. Asynchronous cross-border temporal interaction module.

$$R_i^{(s)} = \sum_{m,n} \sum_{j \in \Omega_s(i)} \alpha_{ij,s}^{m \rightarrow n} V_s^{m,n} U_j^m, \quad s \in \{\text{ins, sho, lon}\}. \quad (19)$$

The instantaneous branch captures immediate responses caused by exchange-rate fluctuations, international movements, and sudden news events. The short-term branch models rapid transmission processes, such as news-induced sentiment changes and logistics disruptions affecting supply expectations. The long-term branch focuses on delayed propagation mechanisms, including trade flow variations, logistics evolution, supply–demand expectation adjustments, and subsequent reactions. Finally, a two-layer gating network is utilized to dynamically integrate different temporal scales according to current states:

$$g_i^{(s)} = \frac{\exp(w_s^T \sigma(W_g[R_i^{(s)}, U_i^{\text{mar}}] + b_g))}{\sum_r \exp(w_r^T \sigma(W_g[R_i^{(r)}, U_i^{\text{mar}}] + b_g))}, \quad T_i = \sum_s g_i^{(s)} R_i^{(s)}. \quad (20)$$

Since $g_i^{(s)} > 0$ and $\sum_s g_i^{(s)} = 1$, the fused representation T_i remains within the convex combination space of different lag representations, satisfying

$$\|T_i\|_2 \leq \sum_s g_i^{(s)} \|R_i^{(s)}\|_2 \leq \max_s \|R_i^{(s)}\|_2. \quad (21)$$

This inequality demonstrates that the gating mechanism prevents excessive amplification of abnormal responses from individual temporal scales. Meanwhile, because $\rho(\Delta\tau)$ increases with temporal distance, historical observations with larger time intervals naturally receive reduced attention weights under similar semantic relevance. However, signals containing long-term trade transmission information can still be preserved through appropriately learned smaller λ_{mn} . Therefore, this module is capable of distinguishing

immediate impacts from exchange rates and unexpected news, short-term transmission from logistics anomalies, and delayed effects from trade flow variations. By maintaining authentic temporal relationships, an asynchronous propagation representation describing the pathway of trade–logistics–exchange rate–overseas is established, providing temporally directional and cross-dependent information for subsequent dynamic trade and policy event impact modeling.

3.4.4. Module III: Dynamic Trade and Policy Event Perception

After the asynchronous cross-border temporal interaction module, the propagated cross-border state at time t is represented as $T_t \in \mathbb{R}^{H_t \times W_t \times C_t}$. Meanwhile, external events, including tariff adjustments, trade restrictions, sanctions, policy announcements, port interruptions, shipping disruptions, extreme weather conditions, and major international news, are encoded as event tensors $E_t \in \mathbb{R}^{H_e \times W_e \times C_e}$.

As shown in Figure 3, the event perception network consists of an L_e -layer event encoder, an L_g -layer event gating network, and an L_r -layer residual correction network. The channel dimension of the l -th layer is denoted as $C_e^{(l)}$, and the final event representation is projected into the same latent space dimension D_e as the cross-border propagation representation. To avoid estimating event impacts solely according to textual sentiment information, event types, intensity levels, associated countries, related commodities, and uncertainty characteristics are jointly incorporated into the event encoder. The integrated event representation is formulated as

$$F_t^e = \mathcal{E}_\theta \left([E_t, P_t^{\text{type}}, P_t^{\text{country}}, P_t^{\text{asset}}, I_t, U_t] \right), \quad (22)$$

where $\mathcal{E}_\theta$ denotes the multi-layer event encoding network, and I_t and U_t represent event intensity and event uncertainty, respectively. Subsequently, a conditional interaction layer is developed to establish correlations between event semantics and the current trade, and logistics propagation states. In this manner, only events with actual relevance to the current cross-border condition can generate significant responses:

$$S_t^e = \tanh(W_s F_t^e + W_t T_t + W_x (F_t^e \odot T_t) + b_s). \quad (23)$$

Based on the event-state interaction representation, an event gating network is constructed. The hidden layer widths of the gating network are denoted as $D_g^{(1)}, \dots, D_g^{(L_g)}$, and the event impact coefficient γ_t is obtained as

$$\gamma_t = \frac{1}{1 + \exp[-\mathcal{G}_\omega(S_t^e, I_t, U_t, T_t)]}. \quad (24)$$

Since the Sigmoid function constrains the output within $0 < \gamma_t < 1$, event information is prevented from completely dominating the normal evolution process. Furthermore, a trend–event residual correction mechanism is introduced to explicitly separate normal dynamics from external event-induced disturbances. The baseline trend is first predicted from the cross-border temporal representation:

$$B_t = \mathcal{F}_b(T_t), \quad (25)$$

where $\mathcal{F}_b(\cdot)$ represents the baseline trend prediction network. The event residual is then generated by jointly considering the gated event representation and the baseline trend:

$$R_t^e = \mathcal{F}_r([\gamma_t S_t^e, B_t, B_t \odot S_t^e]), \quad (26)$$

where $\mathcal{F}_r(\cdot)$ denotes the residual correction network. The final event-adjusted state is obtained as

$$Y_t^e = B_t + \gamma_t R_t^e. \quad (27)$$

The proposed event correction mechanism provides an explicit stability constraint. Assuming that the residual network $\mathcal{F}_r$ satisfies the Lipschitz condition with constant L_r , and the event disturbance is represented as δE_t , the variation in the final event-adjusted state satisfies

$$\|\delta Y_t^e\|_2 \leq \|\delta B_t\|_2 + \gamma_t L_r \|\delta E_t\|_2. \quad (28)$$

Because $\gamma_t < 1$, the influence of event disturbances is bounded after the gating operation. Therefore, weakly correlated news, uncertain information sources, or temporary event noise cannot directly cause excessive deviations in the predicted state. Conversely, when events such as tariff adjustments, economic sanctions, or port disruptions exhibit strong consistency with trade, logistics, and conditions, γ_t is automatically increased, allowing event residuals to effectively correct the baseline trend. Through this design, different external events are no longer treated equivalently. The stability of normal evolution can be maintained, while significant event-driven shocks with clear cross-border economic transmission effects are effectively captured. Consequently, an event-conditioned state representation is generated and provided to the subsequent context-aware adaptive fusion module for joint prediction of direction, volatility, and risk levels.

