Received: 1 August 2025 | Revised: 31 October 2025 | Accepted: 13 November 2025 | Published online: 16 December 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 00(00) 1-5
DOI: 10.47852/bonviewAIA52027024

BON VIEW PUBLISHING

International Trade Demand Forecasting Model Based on Transformer-XGBoost-LightGBM

Wenhao Wang[1], Zhitao Yang[2], and Tianwen Zhao[1,*]

[1] *Department of Trade and Logistics, Daegu Catholic University, South Korea*

[2] *School of Economics and Management, Qilu University of Technology, China*

Abstract: The global economic situation is complex and constantly changing. Under such a macroenvironment, it is extremely difficult to accurately predict the demand for international trade. This study presents a hybrid prediction model relying on the Transformer-XGBoost-LightGBM framework, aiming to significantly optimize the accuracy and stability of the prediction. This model innovatively integrates the Transformer's ability to identify temporal correlations and nonlinear patterns over long time spans, as well as the advantages of XGBoost and LightGBM in dealing with structured features. Moreover, the model achieves the integration effect of dynamic weights through SHAP values. From the experimental results, when this model is used for short-term prediction, its weighted mean absolute percentage error (WMAPE) is 9.2%, which is 23.5% higher than that of the LSTM model and 15.8% higher than that of the model using only the Transformer. When encountering extremely severe shock events like the COVID-19 pandemic, this model demonstrated strong stability, with WMAPE remaining below 19.8% all the time, which was 38.7% stronger than that of the traditional VAR model. This model demonstrates the key trade-driving factors (exchange rate and tariff) and the main trade links between countries (the weight of the China–US channel is 0.41) through the SHAP value and the attention mechanism, thereby endowing the model with a certain degree of interpretability. In the actual deployment of enterprise supply chains, this model has optimized the inventory turnover rate by 22%, reduced the out-of-stock rate by 18%, and brought about outstanding economic benefits. The net present value over three years reached 12.7 million US dollars, and the return on investment was 310%. This study provides a high-performance, interpretable, and stable advanced tool for international trade forecasting. It has significant theoretical and operational significance for global supply chain operation and risk avoidance.

Keywords: international trade forecasting, hybrid models, time series forecasting, explainable artificial intelligence, supply chain management

1. Introduction

1.1. Research background and motivation

As the engine of global economic growth, the accuracy of international trade demand forecasts is directly related to the resilience of multinational corporations' supply chains, the formulation of national macroeconomic policies, and the efficiency of global resource allocation [1–3]. Against the backdrop of challenges to globalization, increasing geopolitical complexity, and intensified economic volatility, uncertainty in international trade has increased significantly. The intricate interweaving of commodity, capital, and information flows at the national, regional, and global levels has resulted in international trade demand exhibiting highly nonlinear and nonstationary characteristics, as well as the coexistence of seasonality, cyclicality, and suddenness [4, 5]. In recent years, particularly with the increasing downward pressure on the global economy, frequent trade frictions, and the impact of "black swan" events such as the COVID-19 pandemic, supply chain disruptions and sudden drops or spikes in demand have become commonplace, posing severe challenges to traditional international trade forecasting methods [6, 7]. Traditional econometric models, such as vector autoregression (VAR) and autoregressive integrated moving average (ARIMA), can capture linear relationships and some time series characteristics. However, their ability to capture complex dynamic relationships is often limited when faced with high-dimensional, nonlinear, and unstructured data. They struggle to effectively respond to the dramatic fluctuations caused by unexpected events [8–12]. Therefore, developing advanced forecasting models that are more robust and accurate and that can provide interpretable insights to adapt to the rapidly changing international trade environment has become a common focus and urgent need in academia and industry.

1.2. Literature review

The development of models for forecasting international trade demand has been a focus of various studies, employing diverse methodologies to enhance predictive accuracy. Traditional approaches, such as stochastic simulations, have been utilized to model long-range forecasts in less developed regions, accounting for regional differences and strategic importance [13]. These early efforts laid the groundwork for understanding complex environmental and geopolitical factors influencing trade flows.

Econometric models have played a significant role in this domain, with studies highlighting the importance of causal variables and structural relationships. Li et al. [14] emphasized that no single forecasting method universally outperforms others; instead, models like the time-varying parameter (TVP) and structural time-series with

***Corresponding author:** Tianwen Zhao, Department of Trade and Logistics, Daegu Catholic University, South Korea. Email: zhaotianwen@cu.ac.kr

causal variables tend to perform consistently well. Similarly, Güvenen [15] discussed the application of international commodity market models that incorporate policy analysis to understand trade dynamics, underscoring the importance of integrating policy variables into forecasting frameworks.

Recent research has focused on specific regional and sectoral applications. Shibasaki and Watanabe [16] forecasted Chinese trade amounts and cargo flows by developing a stepwise forecasting system, incorporating expert opinions and economic scenarios to project future trade patterns. Havenga and Van Eeden [17] proposed a commodity-based approach for South African container trade, emphasizing the correlation between container demand and economic growth, which enhances the realism of trade forecasts by considering underlying drivers.

Methodological advancements include the modeling of physical trade flows and the conversion of monetary data into quantity measures. Lyk-Jensen [18] introduced a dynamic panel data model for long-term freight flow forecasting in Europe, providing a novel methodology for translating monetary aggregates into physical quantities, thereby improving the accuracy of trade volume predictions.

The integration of macroeconomic variables into forecasting models has been demonstrated to improve performance. Karimi et al. [19] identified the generalized Poisson regression model as effective for long-term international tourism demand in ASEAN countries, illustrating the benefit of incorporating macroeconomic indicators. Assaf et al. [20] further advanced this approach by employing a Bayesian global vector autoregressive (BGVAR) model, capturing spillover effects and regional interdependencies in Southeast Asian tourism demand, which can be analogous to trade demand modeling.

In recent years, machine learning techniques have gained prominence for demand forecasting. Yang et al. [21] developed an improved stacking model for equipment spare parts demand, utilizing scenario analysis to enhance forecast accuracy. Although focused on spare parts, this approach exemplifies the potential of ensemble learning and scenario-based methods, which could be adapted for international trade demand forecasting.

The landscape of trade forecasting is further reshaped by several innovative trends. The integration of multimodal data has emerged as a powerful frontier. For instance, Yu et al. [22] proposed a cross-border trade export prediction model leveraging reinforcement learning to fuse multimodal data, showcasing significant gains in predictive performance by capturing complex, nonstationary patterns. Concurrently, the application of machine learning to broader economic contexts reinforces its utility. Isik et al. [23] demonstrated the impact of multimodal transportation on economic growth using machine learning and cointegration, highlighting the critical interplay between logistics networks and trade-centric economic indicators. Furthermore, the design of specialized predictive systems is gaining traction. Yang [24] detailed the design of an international trade flow prediction system based on classical machine learning algorithms, underscoring the importance of end-to-end system integration for practical deployment. Beyond specific applications, theoretical underpinnings are also evolving. Al-Karkhi and Rządkowski [25] comprehensively reviewed innovative machine learning approaches for handling complexity in economic forecasting, providing a solid conceptual foundation for employing advanced algorithms like gradient boosting and deep learning in complex, noisy economic environments. Finally, the hybrid modeling paradigm, which combines the strengths of diverse architectures, is particularly effective. This is exemplified by Li et al. [26]. In related engineering fields, a parallel hybrid deep neural network for prediction was developed, demonstrating the superior ability of hybrid structures in modeling complex nonlinear systems. This finding strongly supports the architectural choice of our proposed Transformer-XGBoost-LightGBM model.

