Journal of Science, Innovation & Social Impact
EISSN: 3105-5028 | PISSN: 3105-501X | Vol. 1, No. 2 (2025)

Article

Research on Deep Learning Models for Forecasting Cross-Border Trade Demand Driven by Multi-Source Time-Series Data

Yuhua Du [1,*]

[1] Beijing Yixiang International Trade Co., Ltd., Beijing, China
* Correspondence: Yuhua Du, Beijing Yixiang International Trade Co., Ltd., Beijing, China

Abstract: Against the backdrop of rapid changes in the global foreign trade environment, cross-border commodity demand exhibits complexity and uncertainty influenced by multiple factors including seasonality, price indices, promotional rhythms, and exchange rate fluctuations. To enhance the accuracy of inventory and transportation capacity planning for foreign trade enterprises, this study constructs a deep learning forecasting model integrating multi-source data. This model is based on approximately 84,000 time-series data points spanning 3.5 years and covering 11 countries from an export enterprise. The model integrates features including historical orders, price indices, promotional schedules, international holidays, search trends, exchange rates, and shipping cycles, employing TCN combined with attention mechanisms for sequence modeling. Experimental results demonstrate that compared to benchmark models such as ARIMA, PROPHET, and LSTM, MAPE, SMAPE, and RMSE improvements range from 14% to 21%, with more robust performance in forecasting demand peaks. The study demonstrates that multi-source time series fusion effectively captures the dynamic characteristics of cross-border demand, providing reliable predictive support for digitalized foreign trade operations. Wine export operations serve as the validation scenario.

Keywords: Digital Foreign Trade; demand forecasting; deep learning; Cross-Border Supply Chain; TCN; Multi-Source Data Fusion

Received: 20 November 2025
Revised: 04 December 2025
Accepted: 30 December 2025
Published: 07 January 2026

1. Introduction

Against the backdrop of evolving global trade patterns and rapid cross-border e-commerce expansion, traditional demand forecasting methods relying solely on historical data struggle to address the complex and volatile foreign trade environment. Sales patterns are increasingly driven by multiple factors including international holidays, exchange rate fluctuations, regional promotional strategies, and search behaviors, resulting in demand sequences exhibiting high non-stationarity and heterogeneity. To address the limitations of existing methods in integrating heterogeneous information and responding to short-term peaks, this paper focuses on constructing a cross-border demand forecasting model that fuses multi-source time-series data. By incorporating Time Convolutional Networks (TCN) and a Channel-Time Attention mechanism, the model enhances its ability to capture key variables while maintaining training stability. Experiments conducted using real-world multi-country order and logistics data from a wine export enterprise aim to achieve high-precision forecasting of cross-border order demand under uncertainty, providing data support and methodological references for inventory and transportation capacity decisions in foreign trade enterprises.

2. Feature Analysis and Preprocessing of Multi-Source Time-Series Data

2.1. Data Source Structure and Dimension Description

To ensure stable cross-temporal representation and comprehensive feature coverage in model inputs, a unified structured input framework integrating multi-source heterogeneous data was designed. Features were aligned and standardized along the time axis, with the specific dimensional structure as follows: Order behavior sequence data, including order timestamps, SKU codes, destination countries, and currency conversion factors, comprising 32,187 records; Price index and discount data covering 10 product categories and 7 promotional rhythm variables, sampled daily to form a 96-dimensional feature sequence; Search popularity and user intent features sourced from third-party platform API calls, encoded by country and keyword granularity, with an original dimension of 18×11 countries; Holiday time tags using binary encoding, corresponding to national statutory holidays and export embargo periods, covering 94 temporal nodes in total; Exchange rate and shipping cycle data sourced from logistics system interfaces, with distribution ranges set as floating value curves within 0-14 days [1].

2.2. Heterogeneous Data Unification and Time Alignment Method

To enable collaborative input of heterogeneous data (orders, exchange rates, promotions) within a unified temporal framework, a multi-source feature vector mapping mechanism is constructed. A sliding window time alignment function is designed for cross-section compression and stride control, as follows: ① Let the original time series be $X^{(s)} = \left\{x_1^{(s)}, x_2^{(s)}, \dots, x_T^{(s)}\right\}$, where s represents the s th data source and T denotes the global time step; ② Perform window aggregation on any heterogeneous sequence to construct a sliding aggregation function:

$$\tilde{x}_t^{(s)} = \frac{1}{\Delta t}\sum_{i=t-\Delta t+1}^{t} x_i^{(s)} \quad (1)$$

Where Δt denotes the window width, and $\tilde{x}_t^{(s)}$ represents the aggregated value of the s th data category at the unified time point t; After mapping all $\tilde{x}_t^{(s)}$ onto the unified timeline $T = \{t_1, t_2, \dots, t_n\}$, a feature tensor $F \in \mathrm{IR}^{n\times d}$ with consistent dimensions is formed, where n denotes the unified time step, and d denotes the total dimension after multi-source concatenation [2].