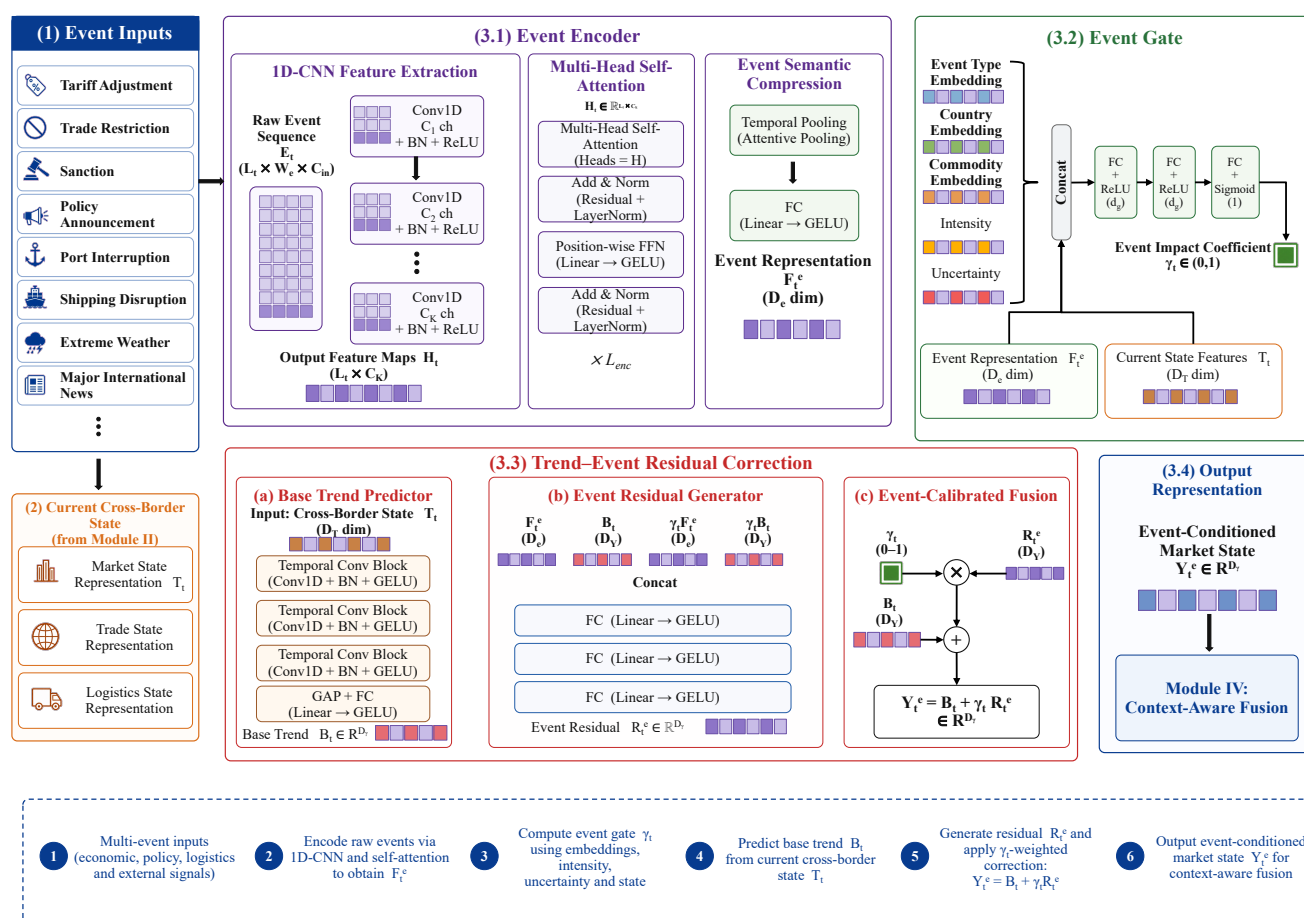


Figure 3. Dynamic trade and policy event perception module.

3.4.5. Module IV: Context-Aware Fusion and Multi-Task Objective

The fusion inputs are the reliability-pooled sensing representation $H_t^{\text{rel}} = \sum_m q_t^m \tilde{Z}_t^m$, asynchronous representation T_t , event-adjusted representation Y_t^e , and current market context U_t^{mar} . After separate linear projections to 128 dimensions, their context-dependent weights are computed by a two-layer gate,

$$\beta_t = \text{Softmax}\left(W_2 \text{GELU}\left(W_1[H_t^{\text{rel}}, T_t, Y_t^e, U_t^{\text{mar}}] + b_1\right) + b_2\right), \quad F_t = \sum_{k=1}^4 \beta_{t,k} P_k H_{t,k}, \quad (29)$$

where the gating hidden width is 128 and dropout is 0.2. Separate heads map F_t to three-class direction logits, a non-negative volatility estimate (Softplus output), and three-class risk logits. The task losses are cross-entropy for direction, Huber loss for volatility, and class-weighted cross-entropy for risk. Their joint objective uses learned homoscedastic uncertainty weights,

$$\mathcal{L}_{\text{total}} = e^{-s_d} \mathcal{L}_{\text{dir}} + e^{-s_v} \mathcal{L}_{\text{vol}} + e^{-s_r} \mathcal{L}_{\text{risk}} + s_d + s_v + s_r + \lambda_c \mathcal{L}_{\text{cons}}, \quad (30)$$

with $s_d = s_v = s_r = 0$ at initialization (equivalent to unit task weights) and $\lambda_c = 0.1$. Class weights for the risk head are calculated from the training fold only. This formulation makes the previously described “dynamic” task weighting explicit and prevents the scale of the regression loss from dominating the two classification losses.

4. Results and Discussion

4.1. Experimental Configuration

4.1.1. Hardware and Software Platform

All experiments were conducted on a high-performance GPU computing platform. The hardware configuration consisted of an Intel Xeon Gold 6430 processor, an NVIDIA GeForce RTX 4090 GPU with 24 GB memory, 128 GB DDR5 RAM, and a 2 TB NVMe solid-state drive for storing experimental data and model parameters. The GPU was primarily utilized for parallel training and inference of Transformer, PatchTST, Multimodal Transformer, and the proposed framework, while the CPU was responsible for data preprocessing, temporal window construction, and data loading, exchange-rate, trade, logistics, and news-policy information. All baseline models and the proposed method were executed under identical hardware conditions to minimize the influence of computational resource differences on experimental comparisons. The software environment was configured with the Ubuntu 22.04 LTS 64-bit operating system. All models were implemented based on Python 3.10 and PyTorch 2.2, with CUDA 12.1 and cuDNN 8.9 configured for GPU acceleration. Data cleaning, statistical analysis, and feature processing were performed using NumPy, Pandas, and Scikit-learn. Time-series models and traditional machine learning baselines were implemented using Statsmodels, XGBoost, and Scikit-learn, while semantic representations of textual news and policy events were extracted through the Transformers framework. During model optimization, TensorBoard was employed to record variations in loss functions and evaluation metrics, and random seeds of Python, NumPy, and PyTorch were fixed to ensure experimental reproducibility. The complete dataset was divided chronologically into training, validation, and testing subsets with proportions of 60%, 20%, and 20%, respectively, thereby avoiding future information leakage caused by random partitioning. Within the training subset, five-fold temporal cross-validation was further adopted for hyperparameter selection. In each fold, training samples were strictly earlier than validation samples, and the final hyperparameters were determined according to the average performance across five folds. The input window length was set to 60 time steps, and future states were predicted at 15 min, 30 min, and 60 min horizons. The batch

size was set to 64, and the maximum number of training epochs was configured as 100. The AdamW optimizer was adopted with an initial learning rate of $\alpha = 1 \times 10^{-4}$ and a weight decay coefficient of 1×10^{-5} . The hidden feature dimension was set to $d = 128$, the number of attention heads was configured as 4, the number of Transformer encoder layers was set to 3, the hidden dimension of the feed-forward network was 256, and the dropout rate was set to 0.2. During training, a learning rate decay strategy was employed, and early stopping was introduced to terminate training when the validation loss did not improve for 10 consecutive epochs. Meanwhile, the gradient clipping threshold was set to 1.0 to improve optimization stability. For multi-task learning, the initial weights of direction, volatility, and risk prediction losses were all set to 1.0, and were subsequently adjusted dynamically according to task uncertainty. Temporal cross-validation is employed strictly for model selection and hyperparameter tuning. To eliminate potential look-ahead bias and serial correlation leakage, an embargo and purge interval equal to the maximum input window length (60 steps) plus the longest forecast horizon (60 min) is enforced at every chronological split boundary. Once the optimal hyperparameters are identified, the model is retrained on the concatenated training and validation folds (the first 80% of the chronological sequence) and evaluated on the untouched final 20% out-of-sample test period across the 15, 30, and 60 min prediction horizons.

4.1.2. Baseline Models and Evaluation Metrics

Nine representative models were selected as baseline methods, namely ARIMA [42], XGBoost [43], LSTM [44], TCN [45], Transformer [46], PatchTST [47], Early Fusion [48], Multimodal Transformer [49], and cross-modal attention network [50]. ARIMA performs prediction by exploiting autoregressive, differencing, and moving-average relationships within historical sequences, providing strong interpretability with a simple structure. XGBoost utilizes gradient-boosted decision trees to model complex nonlinear relationships and demonstrates strong capability in handling structured features. LSTM captures long- and short-term temporal dependencies through gating mechanisms and is suitable for sequential data with persistent correlations. TCN extracts multi-scale temporal patterns through causal and dilated convolutions while maintaining high parallelization efficiency. Transformer captures global temporal dependencies through self-attention mechanisms and is effective for modeling long-range interactions. PatchTST divides time-series sequences into multiple patches and applies Transformer encoding, enabling joint modeling of local patterns and long-term trends. Early Fusion directly concatenates heterogeneous modality features at the input stage, providing a simple and efficient multimodal integration strategy. Multimodal Transformer employs multimodal attention mechanisms to jointly model, trade, logistics, and textual information, thereby exploiting complementary relationships among different sources. Cross-modal attention network learns association weights among heterogeneous information sources through cross-modal attention, highlighting informative cross-source interactions while suppressing redundant features.