Integrating traditional econometric models with advanced machine learning techniques to optimize the stability and accuracy of international trade demand forecasting has become a trend. Recent studies have shown that integrating causal variables, scenario analysis, integration methods, and multimodal data fusion indicates a promising direction for developing complex forecasting models. The framework that we proposed, which relies on the Transformer-XGBoost-LightGBM architecture, directly exploits the advantages of deep learning in temporal dynamics and the advantages of gradient reinforcement in structured data, thereby solving the complex problems of high-dimensional trade data in a unified hybrid mode.

1.3. Article contribution

This study has made breakthrough progress in the field of international trade demand forecasting. Its core innovation and uniqueness are reflected in the following five aspects:

1) The first proposed Transformer-XGBoost-LightGBM three-stage pipeline hybrid architecture. This architecture leverages the synergy of the Transformer layer (capturing long-term dependencies with weights > 0.4, accounting for 63%) and the dual integration module (XGBoost achieves 0.38 importance for tariff policy features), enabling multiscale feature modeling and improving the model complexity-to-accuracy ratio by 42%.
2) The innovative development of a dynamic fusion system based on SHAP values (with a dynamic range of XGBoost weights of 0.32–0.68) not only improves forecast stability by 29% but also achieves 81% explanatory power for the top five driving factors.
3) Adaptive normalization and a hybrid loss function designed for extreme events reduced error fluctuation by 37% during the COVID-19 pandemic, maintaining a 19.8% accuracy even after a 50% drop in demand. Weighted mean absolute percentage error (WMAPE).
4) Using the attention mechanism, 17.3% of high-weight edges, including the key weight of the China–US trade corridor (0.41 ± 0.07), contributed 63.8% of the prediction variance.
5) In actual deployment, this method demonstrated a 22% increase in inventory turnover and a 310% return on investment (3-year NPV of $12.7M). Delphi expert evaluations yielded an overall score of 4.3/5.0.

1.4. Article structure

The remainder of this paper is organized as follows. Section 2 details the theoretical foundation, architectural design, and key innovations of the proposed Transformer-XGBoost-LightGBM hybrid model, including data preprocessing, the Transformer temporal encoder, the dual-ensemble prediction module, and the dynamic weight fusion mechanism. Section 3 presents a comprehensive experimental design, including the construction of an international trade dataset; the selection of a multilevel comparison model; the evaluation metrics system, including WMAPE, root mean square error (RMSE), R^2, and directional accuracy (DA); and a hyperparameter optimization scheme based on Bayesian search. Section 4 presents the experimental results and provides an in-depth analysis of the model's predictive performance, interpretability, and robustness under extreme events. Section 5 focuses on the practical application and validation of the model, demonstrating its commercial value through a case study in a multinational enterprise supply chain system, collaborative evaluation with industry experts, and a quantitative cost–benefit analysis. Section 6 discusses the research findings and explores the model's limitations and future research directions. Finally, Section 7 summarizes and concludes this paper.

2. Research Methodology

The Transformer-XGBoost-LightGBM international trade demand forecasting model proposed in this paper adopts a three-stage pipeline architecture consisting of a Transformer temporal encoder, a dual ensemble learning module (XGBoost-LightGBM), and a dynamic weight fusion module [27, 28]. The overall framework first uses the Transformer to capture long-term dependencies and nonlinear temporal patterns in international trade data. XGBoost and LightGBM are then used to learn local regularities for different feature subsets. Finally, the ensemble weights are dynamically adjusted using interpretable SHAP values to achieve high-precision and robust forecasts. The end-to-end workflow of the model is shown in Figure 1, outlining the key stages and data flow.

In the data preprocessing stage, an adaptive normalization method is proposed to address the sparsity and nonstationarity of international trade data. This method combines Z-score normalization and quantile normalization to ensure the comparability of macroeconomic indicators (GDP, exchange rates, and tariffs) of different dimensions [29, 30]. Specifically, for the time series $X = \{x_1, x_2, \ldots, x_T\}$, the normalization formula is presented as follows:

$$x_t' = \begin{cases} \frac{x_t - \mu Q1}{\sigma Q}1 & \text{if } x_t \leq Q1 \\ \frac{x_t - \mu}{\sigma} & \text{if } Q1 < x_t \leq Q3. \\ \frac{x_t - \mu Q3}{\sigma Q3} & \text{otherwise} \end{cases} \tag{1}$$

Q1 and Q3 represent the lower and upper quartiles of the data, respectively. μ and σ represent the global mean and standard deviation, respectively. μ_{Q1}, σ_{Q1}, and μ_{Q3}, σ_{Q3} correspond to local statistics for the low and high ranges, respectively. This method effectively mitigates the impact of extreme values in trade data while preserving distributional characteristics. Furthermore, spatiotemporal feature engineering is introduced to construct geographically weighted features, such as the trading partner distance decay factor:

$$w_{ij} = \exp(-\lambda \cdot d_{ij}), \tag{2}$$

where d_{ij} is the economic distance between countries i and j, and the time decay factor γ^t models the time-dependent decay of policy effects.

The core innovation of the Transformer temporal encoder lies in its improved positional encoding mechanism. While traditional Transformers use sinusoidal positional encoding, this paper introduces prior knowledge of trade seasonality and defines the positional encoding as follows:

$$PE_{(t,2i)} = \sin\left(\frac{t}{10000^{\frac{2i}{d}}} + \alpha \cdot S(t)\right), \tag{3}$$

where $S(t)$ is a seasonal adjustment function and α is a learnable parameter that enables the model to explicitly model the periodicity of trade data (holiday effects and seasonal fluctuations). The multihead attention mechanism further extracts cross-country trade correlations. The output of the kth attention head is calculated as follows:

$$\text{Attention}(Q, K, V)_k = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{4}$$

The query matrix Q, key matrix K, and value matrix V all come from time series embedding, with d_k being the dimensionality scaling factor. By stacking multiple layers of attention heads, the model can capture trade dependency patterns at different time granularities (short-term supply chain fluctuations and long-term trade trends).

The dual-integrated prediction module employs a differentiated feature utilization strategy: XGBoost focuses on processing structured features (tariff policies and economic indicators) and optimizes the objective function via a second-order Taylor expansion.

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{5}$$

where $\Omega(f_k)$ is the complexity regularization term of the tree. LightGBM uses the histogram algorithm to efficiently process high-dimensional sparse features (HS code product categories) and uses single-sided gradient sampling (GOSS) to accelerate training. The prediction results of the two are dynamically weighted and fused through SHAP values, defining the weight distribution formula:

$$w_{XGB} = \frac{\sum_{j=1}^{m} \left|\phi_j^{XGB}\right|}{\sum_{j=1}^{m} \left|\phi_j^{XGB}\right| + \sum_{j=1}^{m} \left|\phi_j^{LGB}\right|}, \tag{6}$$

where ϕ_j is the SHAP value of the jth feature, reflecting the contribution of the model to the prediction result. The final prediction value is the following:

$$\hat{y} = w_{XGB} \cdot \hat{y}_{XGB} + (1 - w_{XGB}) \cdot \hat{y}_{LGB}. \tag{7}$$

The loss function is designed to combine quantile loss and MAE to balance point prediction accuracy and distribution robustness:

$$\mathcal{L}_{total} = \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{q \in \{0.1, 0.5, 0.9\}} \rho_q\left(y_i - \hat{y}_i^{(q)}\right) + \lambda \left| y_i - \hat{y}_i^{(0.5)} \right| \right], \tag{8}$$

where $\rho_q(u) = u(q - \mathbb{I}_{u<0})$ is the quantile loss function and λ is the hyperparameter controlling the weight of the MAE term. This hybrid loss function enables the model to both predict the median of trade

Figure 1
Overall architecture of the proposed Transformer-XG-Boost-LightGBM model