Baseline selection and tuning follow the same temporal folds and early-stopping rule as the proposed model. ARIMA orders are selected by AIC over $p, q \in \{0, \dots, 5\}$ and $d \in \{0, 1\}$; XGBoost is tuned over 300–800 trees, depth 3–8, learning rate 0.01–0.1, and subsample/column-subsample 0.7–1.0. LSTM uses one to three layers and hidden width 64–256; TCN uses three or four residual blocks with kernel size 3 and dilations 1, 2, 4, 8; Transformer uses two to four layers, four or eight heads, and width 64–256; PatchTST uses patch lengths 8, 16, or 24 with 50% overlap. Early Fusion, Multimodal Transformer, and the cross-modal attention network receive the identical five modalities, causal windows, preprocessing, and availability masks used by the proposed framework. Unimodal ARIMA, XGBoost, LSTM, TCN, Transformer, and PatchTST receive only the market sequence by

design; this distinction is stated explicitly and no multimodal competitor receives privileged features. All search decisions use validation data only.

The evaluation metrics include accuracy, macro-precision, macro-recall, macro-F1, macro-AUC, MAE, RMSE, MAPE, R^2 , PR-AUC, and FPR. For multi-class classification tasks (K classes, where $K = 3$ for direction and risk states), the corresponding metric formulations are defined as:

$$\begin{aligned} \text{Accuracy} &= \frac{\sum_{k=1}^K TP_k}{N} = \frac{\sum_{i=1}^N \mathbb{I}(y_i = \hat{y}_i)}{N}, \\ \text{Precision} &= \frac{1}{K} \sum_{k=1}^K \frac{TP_k}{TP_k + FP_k}, \\ \text{Recall} &= \frac{1}{K} \sum_{k=1}^K \frac{TP_k}{TP_k + FN_k}. \end{aligned} \quad (31)$$

$$F1_k = \frac{2P_k R_k}{P_k + R_k}, \quad \text{Macro-F1} = \frac{1}{K} \sum_{k=1}^K F1_k, \quad (32)$$

$$\begin{aligned} \text{AUC} &= \frac{1}{K} \sum_{k=1}^K \int_0^1 TPR_k(FPR_k) dFPR_k, \\ \text{PR-AUC} &= \int_0^1 \text{Precision}(\text{Recall}) d\text{Recall}, \\ \text{FPR} &= \frac{FP}{FP + TN}. \end{aligned} \quad (33)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (34)$$

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (35)$$

where K denotes the number of target classes ($K = 3$), N represents the total number of evaluation instances, and $\mathbb{I}(\cdot)$ denotes the indicator function. For each class $k \in \{1, \dots, K\}$, TP_k , FP_k , and FN_k denote the number of true positives, false positives, and false negatives in a one-vs-rest formulation, respectively; P_k and R_k represent the class-specific precision and recall, with the reported precision and recall metrics corresponding to their macro-averaged values across all classes. Multi-class AUC is calculated via the one-vs-rest macro-average of individual ROC curves, with TPR_k and FPR_k denoting the class-specific true positive and false positive rates. For binary risk-alert indicators, FP and TN represent false positives and true negatives. In continuous forecasting, y_i and $\hat{y}_i$ indicate the ground-truth and predicted values, respectively, and $\bar{y}$ denotes the mean value of the ground-truth samples.

4.2. Short-Term Cross-Border Direction Prediction

This experiment was designed to evaluate the effectiveness of different methods for short-term cross-border direction prediction and to investigate the influence of multisource cross-border sensing information, temporal dependency modeling, and cross-modal interaction mechanisms on trend identification. Strict temporal cross-validation was adopted, and five evaluation metrics, namely accuracy, precision, recall, macro-F1, and AUC, were employed to comprehensively assess the capability of each model to distinguish future upward, stable, and downward states.

As shown in Table 2 and Figure 4, the experimental results indicate that the conventional statistical model ARIMA achieved the lowest predictive performance, with accuracy, macro-F1, and AUC values of 0.621, 0.601, and 0.672, respectively. This result can be primarily attributed to the linear autocorrelation assumptions underlying ARIMA. Although periodic variations in relatively stable time series can be captured effectively, nonlinear changes induced by external factors such as cross-border trade flows, logistics states, and policy events remain difficult to model. XGBoost achieved a clear improvement over ARIMA, with accuracy increasing to 0.694, demonstrating that nonlinear combinations based on decision trees are capable of exploiting complex relationships among structured and cross-border variables. Nevertheless, this method largely depends on manually constructed features and lacks explicit capability for modeling continuous temporal propagation processes, which limits its effectiveness in highly dynamic environments. LSTM and TCN further improved prediction performance. LSTM achieved an accuracy of 0.716 and an AUC of 0.773, benefiting from its gated structure for retaining long-term historical information. TCN further increased these metrics to 0.731 and 0.791, respectively, because causal and dilated convolutions expand the temporal receptive field while maintaining computational efficiency, making the architecture more suitable for capturing multi-scale variation patterns in cross-border. Transformer and PatchTST achieved accuracy values of 0.748 and 0.761, respectively, indicating that attention-based architectures can overcome the sequential computation constraints of recurrent structures and identify informative signals across different temporal positions through global dependency modeling. In particular, PatchTST further improves performance by jointly learning local fluctuation patterns and long-term trends through patch-based temporal representations.

Table 2. Performance comparison of different methods for short-term cross-border direction prediction. Results are reported as mean $\pm$ 95% confidence interval.

Method	Accuracy $\uparrow$	Precision $\uparrow$	Recall $\uparrow$	Macro-F1 $\uparrow$	AUC $\uparrow$
ARIMA	0.621 $\pm$ 0.012	0.608 $\pm$ 0.015	0.596 $\pm$ 0.013	0.601 $\pm$ 0.014	0.672 $\pm$ 0.011
XGBoost	0.694 $\pm$ 0.010	0.681 $\pm$ 0.012	0.673 $\pm$ 0.011	0.676 $\pm$ 0.010	0.748 $\pm$ 0.009
LSTM	0.716 $\pm$ 0.011	0.704 $\pm$ 0.013	0.698 $\pm$ 0.012	0.700 $\pm$ 0.011	0.773 $\pm$ 0.010
TCN	0.731 $\pm$ 0.009	0.719 $\pm$ 0.010	0.712 $\pm$ 0.011	0.715 $\pm$ 0.009	0.791 $\pm$ 0.008
Transformer	0.748 $\pm$ 0.010	0.736 $\pm$ 0.011	0.729 $\pm$ 0.010	0.732 $\pm$ 0.010	0.814 $\pm$ 0.009
PatchTST	0.761 $\pm$ 0.009	0.752 $\pm$ 0.010	0.741 $\pm$ 0.009	0.746 $\pm$ 0.010	0.829 $\pm$ 0.008
Early Fusion	0.772 $\pm$ 0.008	0.764 $\pm$ 0.009	0.755 $\pm$ 0.010	0.759 $\pm$ 0.009	0.842 $\pm$ 0.007
Multimodal Transformer	0.791 $\pm$ 0.008	0.783 $\pm$ 0.009	0.776 $\pm$ 0.008	0.779 $\pm$ 0.008	0.861 $\pm$ 0.007
Cross-Modal Attention Network	0.803 $\pm$ 0.007	0.796 $\pm$ 0.008	0.788 $\pm$ 0.008	0.792 $\pm$ 0.007	0.874 $\pm$ 0.006
Proposed Method	0.842 $\pm$ 0.006	0.836 $\pm$ 0.007	0.829 $\pm$ 0.006	0.832 $\pm$ 0.006	0.913 $\pm$ 0.005

Bold indicates the best results.

Compared with unimodal time-series models, multimodal approaches exhibited stronger predictive capability. Early Fusion achieved an accuracy of 0.772, indicating that the joint use of trade, logistics, and event information provides additional cross-border state information. However, direct feature concatenation cannot effectively distinguish the importance of heterogeneous data sources and remains sensitive to low-quality or conflicting information. Multimodal Transformer further improved accuracy and AUC to 0.791 and 0.861, respectively, demonstrating that multimodal attention mechanisms can effectively exploit complementary relationships among heterogeneous information sources. Cross-modal attention network achieved an accuracy of 0.803 and an AUC of 0.874. Through explicit cross-modal attention weighting, interactions among critical information sources were strengthened, resulting in superior performance compared with conventional multimodal fusion strategies. The proposed method achieved the best predictive performance, with accuracy, macro-F1, and AUC reaching 0.842, 0.832, and 0.913, respectively, corre-

sponding to improvements of 3.9, 4.0, and 3.9 percentage points over cross-modal attention network. These gains are mainly attributed to the reliability, consistency, and asynchronous propagation modeling mechanisms specifically designed for cross-border characteristics. From a structural perspective, conventional fusion methods generally assume comparable contributions among modalities, whereas the proposed framework dynamically reduces the influence of missing, delayed, and noisy information through reliability-aware weighting and identifies genuine relationships among trade, logistics, news, and signals through cross-modal consistency modeling. In addition, the asynchronous temporal interaction mechanism avoids forcibly synchronizing heterogeneous data with different sampling frequencies, thereby enabling the propagation pathway from trade variation to logistics change and subsequent response to be learned more effectively. The dynamic event perception module further enhances responsiveness to tariff adjustments, trade restrictions, and sudden policy events. Consequently, compared with methods relying on historical trends or static fusion strategies, the proposed framework more accurately characterizes short-term direction changes jointly driven by multisource cross-border factors, resulting in more stable and reliable predictions.