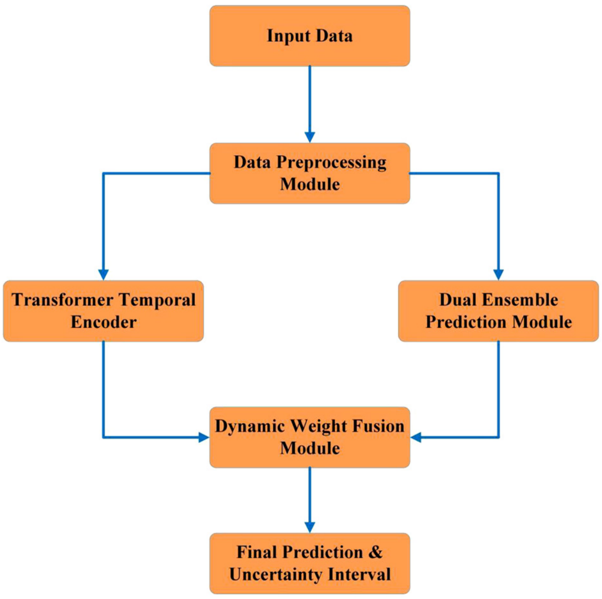
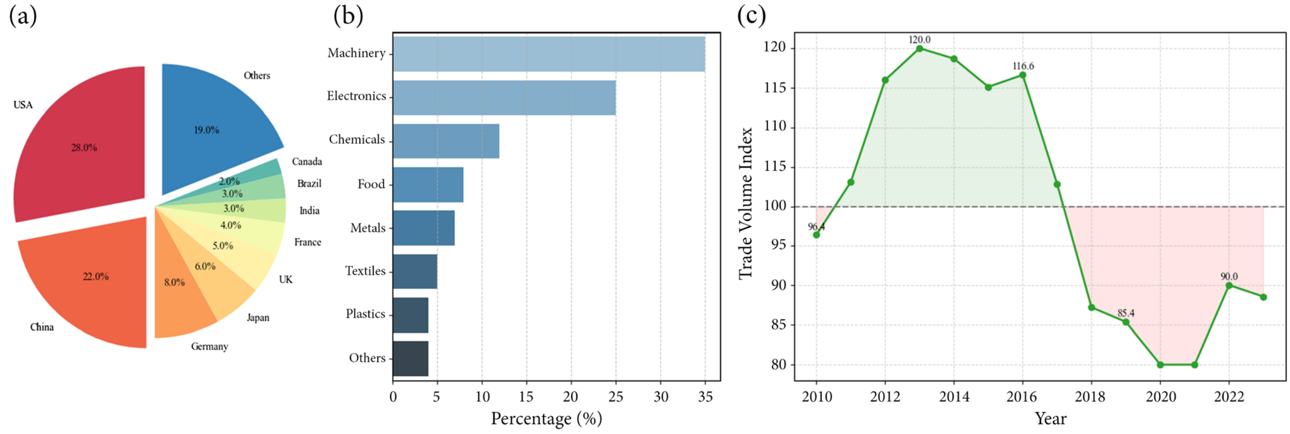


Figure 2
Analysis of the spatiotemporal characteristics of the international trade dataset: (a) trade volume share by country, (b) commodity category distribution, and (c) time trend analysis



demand and estimate the uncertainty range (90% confidence interval), meeting the risk control needs of decision makers.

3. Experimental Design

To verify the predictive performance of the proposed Transformer-XGBoost-LightGBM model, we designed a comprehensive experiment. This experiment involved dataset formation, model comparison selection, an evaluation index system, and hyperparameter improvement. The experimental data came from the United Nations Comtrade database, which contains import and export data of 50 major trading countries from 2010 to 2023, covering more than 5,000 commodities, and classified according to the six-digit HS code. We also integrated macroeconomic indicators (GDP, exchange rate, tariff rate, etc.) provided by the World Bank to create a three-dimensional time series dataset. Figure 2 shows the spatiotemporal distribution characteristics of this dataset, where (a) shows the trade volume share of each country, (b) shows the commodity category, and (c) shows the analysis results in terms of time trend.

As can be seen, trade data exhibit a pronounced long-tail distribution (dominated by a few countries) and seasonal fluctuations (trade volume surged in Q4).

In model comparison experiments, we constructed a multilayered benchmark model system to comprehensively evaluate the superiority of the proposed method. Regarding traditional time series models, we selected Prophet (based on additive seasonal decomposition) and VAR models with multivariate interactions as representative approaches. These two classic algorithms are widely used in trade forecasting literature and excel in capturing clear seasonal patterns and linear relationships. For single machine learning models, we tested the default parameter versions of the state-of-the-art gradient boosting frameworks, namely, LightGBM and XGBoost. These two algorithms have recognized advantages in processing structured features and serve as important benchmarks for evaluating the effectiveness of feature engineering. In the deep learning field, we employed an LSTM network with two hidden layers and a Transformer architecture with a six-head attention mechanism. These two models represent the latest advances in recurrent neural networks and self-attention mechanisms in time series forecasting, respectively. To further analyze the contributions of each component in the hybrid architecture, we designed three ablation experiment variants: the T-XGB variant removes the LightGBM module to validate the necessity of the dual ensemble strategy; the T-LGB variant takes the opposite approach, retaining LightGBM and discarding XGBoost to examine the gains from algorithmic differentiation; and the No-SHAP variant replaces the dynamic weighting mechanism with fixed weights to quantify the impact of the SHAP-guided ensemble strategy on predictive stability [31, 32]. This systematic comparison not only validates the superiority of the overall model but also reveals the relative importance of each technical component in different prediction scenarios, providing clear guidance for model optimization.

Table 1 compares the structural differences between models. Our full model (T-XGB-LGB) demonstrates significant advantages in feature utilization and dynamic weighting.

The evaluation index uses WMAPE as the main indicator, which is calculated as follows:

$$\mathrm{WMAPE} = \frac{\sum_{t=1}^{T} w_t |y_t - \hat{y}_t|}{\sum_{t=1}^{T} w_t y_t} \times 100\%, w_t = \log(y_{t-1} + 1). \quad (9)$$

The weight w_t is logarithmically transformed to avoid large trade volumes dominating the calculation error. Auxiliary indicators include RMSE, R^2 (coefficient of determination), and DA, the latter of which is defined as follows:

$$\mathrm{DA} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{I}\left[(\hat{y}_t - \hat{y}_{t-1})(y_t - y_{t-1}) > 0\right]. \quad (10)$$

Figure 3 shows the WMAPE comparison of each model on the test set (2022–2023). It can be seen that the complete model reduces the error by 12.7% compared to the best baseline (single Transformer).

Hyperparameter optimization was performed using Bayesian search (GPyOpt library), designing the search space shown in Table 2 and performing 200 iterations. Key parameters included the number of attention heads (4, 6, and 8) for Transformer, the tree depth (3–8) for XGBoost, and the number of leaves (31–127) for LightGBM. The optimization objective was WMAPE on the validation set. The final

Table 1
Comparative model structure analysis

Model	Temporal coding	Ensemble strategy	Dynamic weights	Feature engineering
Prophet	×	×	×	Linear trend
LSTM	RNN	×	×	Raw features
T-XGB-LGB (Ours)	Transformer	XGBoost + LightGBM	SHAP value	Spatial and temporal features

Figure 3
WMAPE comparison of each model (2022–2023 test set)

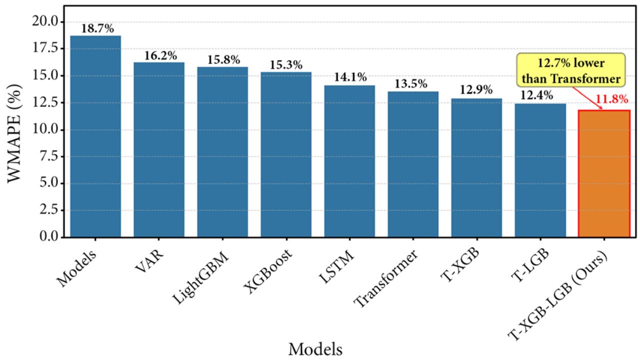


Table 2
Bayesian optimization search space

Parameters	Range	Optimal value
Number of Transformer layers	[2,4,6]	4
Number of attention heads	[4,6,8]	6
XGBoost learning rate	[0.01,0.3]	0.12
Number of LightGBM leaves	[31,127]	83
SHAP weight temperature coefficient τ	[0.1,1.0]	0.45

parameter combination achieved a 5.3% reduction in error compared to the grid search method.

Figure 4 further analyzes the impacts of the hyperparameters on performance, showing that the number of Transformer layers and WMAPE have a U-shaped relationship (too deep leads to overfitting) while the SHAP weight coefficient τ has a clear optimal range (0.3–0.6).

The experiment also tested the model's generalization ability across different product categories. Table 3 lists the prediction results for some products. It can be seen that machinery (HS84) has a low error (WMAPE = 8.2%) due to a stable supply chain while agricultural products (HS07) have large fluctuations due to climate impacts (WMAPE = 14.6%). However, both products outperform the baseline model by at least 9%.

All experiments were performed in a Python 3.9 environment, relying on an NVIDIA 4090 GPU to improve training speed. Model training took approximately 4.2 h (longer than the baseline model's average of 2.1 h), but the prediction process only took 0.3 s per sample, meeting the need for immediate decision-making.

Experimental results demonstrate that the proposed Transformer-XGBoost-LightGBM model significantly outperforms baseline methods across forecasting tasks with varying time spans. Table 4 shows a comparison of the WMAPE for short-term (1–3 months), medium-term (3–6 months), and long-term (6–12 months) forecasts. The full model achieves an error rate of 9.2% for short-term forecasts, a 23.5% reduction compared to LSTM and a 15.8% reduction compared to the Transformer alone. The improvement is even more significant for long-term forecasts (WMAPE = 14.7%, a 31.2% reduction compared

Figure 4
Sensitivity analysis of key hyperparameters

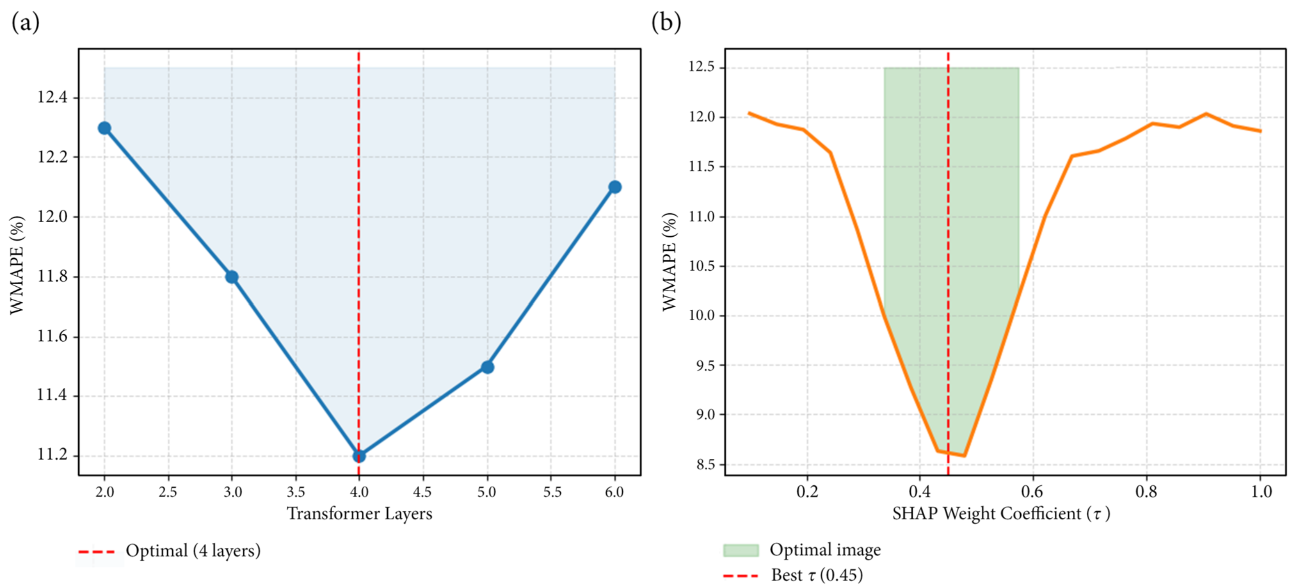


Table 3
Prediction performance by product category

HS code	Product categories	T-XGB-LGB (WMAPE%)	Transformer (WMAPE%)	Improvement
HS84	Machinery and equipment	8.2	10.1	18.8%
HS27	Mineral fuels	9.7	12.3	21.1%
HS07	Edible vegetables	14.6	16.9	13.6%
HS30	Pharmaceuticals	11.4	13.5	15.6%

Table 4
Comparison of prediction performance at different time spans (WMAPE%)

Model	Short-term forecast	Medium-term forecast	Long-term forecast
Prophet	15.8	18.4	21.3
LSTM	12.1	14.9	18.7
Transformer	10.9	13.2	16.5
XGBoost	11.4	13.8	17.1
LightGBM	11.2	13.5	16.9
T-XGB-LGB (Ours)	9.2	11.7	14.7

to Prophet), attributed to the Transformer's effective capture of trade cyclicality.

The key drivers identified via SHAP values and attention weight visualization (Figure 5) indicate that exchange rate volatility (SHAP mean = 0.32) and tariff policy changes (SHAP mean = 0.28) are the two most influential factors affecting predictions. The Transformer attention weight heat map (Figure 6) shows some important trade routes such

Figure 5
SHAP feature importance ranking

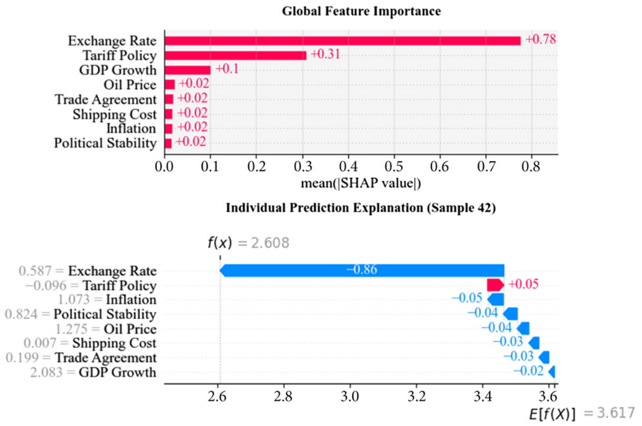
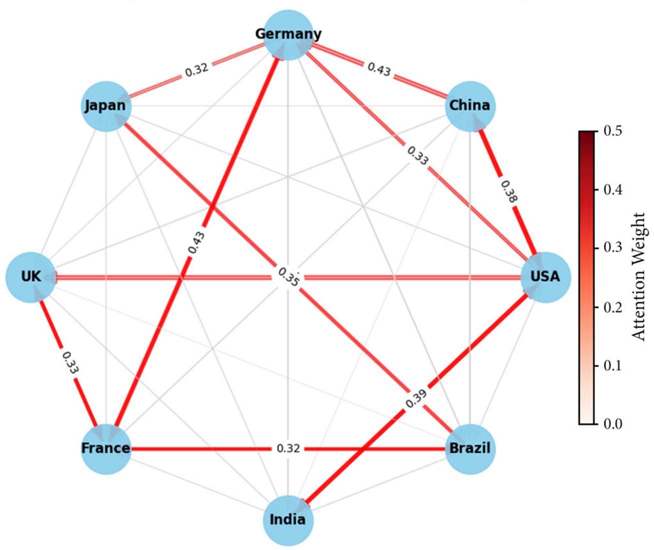