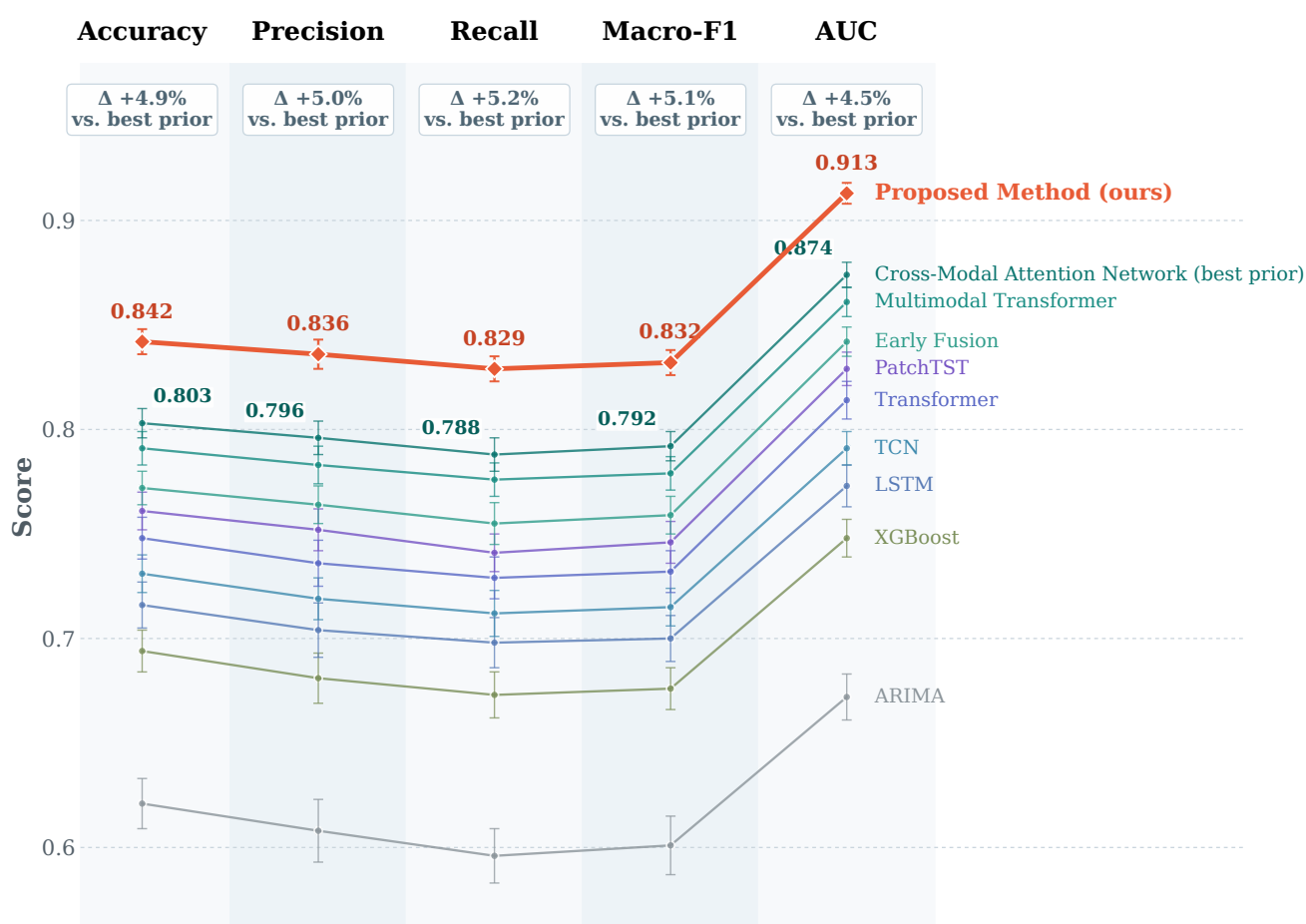


Figure 4. Performance comparison of different methods for short-term cross-border direction prediction.

4.3. Short-Term Volatility Forecasting and Cross-Border Risk Prediction

This experiment was designed to evaluate the comprehensive performance of different models for short-term volatility forecasting and cross-border risk prediction, with particular emphasis on their capability to model continuous state variations and identify high-risk events. Compared with direction prediction, volatility forecasting focuses more strongly on uncertainty and the magnitude of price disturbances, whereas risk prediction requires ab-

normal states to be identified at an early stage. Therefore, this experiment provides a further assessment of the ability of different methods to perceive complex cross-border shocks.

As shown in Table 3 and Figure 5, the experimental results show that ARIMA exhibited relatively limited performance in both tasks. For volatility forecasting, its MAE, RMSE, and MAPE were 0.084, 0.126, and 18.7, respectively, while the macro-F1 and AUC values for risk prediction reached only 0.584 and 0.651. This limitation is mainly attributable to the linear stationarity assumptions of ARIMA, where future variations are inferred from autocorrelation structures within historical sequences. Consequently, non-stationary volatility induced by sudden policy changes, logistics anomalies, and cross-border supply-chain disruptions cannot be effectively accommodated. XGBoost improved predictive performance through nonlinear ensembles of decision trees, increasing R^2 to 0.681 for volatility forecasting and AUC to 0.721 for risk prediction. These results indicate that tree-based feature partitioning can capture complex mappings between variables and risk states. However, because static feature representations are primarily employed, the dynamic evolution of risk caused by temporally propagating cross-border events remains difficult to model. LSTM further improved performance, reducing MAE to 0.066 and increasing risk AUC to 0.762. Its gated recurrent structure enables long-term historical information to be retained, thereby providing improved modeling capability for continuous volatility. TCN achieved further improvements, with MAE decreasing to 0.061 and AUC increasing to 0.791. Its causal and dilated convolutional structure enables local fluctuation patterns and longer-term temporal trends to be extracted simultaneously, resulting in stronger performance for short-term risk state identification.

Table 3. Performance comparison on short-term volatility forecasting and cross-border risk prediction. Results are reported as mean $\pm$ 95% confidence interval.

Method	MAE $\downarrow$	RMSE $\downarrow$	MAPE $\downarrow$	R^2 $\uparrow$	Macro-F1 $\uparrow$	AUC $\uparrow$	PR-AUC $\uparrow$	Recall $\uparrow$	FPR $\downarrow$
ARIMA	0.084 $\pm$ 0.004	0.126 $\pm$ 0.006	18.7 $\pm$ 1.2	0.612 $\pm$ 0.015	0.584 $\pm$ 0.012	0.651 $\pm$ 0.010	0.562 $\pm$ 0.011	0.571 $\pm$ 0.013	0.218 $\pm$ 0.009
XGBoost	0.072 $\pm$ 0.003	0.108 $\pm$ 0.005	15.4 $\pm$ 1.0	0.681 $\pm$ 0.012	0.643 $\pm$ 0.011	0.721 $\pm$ 0.009	0.631 $\pm$ 0.010	0.636 $\pm$ 0.012	0.185 $\pm$ 0.008
LSTM	0.066 $\pm$ 0.003	0.097 $\pm$ 0.004	13.9 $\pm$ 0.9	0.724 $\pm$ 0.011	0.681 $\pm$ 0.010	0.762 $\pm$ 0.008	0.674 $\pm$ 0.009	0.669 $\pm$ 0.010	0.162 $\pm$ 0.007
TCN	0.061 $\pm$ 0.003	0.091 $\pm$ 0.004	12.8 $\pm$ 0.8	0.758 $\pm$ 0.010	0.703 $\pm$ 0.009	0.791 $\pm$ 0.008	0.701 $\pm$ 0.008	0.694 $\pm$ 0.009	0.148 $\pm$ 0.007
Transformer	0.056 $\pm$ 0.002	0.084 $\pm$ 0.004	11.5 $\pm$ 0.7	0.792 $\pm$ 0.009	0.731 $\pm$ 0.009	0.823 $\pm$ 0.007	0.732 $\pm$ 0.008	0.726 $\pm$ 0.009	0.132 $\pm$ 0.006
PatchTST	0.051 $\pm$ 0.002	0.076 $\pm$ 0.003	10.4 $\pm$ 0.6	0.824 $\pm$ 0.008	0.754 $\pm$ 0.008	0.846 $\pm$ 0.007	0.756 $\pm$ 0.007	0.749 $\pm$ 0.008	0.119 $\pm$ 0.006
Multimodal Transformer	0.046 $\pm$ 0.002	0.069 $\pm$ 0.003	9.3 $\pm$ 0.5	0.856 $\pm$ 0.007	0.781 $\pm$ 0.007	0.872 $\pm$ 0.006	0.784 $\pm$ 0.007	0.776 $\pm$ 0.007	0.103 $\pm$ 0.005
Proposed Method	0.038 $\pm$ 0.001	0.057 $\pm$ 0.002	7.8 $\pm$ 0.4	0.901 $\pm$ 0.006	0.824 $\pm$ 0.006	0.918 $\pm$ 0.005	0.836 $\pm$ 0.006	0.819 $\pm$ 0.006	0.081 $\pm$ 0.004