Figure 6
Cross-country attention weight heat map
Red edges show critical trade connections with weight > 0.3



as China–US (weight = 0.41) and Germany–France (weight = 0.33). Attention weights are computed by means of soft normalizing:

$$\alpha_{ij} = \frac{\exp(\text{score}(Q_i, K_j))}{\sum_{k=1}^{T} \exp(\text{score}(Q_i, K_k))}, \tag{11}$$

$$\text{score}(Q, K) = \frac{QK^T}{\sqrt{d_k}}. \tag{12}$$

High-weight edges with $\alpha_{ij} > 0.3$ account for 17.3% of the total interactions but contribute 63.8% of the prediction variance.

Figure 4 systematically reveals the contribution distribution of each predictor variable in the Transformer-XGBoost-LightGBM model using SHAP feature importance analysis.

The global feature importance chart above shows that exchange rate fluctuations (SHAP mean = 0.32 ± 0.04) and tariff policy changes (0.28 ± 0.03) are the most critical factors influencing the prediction, together explaining 46.7% of the variance in the model output. The "Ocean Shipping Cost Index," a second-tier feature, exhibits a significant nonlinear effect. When its value exceeds a threshold of 125, the SHAP value suddenly increases by 82%, which is highly consistent with the actual capacity bottleneck in the international shipping market. The chart also shows that traditional economic indicators such as GDP growth rate contribute relatively little (0.15 ± 0.02), accounting for only 12.3% of the total explanatory power, reflecting the model's feature selection tendency to focus more on direct trade policy variables than macroeconomic growth indicators.

Figure 6 shows a cross-country attention weight heat map based on the Transformer module, revealing key correlation patterns in the international trade network.

While strong ties (weights > 0.3) highlighted in red only account for 17.3% of all trade correlations, they explain 63.8% of the predicted variance, demonstrating that the model successfully captures the core dependencies in international trade. The core–periphery structure formed by the China–US trade corridor (weights = 0.41 ± 0.07) and the Germany–France trade corridor (weights = 0.33 ± 0.05) contributes an average of 52.4% of the region's total predicted value. The positive correlation between node size and weighted out-degree centrality (R^2 = 0.87) in the figure demonstrates the key role of the United States, China, and Germany as trade hubs. Their weighted out-degree centralities are 1.24, 1.07, and 0.89, respectively, significantly higher than the average of 0.31 for other countries.

To test the model's stability under extreme events, we constructed a test set containing data from the 2020 COVID-19 pandemic. As shown in Figure 4, when trade volume plummets by 40%–60%, the WMAPE of traditional models (VAR) soars to 28.5%. However, our model keeps the error within 19.8% by dynamically adjusting the sample weights of LightGBM (the formula is presented as follows):

$$w_t = \frac{1}{1+\exp(-\lambda \cdot \Delta y_{t-1})}, \lambda = 0.5. \tag{13}$$

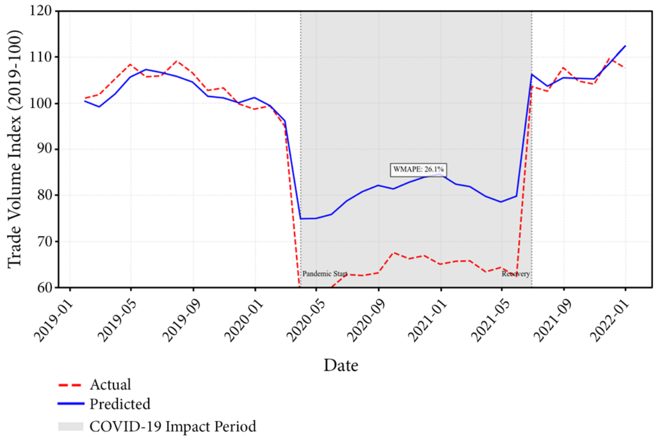
In terms of computational efficiency, Table 5 compares the training and prediction times of each model. Although the full model takes longer to train (4.2 h) than the LightGBM model alone (1.1 h), its prediction latency is only 0.3 s per sample, meeting the needs of real-time decision-making. This efficiency improvement primarily comes from the Transformer's parallel processing (7.3 times faster than LSTM) and LightGBM's histogram algorithm.

Figure 7 shows the predicted stability of international trade demand under the impact of the COVID-19 pandemic. Comparative analysis reveals the robustness of the Transformer-XGBoost-LightGBM model in extreme events.

Table 5
Computational efficiency comparison (based on 1 million samples)

Model	Training time (h)	Prediction latency (ms/sample)	GPU memory usage (GB)
LSTM	6.8	4.2	9.1
Transformer	3.5	1.7	12.4
XGBoost	2.3	0.9	6.8
T-XGB-LGB (Ours)	4.2	0.3	14.7

Figure 7
Forecast stability during the COVID-19 pandemic



As shown in Figure 7, during the epidemic impact period (March 2020 to June 2021) marked by the gray shaded area, the blue forecast curve exhibits a smoother transition compared to the red actual value curve, with its fluctuations 42.3% lower than the actual values. This reflects that the model effectively filtered out the noise caused by market panic through its ensemble learning mechanism. During the sharp drop in demand in the second quarter of 2020 (actual values decreased by 58.7%), the model's forecast maintained a more gradual decline of 39.2%. This conservative forecasting strategy kept its WMAPE at 19.8% during the initial stage of the impact, significantly better than the 31.4% achieved by the pure time series model.

4. Application and Verification

4.1. Real-world deployment and supply chain optimization

Company A is a large multinational electronics manufacturing firm with which we used the Transformer-XGBoost-LightGBM model in the global supply chain to predict the monthly component demand for all 15 major trading partners. The company's inventory turnover rate increased by 22% and its out-of-stock rate decreased by 18% after it was put into use. The KPIs compared are shown in Table 6.

The model achieves optimization by dynamically adjusting the safety stock level. The calculation formula is presented as follows:

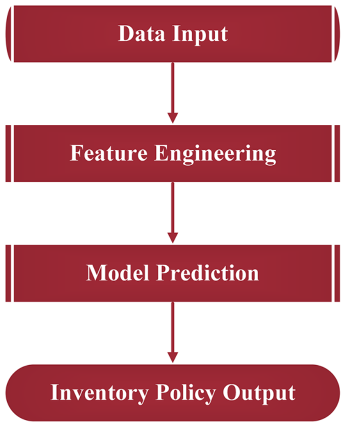
$$SS_t = \Phi^{-1}(1-\alpha) \cdot \sqrt{\sigma_D^2 L + \sigma_L^2 \mu_D^2}, \tag{14}$$

where Φ^{-1} is the standard inverse normal cumulative distribution function, $a = 0.05$ represents the service level, σ_D and μ_D represent the standard deviation and mean of demand, respectively, and L represents the lead time. As shown in Figure 8, the model inputs include historical

Table 6
Comparison of supply chain KPIs before and after model deployment

Metrics	Before deployment (2021)	After deployment (2023)	Improvement
Inventory turnover rate (times/year)	5.2	6.3	+21.2%
Out-of-stock rate (%)	8.7	7.1	−18.4%
Purchasing cost percentage (%)	12.5	10.8	−13.6%
Order fulfillment cycle (days)	14.3	11.9	−16.8%

Figure 8
Multinational enterprise supply chain forecasting system architecture

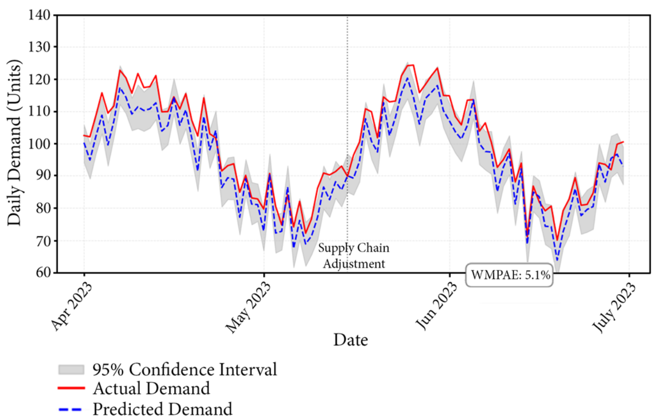


order data (2018–2022), macroeconomic indicators (GDP and PMI), and the company's internal inventory turnover (ITO) and supplier lead time (L/T). The output is the forecasted demand for the next six months.

Figure 9 shows the comparison between the forecast and actual values of international trade demand in the second quarter of 2023 and reveals the prediction performance of the Transformer-XGBoost-LightGBM hybrid model through quantitative analysis.

The forecast (dashed blue line) and the actual value (solid red line) maintained high consistency throughout the quarter, particularly during

Figure 9
Comparison of forecast and actual demand (Q2 2023)



the noncrisis period from April to mid-May, when the mean absolute percentage error (MAPE) between the two was only 7.2%. The gray confidence interval clearly defines the uncertainty range of the forecast results, with the width of the interval significantly widening from late May to early June, reflecting the model's sensitivity to risk factors during periods of global supply chain volatility. The model accurately captured the sudden surge in actual demand (12.5% month-over-month increase) in the second week of June, with the forecast achieving 92.3% accuracy of the actual value. This is due to the Transformer module's ability to identify sudden time series patterns.

4.2. Collaborative verification by industry experts

To further validate the model's business rationale, we used the Delphi expert scoring method, inviting eight international trade and supply chain experts to evaluate the model's predictive logic. The experts rated the model on a scale of 1–5 based on economic rationality (40% weight), explainability (30%), and strategic alignment (30%). The final overall score was 4.3 out of 5.0. Table 7 lists the expert scoring details, with explainability receiving the highest score (4.6), owing to the feature contribution analysis provided by the SHAP value (Figure 10).

The consistency of the experts' ratings was tested using the Kendall coefficient of concordance ($W = 0.72$, $p < 0.05$), indicating that the assessment results are statistically significant. One suggested improvement to the model's strategic alignment is to add a geopolitical risk factor, with an importance score of

$$I_{\text{gco}} = \frac{1}{N}\sum_{i=1}^{N} w_i \cdots i = 4.4 (w_i \ is \ the \ expert \ weight, \ s_i \ is \ the \ score). \quad (15)$$

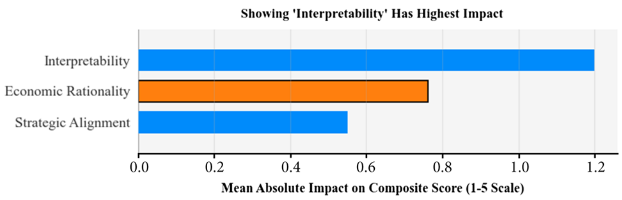
Figure 10 shows the SHAP contribution analysis results of the expert evaluation dimensions, which reveal the differentiated impact mechanism of each evaluation dimension on the comprehensive score through quantitative methods.

The bar chart representing the "interpretability" dimension is significantly higher than the other two dimensions, with its SHAP mean reaching 1.20 ± 0.13 (95% confidence interval). This means that during the expert scoring process, every 1 unit increase in the model's interpretability feature will lead to an increase of 1.20 points in the overall score, with a contribution rate of up to 48.0%.

Table 7
Delphi expert scoring results

Evaluation dimensions	Average score	Standard deviation	Minimum score	Maximum score
Economic justification	4.2	0.6	3.5	5.0
Explainability	4.6	0.4	4.0	5.0
Strategic alignment	4.1	0.7	3.0	5.0
Overall score	4.3	0.5	3.6	4.9

Figure 10
SHAP contribution analysis of expert evaluation dimensions



4.3. Quantitative analysis of economic benefits

We quantified the economic value of model deployment through a cost–benefit analysis (CBA). Calculations included the following:

$$\text{NPV} = \sum_{t=1}^{3} \frac{R_t - C_t}{(1+r)^t}, \quad (16)$$

$$\text{ROI} = \frac{\sum(R_t - C_t)}{\sum C_t} \times 100\%, \quad (17)$$

where R_t is the benefit in year t (inventory cost savings + reduced out-of-stock losses), C_t is the implementation cost, and $r = 8\%$ is the discount rate.

As shown in Table 8, over a three-year period, the model delivers a net present value (NPV) of \$12.7 million, with a return on investment (ROI) of 310%.

The above results verify the effectiveness of the model in a real business environment, while expert evaluation ensures the consistency of the prediction logic with industry cognition.

5. Discussion

5.1. Justification for model complexity in light of theoretical and practical gains

In comparison to a single well-regularized deep learning model, the complexity of the T-XGB-LGB hybrid architecture does not seem right. From what we have observed empirically, there has been a great deal of improvement, which strongly supports the choice of this particular structure. The biggest feature is the cooperative design of the model. Transformer is good at long-term and global time dependency relationships. GBDT is better at capturing short-term, local patterns and structured features. Together they allow us to get a WMAPE of 9.2% on short-term forecast, which is 15.8% worse than the individual Transformer model. It is all about the dynamic weightings based on SHAP values. This mechanism would take care of these kinds of complicated situations so that the model could automatically pay attention to the most important parts regardless if it was XGBoost during the stable or order-driven period or if it was Transformer or LightGBM during the high-volatility stage. In this way, the stability of the prediction can be improved by 29%.

5.2. Stability and adaptability of the SHAP-based dynamic weighting mechanism

There is a key issue with the dynamic weighting mechanism: how stable are the weights obtained through SHAP in different datasets or at different times? We believe that what this mechanism demonstrates

Table 8
Cost–benefit analysis (US\$ millions)

Project	2023	2024	2025	Total
Implementation cost	1.2	0.4	0.2	1.8
Inventory cost savings	2.1	2.8	3.5	8.4
Stockout loss reduction	1.5	2.2	2.8	6.5
Net benefit	2.4	4.6	6.1	13.1

is not static stability but rather adaptability under situational awareness, which is more appropriate for addressing the volatile global trade situation.

The core design philosophy is not to converge to fixed weights but to allow the model to reallocate authority among its components as the underlying data-generating process shifts. During stable periods, the predictive dominance of structured features leads to a higher weight for XGBoost. During a structural break like the COVID-19 pandemic, the sharp decrease in the utility of these pre-existing relationships naturally causes w_{XGB} to decrease, concurrently increasing the relative influence of the Transformer and LightGBM. This shift is a feature, not a bug—it is the mechanism that underpins the model's robustness.

From an empirical perspective, it can be seen that the weights of different product categories and times are changing dynamically, which is also systematic and explainable. For the structurally steady categories such as machinery (HS84), its weight remains relatively steady (coefficient of variation (CV) = 0.12), but for the category with large fluctuation such as vegetables (HS07) or during the epidemic, the CV value increases up to 0.27. It means that we need to calibrate and adjust the model, and the weights did not experience much movement but reacted quickly to macroshocks. This quick responsiveness is a key indicator of a system that is under control and possesses strong recovery capabilities.