Bold indicates the best results.

Attention-based architectures exhibited further advantages in complex environments. Transformer increased the AUC for risk prediction to 0.823, primarily because self-attention establishes global relationships across different temporal positions and thereby captures long-range dependencies. PatchTST further achieved an MAE of 0.051 and an R^2 of 0.824. Its patch-based representation reduces the difficulty of long-sequence modeling and facilitates simultaneous learning of local volatility structures and long-term trends. After integrating trade, logistics, and textual event information, Multimodal Transformer achieved an MAE of 0.046 and an AUC of 0.872, indicating that external cross-border state information provides valuable complementary evidence for risk assessment. The proposed method achieved the best overall performance, with MAE, RMSE, and MAPE values of 0.038, 0.057, and 7.8 for volatility forecasting, together with an R^2 of 0.901. For risk prediction, macro-F1, AUC, PR-AUC, and recall reached 0.824, 0.918, 0.836, and 0.819, respectively, while FPR was reduced to 0.081. Compared with Multimodal Transformer, the proposed method improved volatility MAE and risk AUC by approximately 17.4% and 5.3%, respectively. These advantages are primarily derived from the multi-level dynamic modeling mechanism developed for cross-border risk formation. Conventional multimodal models generally assume that all input sources contribute with relatively stable reliability, whereas actual cross-border environments may involve delayed trade statistics, missing

logistics trajectories, and uncertain news events. Therefore, reliability and consistency modeling are employed to dynamically regulate information-source contributions and reduce interference from abnormal signals. Meanwhile, the asynchronous temporal interaction mechanism exploits differences in information propagation speed to identify lagged relationships among logistics variations, policy shocks, and responses, while the dynamic event perception module strengthens responsiveness to extreme-risk states. From a modeling perspective, attention mechanisms improve long-range dependency representation, whereas reliability constraints and event gating further restrict excessive responses induced by noisy or abnormal inputs. As a result, more stable discrimination between normal fluctuations and extreme-risk conditions is achieved, leading to improved volatility forecasting accuracy and more reliable cross-border risk prediction.



Figure 5. Performance comparison on short-term volatility forecasting and cross-border risk prediction.

4.4. Ablation Study of Different Components

The ablation experiment was designed to quantify the independent contribution of each component to short-term cross-border direction prediction and to verify the necessity of reliability modeling, consistency modeling, asynchronous temporal interaction, event perception, and context-aware fusion under complex cross-border conditions. Different degraded variants were constructed by removing individual components from the complete framework and comparing their performance with that of the full model.

As shown in Table 4 and Figure 6, the full model achieved the best performance, with accuracy, macro-F1, and AUC values of 0.842, 0.832, and 0.913, respectively, demonstrating clear complementarity among the different components. After reliability modeling was removed, performance decreased to an accuracy of 0.816, a macro-F1 of 0.804, and an AUC of 0.882, indicating that variations in data quality among heterogeneous information sources significantly affect cross-border prediction. Because trade statistics may be delayed, logistics trajectories may be incomplete, and news information may contain noise, direct dependence on raw modality inputs can increase sensitivity to low-quality signals. Reliability estimation therefore contributes to suppressing the negative influence of abnormal information. When consistency modeling was removed, the resulting performance remained

slightly higher than that obtained after removing reliability modeling, but accuracy and AUC still decreased to 0.821 and 0.887, respectively. This result demonstrates that cross-modal relational constraints are important for identifying genuine economic transmission relationships. When news, trade, logistics, and signals are mutually aligned, multisource information can provide reinforcing evidence; conversely, consistency assessment helps avoid misleading associations when abrupt changes are observed in only a single modality. When both reliability and consistency mechanisms were removed, the most substantial degradation was observed, with Accuracy decreasing to 0.802 and AUC to 0.871. This finding confirms the synergistic relationship between the two mechanisms and indicates that data quality and cross-modal relationships cannot be treated independently when addressing the complexity of cross-border information.

Table 4. Ablation study of different components in the proposed framework. Results are reported as mean $\pm$ 95% confidence interval.

Variant	Accuracy $\uparrow$	Macro-F1 $\uparrow$	AUC $\uparrow$
w/o Reliability Modeling	0.816 $\pm$ 0.007	0.804 $\pm$ 0.008	0.882 $\pm$ 0.007
w/o Consistency Modeling	0.821 $\pm$ 0.007	0.809 $\pm$ 0.007	0.887 $\pm$ 0.006
w/o Reliability and Consistency	0.802 $\pm$ 0.008	0.791 $\pm$ 0.009	0.871 $\pm$ 0.008
w/o Asynchronous Temporal Interaction	0.824 $\pm$ 0.007	0.813 $\pm$ 0.007	0.891 $\pm$ 0.006
w/o Event Perception	0.829 $\pm$ 0.006	0.818 $\pm$ 0.007	0.897 $\pm$ 0.006
w/o Context-Aware Fusion	0.834 $\pm$ 0.006	0.823 $\pm$ 0.006	0.904 $\pm$ 0.005
Full Model	0.842 $\pm$ 0.006	0.832 $\pm$ 0.006	0.913 $\pm$ 0.005

Bold indicates the best results.