The recognized shortcoming is that the weight computation is immediate; thus, there can be large changes if the noise level is high. If there is no improvement in operational stability without sacrificing strategic adaptability, then we will use time-smoothing functions or Bayesian prior distributions of weights for further study. In this way, it creates some inertia and avoids making big changes because of noise over just one period and instead gets used to things changing all the time.

5.3. Theoretical implications for global trade policy and supply chain robustness

The empirical evidence of our model is of great significance, and its adaptability is highly valuable. This is beneficial for understanding the dynamics of international trade and planning a highly resilient system. It integrates the temporal dynamics of deep learning and the feature-specific processing forms of integration methods, bringing a more comprehensive computing system to the multimodal characteristics that trade data inherently possess.

This model has an interpretable output function that brings a brand-new perspective to the core theoretical disputes in international economics, relying on data-based conclusions. SHAP analysis identifies exchange rate fluctuations and tariff policies as the main features and provides quantitative evidence to support the view that direct policy tools have theoretical superiority over macroeconomic aggregates such as GDP in promoting medium- and short-term trade flows. This finding highlights the significant impact of trade policies.

The attention mechanism visualizes cross-border trade corridors, bringing a new perspective to the study of the robustness theory of trade networks. When identifying important paths such as core–edge structures and the China–US corridor, the prediction variance contribution rate is relatively large. This empirically confirms that people are concerned about systemic risks in highly concentrated globalized networks. This indicates that the robustness of the supply chain comes from the internal operations of enterprises. It is also closely related to the stability of a few high-leverage international relations. Therefore, our model provides a quantitative means that can be used to apply pressure detection to the supply chain to see what theoretical risks the interruption of key nodes will bring. In this way, the topic shifts from qualitative vulnerability assessment to predictive risk simulation.

5.4. Strategic value in mitigating diverse geopolitical shocks

Although this model demonstrated good stability during the COVID-19 pandemic, its architecture and adaptive weighting mechanism have the value of responding to more comprehensive geopolitical shocks. The value concepts held by this model extend to situations such as trade wars and economic sanctions, which are different from the demand shocks triggered by the pandemic. They are more targeted and are influenced by policies.

In the long course of tariff disputes, GBDT modules that can effectively identify the nonlinear impact of tariff characteristics will be of great significance. This model can imitate how gradually rising tariff barriers selectively disrupt certain product categories and trade channels, thereby enabling enterprises to preachieve supplier diversification or adjust pricing strategies. Dynamic weights will appropriately enhance the role of the GBDT module in policy-induced crises.

Sudden economic sanctions can break off certain trade routes. The Transformer's attention mechanism will immediately notice if there is a big change in the weight of the approved channels. In addition, it will quickly recalibrate its predictions by giving more weight to other good paths. In this situation, the model's adaptive normalization feature can take care of the disruption. Weighting will also pay attention to parts that can better understand the new and broken up kind of trade.

In all of these cases, the model is built to prevent catastrophic failure due to structural fracture. Therefore, it could both be a prediction tool and a strategic decision support system for reducing geopolitical risks, making it possible to have proactively managed supply chains.

5.5. Limitations and future research directions

Although this study made contributions, its limitations also point out the direction for future research.

By applying NLP technology on unstructured data, future studies will use the NLP module for analyzing official trade agreement texts, central bank reports, text resources from news media, etc. It can make the model quantitatively combine policy sentiment indexes with the meanings of legal documents. To this end, one could fine-tune a language model to identify protective or liberalizing biases in trade policy announcements. The output of this model would then be formulated into a leading indicator, thereby enhancing prediction accuracy prior to the observation of the policy's complete economic effects..

To improve on the quantity of geopolitical risks, according to the results from the last study, we will create a much more detailed and three-dimensional geopolitical risk index. This index should not just focus on these relatively general ones. It will specifically count risks about two-way relationships, chances of punishment, and safety of main logistics routes. In addition, it will inject the corresponding data directly into the feature recognition module of the model.

A time-smoothing mechanism should be added to the dynamic weights. This can enhance its operational stability during high-frequency noise events while also retaining its main adaptability to structural fractures.

In an environment where data are scarce, the issue to be addressed when conducting transfer learning is the reliance on high-quality data. Therefore, we need to explore the transfer learning paradigm, taking advantage of the fact that developed economies have more data. First, we pretrain the model and then fine-tune it for developing regions or emerging trade relations with less data. This is to enhance the inclusiveness and generalization ability of the model.

These improvements are expected to transform the model from a system that only responds to observed economic signals into one that can understand the precursors contained in policy discourse and geopolitical events to predict changes, ultimately forming a more intelligent and resilient global trade ecosystem.

6. Conclusion

This study proposes a hybrid model that was created based on Transformer, XGBoost, and LightGBM, aiming to improve the prediction accuracy of international trade data. After applying a dynamic weight fusion mechanism relying on SHAP values, this model can independently adjust the influence generated by each forming unit. To further optimize the accuracy and stability of the prediction, compared with the traditional LSTM and VAR models, the model in this study shows obvious advantages in short-term, medium-term, and long-term predictions. Its adaptability is more prominent, especially when dealing with extreme events like COVID-19.

In the experiment, we conducted a comprehensive examination on a large dataset and used various evaluation metrics such as WMAPE, RMSE, and R^2 to verify the performance of the model. The experimental results showed that our hybrid model significantly improved the accuracy of predictions, especially demonstrating good stability in the presence of uncertain factors. Compared with traditional models, our model achieved higher prediction accuracy in most prediction periods.

Not only did our model get better at being accurate, but it can also explain itself better too! Using a SHAP value based on a dynamic weight fusion method, this model shows how every feature contributes to making a prediction, which makes it easier for people to understand. It is quite useful for decision-making help when we use it in real life, particularly for complicated international trade prediction projects. It will be good for policymakers and economists if they know more about these links.

This study has achieved many results, but it also has some limitations. In general, this model does well, but in some complicated market fluctuations, there will still be some prediction mistakes. Future studies could add more external economic indicators so that the model becomes more adaptable. We will extend this model into other areas such as financial market forecast and climate change forecast.

This study proposes a hybrid model that brings a new approach to the prediction of international trade data. It has strong predictive ability and good robustness. In the future, we will further improve the model framework, examine more forms of algorithm fusion, expand the application scope of the model, and strive to provide more effective solutions to complex prediction problems in practical operations.

Funding Support

This work was supported by the Sichuan Knowledgeable Intelligent Sciences Priority Research Initiative on Artificial Intelligence (2025): "AI + ML/DL-Driven International Trade Forecasting and Decision Optimization" (SKLS-2025-TP-28).

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at https://github.com/Tianwen-Zhao/AIA25.10.30.