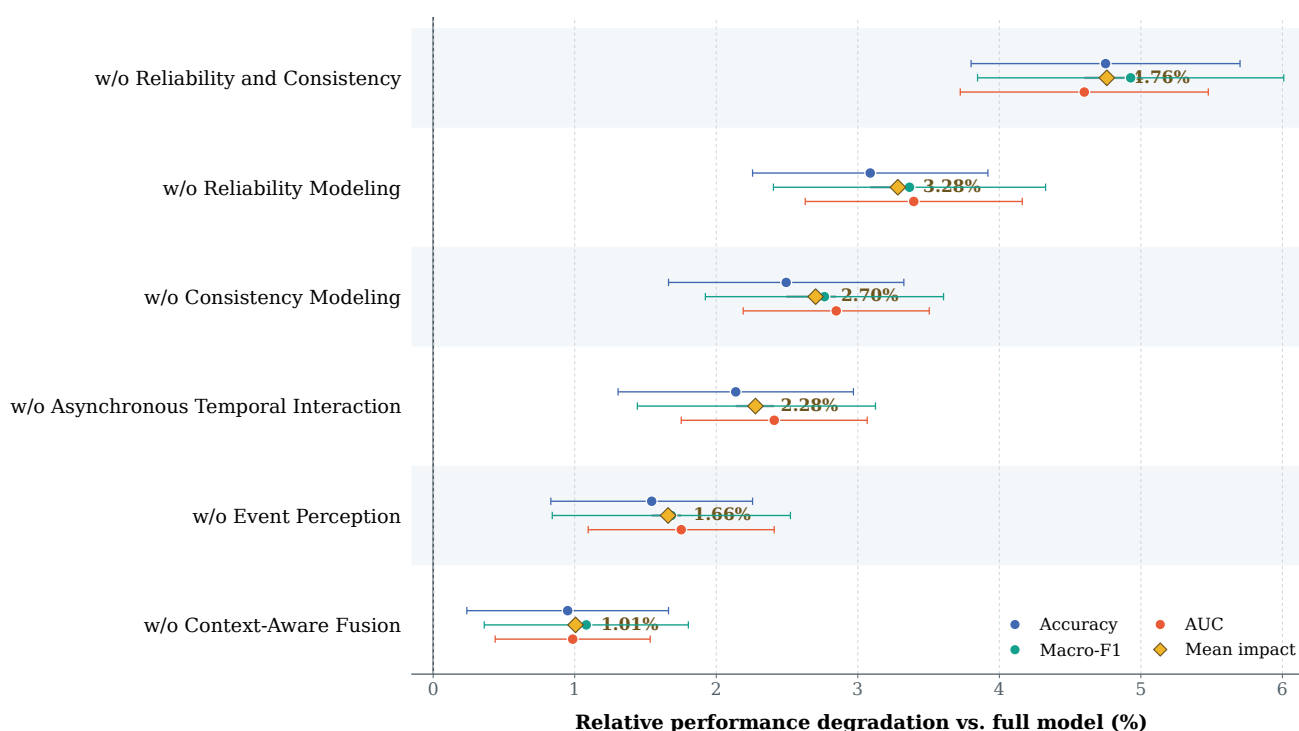


Figure 6. Impact analysis of different components through ablation study.

Removing the asynchronous temporal interaction module reduced accuracy to 0.824 and AUC to 0.891, demonstrating that heterogeneous cross-border information does not influence target simultaneously and that significant propagation delays exist among different signals. Conventional synchronous fusion may distort the actual temporal relationships among trade variation, logistics adjustment, policy release, and response, whereas asynchronous temporal modeling captures differences in propagation speed through explicit temporal dependencies. Removal of the event perception module reduced accuracy

and AUC to 0.829 and 0.897, respectively, indicating that sudden events such as tariff adjustments, trade restrictions, and port disruptions play an important role in short-term direction changes. Conventional time-series models mainly learn historical trends and may continue extrapolating normal dynamics under external shocks, whereas event perception allows abnormal disturbances to be identified and corresponding states to be corrected. Removing context-aware fusion produced a relatively smaller performance decline, with accuracy and AUC decreasing to 0.834 and 0.904, respectively, indicating that the relative importance of different modalities varies across regimes and that dynamic information weighting contributes to improved stability. Structurally, the reliability and consistency mechanisms reduce the influence of noisy modalities through dynamic weighting constraints, asynchronous temporal interaction strengthens long-range dependency representation through attention-based modeling, event perception limits and calibrates event-driven disturbances through gating mechanisms, and context-aware fusion further optimizes information combinations under different states. Consequently, the complete framework jointly addresses data credibility, temporal asynchrony, and external shocks in cross-border prediction, resulting in more stable and accurate performance.

4.5. Ablation Study on Data Augmentation Strategies

To systematically isolate and quantify the individual contributions of the training-time perturbation mechanisms described in Section 3.3, we conducted a comprehensive leave-one-out ablation study across all three forecast horizons ($h \in \{15, 30, 60\}$ minutes). Throughout all variants, the network backbone architecture, random seeds, temporal chronological cross-validation folds, early stopping criteria, and out-of-sample test splits were kept identical to ensure strict comparability. Specifically, the full augmentation pipeline was benchmarked against six single-component removal variants: (1) without modality masking, (2) without temporal masking, (3) without Gaussian feature noise, (4) without causal time shifting, (5) without cross-modal shuffling, and (6) without event-window oversampling. In addition, a completely unaugmented baseline (“w/o Any Augmentation”) was evaluated under the identical training schedule.

As summarized in Table 5, removing any single augmentation component leads to consistent degradation across direction prediction, volatility forecasting, and risk classification. Among individual perturbations, disabling cross-modal shuffling incurs the most noticeable drop in classification discrimination (e.g., direction macro-F1 dropping from 0.832 to 0.814 at $h = 15$), confirming its essential role in providing contrastive supervision for the consistency perception module. Eliminating causal time shifting degrades the model’s tolerance to cross-modal update delays, causing the 60 min volatility MAE to worsen from 0.046 to 0.052. Furthermore, removing event-window oversampling substantially harms risk prediction recall during abrupt turbulence periods, as evidenced by the drop in risk AUC across all horizons. Finally, stripping all augmentation operations simultaneously (“w/o Any Augmentation”) results in the most severe performance collapse across all tasks (direction accuracy falls to 0.796 at $h = 15$ min, and overall risk AUC drops to 0.865). These empirical results substantiate that our causally aligned multisource augmentation strategies synergistically prevent overfitting to high-frequency dominant signals and bolster model resilience under real-world data missingness and asynchronous shocks.

Table 5. Ablation results of data augmentation components across 15, 30, and 60 min prediction horizons. Results are reported as mean ± 95% confidence interval. Best results are marked in **bold**.

Augmentation Variant	15 min Horizon			30 min Horizon			60 min Horizon		
	Dir. Acc ↑	Vol. MAE ↓	Risk AUC ↑	Dir. Acc ↑	Vol. MAE ↓	Risk AUC ↑	Dir. Acc ↑	Vol. MAE ↓	Risk AUC ↑
w/o Modality Masking	0.828 ± 0.007	0.041 ± 0.002	0.903 ± 0.006	0.817 ± 0.007	0.046 ± 0.002	0.892 ± 0.006	0.803 ± 0.008	0.050 ± 0.003	0.879 ± 0.007
w/o Temporal Masking	0.831 ± 0.006	0.040 ± 0.002	0.907 ± 0.005	0.820 ± 0.007	0.044 ± 0.002	0.896 ± 0.006	0.806 ± 0.007	0.049 ± 0.002	0.883 ± 0.006
w/o Gaussian Noise	0.835 ± 0.006	0.040 ± 0.002	0.910 ± 0.005	0.824 ± 0.006	0.045 ± 0.002	0.898 ± 0.005	0.810 ± 0.007	0.048 ± 0.002	0.887 ± 0.006
w/o Causal Time Shifting	0.823 ± 0.007	0.042 ± 0.002	0.901 ± 0.006	0.812 ± 0.008	0.047 ± 0.003	0.889 ± 0.006	0.798 ± 0.008	0.052 ± 0.003	0.874 ± 0.007
w/o Cross-Modal Shuffling	0.821 ± 0.007	0.043 ± 0.002	0.898 ± 0.006	0.810 ± 0.008	0.048 ± 0.003	0.885 ± 0.006	0.795 ± 0.008	0.053 ± 0.003	0.871 ± 0.007
w/o Event Oversampling	0.833 ± 0.006	0.041 ± 0.002	0.895 ± 0.006	0.821 ± 0.007	0.046 ± 0.002	0.882 ± 0.006	0.808 ± 0.007	0.049 ± 0.002	0.869 ± 0.007
w/o Any Augmentation	0.796 ± 0.009	0.048 ± 0.003	0.865 ± 0.008	0.785 ± 0.009	0.054 ± 0.003	0.852 ± 0.008	0.771 ± 0.010	0.059 ± 0.003	0.839 ± 0.008
Full Pipeline (Ours)	0.842 ± 0.006	0.038 ± 0.001	0.918 ± 0.005	0.831 ± 0.006	0.042 ± 0.002	0.907 ± 0.005	0.819 ± 0.006	0.046 ± 0.002	0.896 ± 0.005

4.6. Discussion

The proposed framework for short-term cross-border prediction is not limited to improving predictive accuracy; more importantly, it also attempts to establish explicit relationships between variations and actual cross-border economic activities. In practical international trade scenarios, price movements are rarely driven by individual variables and are instead jointly influenced by commodity flows, logistics conditions, policy environments, and investor expectations. For example, in international energy, agricultural commodity, and industrial raw material, port congestion or shipping route disruptions in major exporting countries may first be reflected in reduced vessel arrivals and increased transportation delays before noticeable price increases or volatility amplification occur. Conventional price-series models generally observe only the subsequent outcome and provide limited insight into the underlying supply-side changes. By integrating AIS vessel trajectories, port operational states, and trade-flow information, potential supply-chain pressure can be perceived before corresponding prices fully respond, thereby providing earlier decision support for inventory adjustment, procurement planning, and risk management. Different types of information also exhibit substantially different propagation speeds in real cross-border trade operations. Taking international agricultural trade as an example, an export restriction imposed by a major exporting country may affect sentiment within several hours through news and policy channels, while exchange-rate fluctuations may rapidly change import costs. By contrast, actual variations in trade quantities and logistics operations usually require a longer period before supply-side effects become observable. The asynchronous cross-border temporal interaction mechanism avoids treating these signals as simultaneous observations and instead identifies their actual influence processes according to heterogeneous propagation pathways. Such a mechanism is more consistent with practical transmission processes involving policy changes, logistics adjustments, supply expectation revisions, and subsequent price responses, thereby enhancing not only predictive performance but also economic interpretability. Data reliability is equally important in practical applications for institutions and cross-border enterprises. AIS information may temporarily become unavailable because of equipment failures, trade statistics may be published with substantial delays, and international news sources may contain unverified or uncertain information. If all sources are treated as equally credible, predictions can easily be distorted by abnormal signals. Reliability- and consistency-aware modeling allows the framework to determine whether heterogeneous information sources mutually support each other. For example, when a commodity supply disruption is reported by news sources while vessel traffic simultaneously decreases, port congestion increases, and related prices begin to respond, the corresponding event can be assigned greater importance. In contrast, when an isolated news report is inconsistent with logistics states, its contribution can be appropriately reduced. Such behavior is more consistent with real-world risk analysis practices.