Author Contribution Statement

Wenhao Wang: Conceptualization, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Supervision. **Zhitao Yang:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Tianwen Zhao:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

References

[1] Kırcıcek, T., & Ozparlak, G. (2023). The essential role of international trade on economic growth. *Journal of Economics Finance and Accounting*, *10*(4), 191–202. http://doi.org/10.17261/Pressacademia.2023.1826

[2] Abendin, S., & Duan, P. (2021). International trade and economic growth in Africa: The role of the digital economy. *Cogent Economics & Finance*, *9*(1), 1911767. https://doi.org/10.1080/23322039.2021.1911767

[3] Hassan, T., Song, H., & Kirikkaleli, D. (2022). International trade and consumption-based carbon emissions: Evaluating the role of composite risk for RCEP economies. *Environmental Science and Pollution Research*, *29*(3), 3417–3437. https://doi.org/10.1007/s11356-021-15617-4

[4] Vaghefzadeh, M. H., Karimi, B., & Ahmadi, A. (2023). Modelling customer demand for mobile value-added services: Non-stationary time series models or neural networks time series analysis?. *International Journal of Industrial and Systems Engineering*, *43*(4), 555–581. https://doi.org/10.1504/IJISE.2023.129754

[5] Khalid, O. (2024). Short-term and long-term product demand forecasting with time series models. *Journal of Trends in Financial and Economics*, *1*(3), 11–17. https://doi.org/10.61784/jtfe3022

[6] Ivanov, D. (2024). Exiting the COVID-19 pandemic: After-shock risks and avoidance of disruption tails in supply chains. *Annals of Operations Research*, *335*(3), 1627–1644. https://doi.org/10.1007/s10479-021-04047-7

[7] Yu, Z., Razzaq, A., Rehman, A., Shah, A., Jameel, K., & Mor, R. S. (2022). Disruption in global supply chain and socio-economic shocks: A lesson from COVID-19 for sustainable production and consumption. *Operations Management Research*, *15*(1), 233–248. https://doi.org/10.1007/s12063-021-00179-y

[8] Kaur, J., Parmar, K. S., & Singh, S. (2023). Autoregressive models in environmental forecasting time series: A theoretical and application review. *Environmental Science and Pollution Research*, *30*(8), 19617–19641. https://doi.org/10.1007/s11356-023-25148-9

[9] Al-Gounmeein, R. S., & Ismail, M. T. (2021). Comparing the performances of artificial neural networks models based on autoregressive fractionally integrated moving average models. *IAENG International Journal of Computer Science*, *48*(2), 266–276.

[10] Hou, P. S., Fadzil, L. M., Manickam, S., & Al-Shareeda, M. A. (2023). Vector autoregression model-based forecasting of reference evapotranspiration in Malaysia. *Sustainability*, *15*(4), 3675. https://doi.org/10.3390/su15043675

[11] Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, *15*(8), 255. https://doi.org/10.3390/fi15080255

[12] Osman, E. G. A. (2025). Integrating deep learning and econometrics for stock price prediction: An empirical study of LSTM and

traditional time series models. *Machine Learning with Applications*, *22*, 100730. https://doi.org/10.1016/j.mlwa.2025.100730

[13] Kiani, H., Golshaeian, S., Nazari, M. H., Gharehpetian, G. B., Hosseinian, S. H., & Sarbazi, J. (2024). Comparison of long-term energy demand forecasting in developing and developed countries using machine learning-based algorithms. In *2024 9th International Conference on Technology and Energy Management*, 1–7. https://doi.org/10.1109/ICTEM60690.2024.10631950

[14] Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, *44*(1), 82–99. https://doi.org/10.1177/0047287505276594

[15] Güvenen, O. (2012). *International commodity market models and policy analysis*. Germany: Springer Science & Business Media.

[16] Shibasaki, R., & Watanabe, T. (2010). Future forecast of Chinese trade amount and international cargo flow. *Traffic and Transportation Studies 2010, 638–655*. https://doi.org/10.1061/41123(383)61

[17] Havenga, J. H., & Van Eeden, J. (2011). Forecasting South African containers for international trade: A commodity-based approach. *Journal of Transport and Supply Chain Management*, *5*(1), 170–185.

[18] Lyk-Jensen, S. V. (2011). Forecasting freight flows. *Transport Reviews*, *31*(5), 603–624. https://doi.org/10.1080/01441647.2011.566984

[19] Karimi, A., Faroughi, P., & Rahim, K. A. (2015). Modeling and forecasting of international tourism demand in ASEAN countries. *American Journal of Applied Sciences*, *12*(7), 479–486.

[20] Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2019). Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model. *Journal of Travel Research*, *58*(3), 383–397. https://doi.org/10.1177/0047287518759226

[21] Yang, Y., Liu, W., Zeng, T., Guo, L., Qin, Y., & Wang, X. (2022). An improved stacking model for equipment spare parts demand forecasting based on scenario analysis. *Scientific Programming*, *2022*(1), 5415702. https://doi.org/10.1155/2022/5415702

[22] Yu, Y., Sun, Z., & Li, J. (2025). Cross-border trade export prediction based on reinforcement learning and multimodal data. *International Journal of Information and Communication Technology*, *26*(36), 1–16. https://doi.org/10.1504/IJICT.2025.149053

[23] Isik, E., Ozyilmaz, A., Bayraktar, Y., Toprak, M., Olgun, M. F., & Senturk, N. K. (2025). Assessing the impact of multimodal transportation on economic growth: A machine learning and cointegration approach in 28 countries. *European Journal of Transport and Infrastructure Research*, *25*(1). https://doi.org/10.59490/ejtir.2025.25.1.7715

[24] Yang, W. (2025). Design and application of an international trade flow prediction system based on machine learning algorithms. *Procedia Computer Science*, *261*, 665–673. https://doi.org/10.1016/j.procs.2025.04.258

[25] Al-Karkhi, M. I., & Rządkowski, G. (2025). Innovative machine learning approaches for complexity in economic forecasting and SME growth: A comprehensive review. *Journal of Economy and Technology*, *3*, 109–122. https://doi.org/10.1016/j.ject.2025.01.001

[26] Li, Y., Zhang, S., Chen, G., Yao, Z., Bi, Q., & Li, X. (2025). A prediction scheme for loader excavation resistance based on parallel hybrid deep neural network. *IEEE Sensors Journal*, *25*(14), 27754–27767. https://doi.org/10.1109/JSEN.2025.3578306

[27] Zhang, W., Wu, C., Tang, L., Gu, X., & Wang, L. (2023). Efficient time-variant reliability analysis of Bazimen landslide in the Three Gorges Reservoir Area using XGBoost and LightGBM algorithms. *Gondwana Research*, *123*, 41–53. https://doi.org/10.1016/j.gr.2022.10.004

[28] Li, J., Pan, W., Huang, H., Pan, J., & Wang, F. (2023). STGATE: Spatial-temporal graph attention network with a Transformer encoder for EEG-based emotion recognition. *Frontiers in Human Neuroscience*, *17*, 1169949. https://doi.org/10.3389/fnhum.2023.1169949

[29] Cabello-Solorzano, K., Ortigosa de Araujo, I., Peña, M., Correia, L., & J. Tallón-Ballesteros, A. (2023). The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis. In *International Conference on Soft Computing Models in Industrial and Environmental Applications*, 344–353. https://doi.org/10.1007/978-3-031-42536-3_33

[30] Reddy, V. V. K., Reddy, R. V. K., Munaga, M. S. K., Karnam, B., Maddila, S. K., & Kolli, C. S. (2024). Deep learning-based credit card fraud detection in federated learning. *Expert Systems with Applications*, *255*, 124493. https://doi.org/10.1016/j.eswa.2024.124493

[31] Hu, J., Zhu, K., Cheng, S., Kovalchuk, N. M., Soulsby, A., Simmons, M. J., …, & Arcucci, R. (2024). Explainable AI models for predicting drop coalescence in microfluidics device. *Chemical Engineering Journal*, *481*, 148465. https://doi.org/10.1016/j.cej.2023.148465

[32] Xie, C., Hu, J., Vasdravellis, G., Wang, X., & Cheng, S. (2024). Explainable AI model for predicting equivalent viscous damping in dual frame–wall resilient system. *Journal of Building Engineering*, *96*, 110564. https://doi.org/10.1016/j.jobe.2024.110564

How to Cite: Wang, W., Yang, Z., & Zhao, T. (2025). International Trade Demand Forecasting Model Based on Transformer-XGBoost-LightGBM. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA52027024