4.7. Limitation and Future Work

Although a multimodal cross-border prediction framework integrating exchange-rate, trade, logistics, and policy-event information has been established, several issues remain to be further investigated. First, the current dataset mainly consists of publicly accessible international information, trade statistics, logistics trajectories, and policy-event data. Although these sources can represent major cross-border economic activity variations, substantial differences remain in data accessibility across countries and regions, and data completeness is still limited for some emerging markets and local trade corridors. In addition, cross-border markets are jointly influenced by macroeconomic conditions, industrial cycles, and changes in international relations. The current framework mainly focuses on short-term

dynamic propagation processes, while longer-term structural changes remain insufficiently explored. Future research can further incorporate more real-time and fine-grained cross-border sensing information, including port Internet of Things devices, operational data from supply-chain enterprises, and real-time transportation states, thereby improving sensitivity to localized trade anomalies and supply-chain disruptions. Cross-market causal relationship modeling can also be further investigated so that variations can not only be predicted but the influence pathways among different countries, commodities, and industrial chains can also be more deeply interpreted. In addition, as privacy requirements for global trade data continue to increase, privacy-preserving computation and federated learning can be integrated to support cross-institutional and cross-regional collaborative modeling without directly sharing raw commercial data, thereby enabling the development of a more open, secure, and reliable global cross-border intelligence framework.

5. Conclusions

Short-term cross-border prediction plays a crucial role in global supply-chain coordination, international trade risk management, and decision-making. With the increasingly strong interactions among global trade activities, logistics networks, exchange-rate fluctuations, and policy events, conventional prediction approaches that rely solely on historical price sequences are insufficient to fully capture real-time state transitions. To address the challenges of heterogeneous multisource data, varying information reliability, asynchronous cross-modal propagation, and difficult modeling of sudden event impacts, a reliability- and consistency-aware multimodal cross-border prediction framework is proposed. Conditions, exchange rates, trade flows, logistics states, and news-policy events are uniformly regarded as cross-border sensing signals, enabling end-to-end intelligent analysis from data perception and information filtering to dynamic interaction and prediction. The main innovations are achieved through the construction of three key modules. First, a reliability- and consistency-aware cross-border representation module is designed, where the influence of low-quality or conflicting information is reduced by dynamically evaluating missing patterns, update delays, noise levels, and cross-modal matching relationships among different information sources. Second, an asynchronous cross-border temporal interaction module is developed, where continuous-time representation and multi-scale lag modeling are introduced to capture the actual propagation pathways among trade, logistics, exchange rates, and states, allowing instantaneous impacts and delayed effects of different signals to be effectively distinguished. Furthermore, a dynamic trade and policy event perception module is constructed to jointly analyze tariff adjustments, trade restrictions, port disruptions, and international news events with current conditions, thereby achieving adaptive responses to non-stationary shocks. Experimental results demonstrate that the proposed framework achieves superior performance in short-term cross-border direction prediction, obtaining an accuracy of 84.2%, precision of 83.6%, recall of 82.9%, macro-F1 of 83.2%, and AUC of 91.3%, which significantly outperform conventional statistical models, machine learning approaches, and existing multimodal deep learning methods. In volatility forecasting and risk identification tasks, the proposed model further achieves an MAE of 0.038, an RMSE of 0.057, an R^2 of 0.901, and an AUC of 0.918, while reducing the risk prediction FPR to 0.081. Ablation experiments further verify the essential contributions of reliability modeling, consistency constraints, asynchronous temporal interaction, and event perception mechanisms to improving model stability and generalization capability. The results demonstrate that artificial intelligence-driven multisource sensing methods can effectively establish connections among trade activities, logistics operations, and dynamics, providing a new technological pathway for intelligent trade monitoring, supply-chain-risk early warning, and cross-border decision support.

Author Contributions: Conceptualization, Y.W., X.W., X.Z. and M.L.; Methodology, Y.W., X.W. and X.Z.; Software, Y.W., X.W. and X.Z.; Validation, S.C. and B.L.; Formal analysis, S.C. and B.L.; Investigation, S.C. and B.L.; Resources, W.Y.; Data curation, W.Y.; Writing—original draft, Y.W., X.W., X.Z., S.C., B.L., W.Y. and M.L.; Visualization, W.Y.; Supervision, M.L.; Project administration, M.L.; Funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

- Dehnavi, M.K.; Dehnoi, M.K.; Amiri, H.A. Economics, Finance, and International Trade: A Nexus of Global Prosperity. In Proceedings of the 1st International Conference on Management, Accounting, Industrial Engineering and Economics, Wolfsburg, Germany, 6 March 2025.
- Chen, W.; Rojniruttikul, N. Developing a machine learning forecasting framework for exchange rates in the context of cross-border business. *Int. J. eBus. eGov. Stud.* **2025**, *17*, 123–148.
- Rogerson, S.; Svanberg, M.; Santén, V. Supply chain disruptions: Flexibility measures when encountering capacity problems in a port conflict. *Int. J. Logist. Manag.* **2022**, *33*, 567–589. [\[CrossRef\]](#)
- Zhao, X. Application of AI Models Integrating Multi-Source Data in Risk Early Warning and Dynamic Optimization of Smart Supply Chain in Cross-Border E-Commerce. *Procedia Comput. Sci.* **2026**, *281*, 1173–1182. [\[CrossRef\]](#)
- Yang, Z.; Zhang, Y. Enhancing cross-border forest product trade forecasting with LSTM and multi-source data fusion. *Discov. Artif. Intell.* **2025**, *5*, 342. [\[CrossRef\]](#)
- Eskandari, B.; Hudaverdi, B. A hybrid learning approach integrating ARIMA-GARCH-LSTM for value-at-risk forecasting. *Int. J. Financ. Eng.* **2026**, *13*, 2650011. [\[CrossRef\]](#)
- Guo, Z. The marketing of cross-border e-commerce enterprises in foreign trade based on the statistics of mathematical probability theory. *Appl. Math. Nonlinear Sci.* **2023**, *8*, 273–282. [\[CrossRef\]](#)
- Sarker, B.; Abubakkar, M.; Bhuiyan, M.A.I.; Hassan, A.; Hasan, M.; Sharmin, F.; Subha, D.T.; Dalim, H.M.; Begum, M.; Islam, M.S. Leveraging machine learning for supply chain disruption prediction and risk assessment in US industries. *Int. J. Appl. Math.* **2025**, *38*, 2382–2410. [\[CrossRef\]](#)
- Goyal, A.; Verma, A.K.; Sengupta, R. External shocks, cross-border flows and macroeconomic risks in emerging market economies. *Empir. Econ.* **2022**, *62*, 2111–2148. [\[CrossRef\]](#)
- Ben Ameur, H.; Boubaker, S.; Ftiti, Z.; Louhichi, W.; Tissaoui, K. Forecasting commodity prices: Empirical evidence using deep learning tools. *Ann. Oper. Res.* **2024**, *339*, 349–367. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ren, L.; Jia, Z.; Laili, Y.; Huang, D. Deep learning for time-series prediction in IIoT: Progress, challenges, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 15072–15091. [\[CrossRef\]](#) [\[PubMed\]](#)
- Klymenko, V.; Dokienko, L.; Novalska, N.; Sokolova, O.Y. Harmonization of customs procedures in the interaction of transport modes as a factor for enhancing the efficiency of multimodal logistics chains. *Intellectualization Logist. Supply Chain Manag.* **2026**, *37*, 96–106. [\[CrossRef\]](#)
- Wang, C.; Yang, H.; Hu, X.; He, X.; Liu, J.; Liang, C.; Lim, M.K. Deciphering iron ore trade dynamics: Supply disruption risk propagation in global networks through an improved cascading failure model. *Resour. Policy* **2024**, *95*, 105157. [\[CrossRef\]](#)
- Bai, Y.; Asif, M. International trade market forecasting and decision-making system: Multimodal data fusion under meta-learning. *PeerJ Comput. Sci.* **2025**, *11*, e3120. [\[CrossRef\]](#) [\[PubMed\]](#)
- Patino, L.; Hubner, M.; King, R.; Litzenberger, M.; Roupioz, L.; Michon, K.; Szklarski, L.; Pegoraro, J.; Stoianov, N.; Ferryman, J. Fusion of heterogeneous sensor data in border surveillance. *Sensors* **2022**, *22*, 7351. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhao, M. Research on cross-border e-commerce product marketing recommendations based on multimodal signal fusion and deep learning technologies. In *Proceedings of the International Conference on Computer Vision, Pattern Recognition, and Detection (ICVPRD 2025)*; SPIE: Bellingham, WA, USA, 2026; Volume 14121, pp. 434–441.
- Gangwani, P.; Panthi, V. Leveraging multimodal data and deep learning for enhanced stock market prediction. In *AI-Based Advanced Optimization Techniques for Edge Computing*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2025; pp. 93–127. [\[CrossRef\]](#)

18. Zhang, Y.; Fu, C.; Wang, X.; Zhang, Y.; Xiong, Z.; Pan, J.; Yin, J. AI-Driven Sensing for Cross-Lingual Risk Prediction via Semantic Alignment and Multimodal Temporal Fusion. *Appl. Sci.* **2026**, *16*, 3741. [\[CrossRef\]](#)
19. Wu, R.; Wang, H.; Chen, H.T.; Carneiro, G. Deep multimodal learning with missing modality: A survey. *arXiv* **2024**, arXiv:2409.07825.
20. Sun, H.; Zhang, J.; Hong, W.; Fang, Y.; Ma, M.; Shi, K.; Li, M. A Cross-Modal Temporal Alignment Framework for Artificial Intelligence-Driven Sensing in Multilingual Risk Monitoring. *Sensors* **2026**, *26*, 2319. [\[CrossRef\]](#) [\[PubMed\]](#)
21. He, J.; An, X. A Multitask Learning framework for E-commerce time series analytics: Operational risk, demand forecasting, and anomaly detection. *PLoS ONE* **2026**, *21*, e0352809. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Chen, L.; Fan, X. TIC-FusionNet: A multimodal deep learning framework with temporal decomposition and attention-based fusion for time series forecasting. *PLoS ONE* **2025**, *20*, e0333379. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Wang, L.; Hu, Y.; Zhou, Y. Cross-border commodity pricing strategy optimization via mixed neural network for time series analysis. *arXiv* **2024**, arXiv:2408.12115.
24. Yao, X. Demand Forecasting for Cross-border Retail Commodities in the New Retail Era: A Hybrid Model Integrating Temporal Granularity and Categorical Dynamics. In Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security, Hangzhou, China, 21–23 February 2025; pp. 498–504.
25. Zhou, K. Financial model construction of a cross-border e-commerce platform based on machine learning. *Neural Comput. Appl.* **2023**, *35*, 25189–25199. [\[CrossRef\]](#)
26. Zhao, Y.; Guo, Y.; Wang, X. Hybrid LSTM–Transformer architecture with multi-scale feature fusion for high-accuracy gold futures price forecasting. *Mathematics* **2025**, *13*, 1551. [\[CrossRef\]](#)
27. Mach, B.N.; Nguyen, H.C.; Nguyen, T.Q. Enhancing long-term forex market forecasting using transformer-based time series models: A comparative analysis with ensemble tree methods. *Res. Sq.* **2025**. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Fu, L. Forecasting Models for Predicting Global Supply Chain Disruptions in Trade Economics. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 899–906. [\[CrossRef\]](#)
29. Wan, X.; Zhao, Q. Research on Timeliness Prediction and Reliability Analysis of Cross-border Logistics. *J. Mod. Bus. Econ.* **2025**, *2*. [\[CrossRef\]](#)
30. Zhang, P.; Ma, A. Harmonized Development of International Trade in Logistics Services and Ecological Environment. *Int. J. Agric. Environ. Inf. Syst. (IJAEIS)* **2025**, *16*, 1–20. [\[CrossRef\]](#)
31. Hiraide, T.; Hanaoka, S.; Matsuda, T. The efficiency of document and border procedures for international trade. *Sustainability* **2022**, *14*, 8913. [\[CrossRef\]](#)
32. Hasan, M.R.; Rahman, M.A.; Gomes, C.A.H.; Nitu, F.N.; Gomes, C.A.; Islam, M.R.; Shawon, R.E.R. Building robust AI and machine learning models for supplier risk management: A data-driven strategy for enhancing supply chain resilience in the USA. *Adv. Consum. Res.* **2025**, *2*, 1152–1171.
33. Dong, S.; Yao, W.; Zhu, Y. Multimodal Data Fusion and Intelligent Decision Support for Digital Trade Industry Clusters. In Proceedings of the 2025 International Conference on Economic Management and Big Data Application, Shenzhen, China, 25–27 July 2025; pp. 837–843.
34. Yin, J.; Lu, Z.; Xiong, B.; Sun, K.; Liu, R.; Liu, Y.; Li, M. Multisource Port Inspection Sensor Fusion with Causal Representation Learning for Cross-Border Anomaly Monitoring. *Sensors* **2026**, *26*, 4142. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Chen, T.; Qiu, Y.; Wang, B.; Yang, J. Analysis of effects on the dual circulation promotion policy for cross-border e-commerce B2B export trade based on system dynamics during COVID-19. *Systems* **2022**, *10*, 13. [\[CrossRef\]](#)
36. Rong, K.; Ling, Y.; Yang, T.; Huang, C. Cross-border data transfer: Patterns and discrepancies. *J. Int. Bus. Policy* **2025**, *8*, 10–32. [\[CrossRef\]](#)
37. Li, Y.; Wei, Z. Regional logistics demand prediction: A long short-term memory network method. *Sustainability* **2022**, *14*, 13478. [\[CrossRef\]](#)
38. Zhang, X.; Li, Y.; Shi, H.; Liu, R.; Jiang, W.; Wang, S.; Li, M. Risk-Aware Multimodal Sensing Network with Asynchronous Temporal Alignment and Predictive Uncertainty Estimation. *Appl. Sci.* **2026**, *16*, 7540. [\[CrossRef\]](#)
39. Otunba, M.; Adenuga, A.A.; Sikiru, A.O.; Gaffar, O. A Financial Integration Framework for Cross-Border M and A in Banking and Financial Services: A Multi-Continent Approach. *Int. J. Sci. Res. Humanit. Soc. Sci.* **2024**, *1*, 275–296.
40. Vidha, M.; Bahroun, Z.; Saihi, A.; Ahmed, V. Tariff Shocks and Global Supply Chains: Network Propagation, Logistics Resilience, and System Reconfiguration—A Systematic Review. *IEEE Access* **2026**, *14*, 91479–91505. [\[CrossRef\]](#)
41. ISO 3166-1:2020; Codes for the Representation of Names of Countries and Their Subdivisions—Part 1: Country Code. International Organization for Standardization: Geneva, Switzerland, 2020.
42. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
43. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

44. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
45. Lea, C.; Flynn, M.D.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal convolutional networks for action segmentation and detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 156–165.
46. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
47. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv* **2022**, arXiv:2211.14730.
48. Snoek, C.G.; Worring, M.; Smeulders, A.W. Early versus late fusion in semantic video analysis. In Proceedings of the 13th Annual ACM International Conference on Multimedia, Singapore, 6–11 November 2005; pp. 399–402.
49. Tsai, Y.H.H.; Bai, S.; Liang, P.P.; Kolter, J.Z.; Morency, L.P.; Salakhutdinov, R. Multimodal transformer for unaligned multimodal language sequences. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 6558–6569.
50. Lu, J.; Yang, J.; Batra, D.; Parikh, D. Hierarchical question-image co-attention for visual question answering. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016.

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.