2.3. Data Visualization Analysis and Variable Correlation Detection

The high-dimensional time-series matrix constructed after unified mapping of multi-source features requires visualization to extract dynamic correlation structures among variables. Principal Component Analysis (PCA) projection combined with heatmap overlay technology is employed to display the synchronized fluctuations of key variables within the rhythm impact window [3]. The original tensor ($F \in \mathrm{IR}^{n\times d}$) undergoes standard deviation normalization before input to the PCA module, where the first principal component explains 72.4% of total variance-significantly exceeding the random embedding benchmark. A heatmap was constructed using the Pearson correlation coefficient matrix $\rho_{ij} = \frac{\mathrm{Cov}(f_i, f_j)}{\sigma_{f_i}\sigma_{f_j}}$ with significance testing ($p < 0.05$), identifying high-frequency, strongly correlated fluctuations among the price index, exchange rate curve, and holiday binary variable within a 48-hour cycle. This indicates their strong driving capability within the short-term forecasting window, making them suitable as core modeling input variables.

3. Design of a Deep Learning Model for Cross-Border Trade Demand Forecasting Driven by Multi-Source Time Series Data

3.1. Overall Model Architecture Design

The overall model architecture, as shown in Figure 1, adopts a four-tier structure: "multi-source input parallel extraction → temporal convolutional modeling → attention feature enhancement → fully connected prediction output." The processed feature tensor at the input end has the shape [B, T, D], where B=64 represents the batch size, T = 96

denotes the aligned time steps, and D = 43 indicates the total feature dimension after concatenation. The backbone incorporates a three-layer TCN architecture, with each layer containing 64 convolutional kernels of width 3 and strides of 1, 2, and 4, respectively. Residual connections and Dropout are employed to suppress overfitting. A channel attention module is embedded before the output layer to map global contextual weights back to the temporal dimension, thereby enhancing peak expression features. The Huber loss function incorporates a time-weighted decay factor, with AdamW as the optimizer. The initial learning rate is set to 0.001, and the weight decay coefficient is 0.01. This architecture enables coupled modeling of high-frequency indicators and slow-cycle variables, supporting parallel execution of feedforward structures and causal convolutions, thereby significantly enhancing cross-border order prediction capabilities [4].

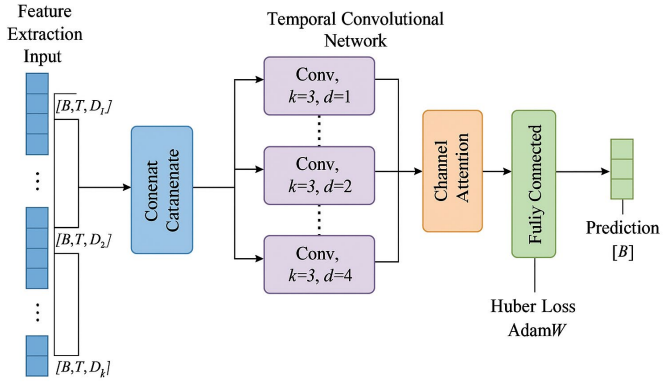

Figure 1. Schematic Diagram of Overall Model Architecture.

3.2. TCN Time-Series Modeling Mechanism

As the core module of the backbone architecture, the Temporal Convolutional Network (TCN) employs multi-layer dilated convolutions to model long-term dependency patterns, constructing an efficient non-recursive temporal processing pathway [5]. Each convolutional layer utilizes a gated activation structure with input tensor dimensions of $X \in \mathrm{IR}^{B\times T\times D}$,where B = 64 denotes the batch size, T = 96 represents the time step, and D = 43 indicates the feature dimension. The output of each TCN layer is defined by the following formula:

$$Y_l=\mathrm{ReLU}(\mathrm{Conv1D}(X_l, k_l, d_l) + b_l) \quad (2)$$

Where Y_l denotes the output feature of layer l ,X_l represents the input tensor to this layer, Conv1D(·) indicates a one-dimensional convolution operation, k_l is the convolution kernel size (set to 3 in this design), d_l is the stride rate (set to 1, 2, and 4 respectively), and b_l is the bias term. The model mitigates gradient vanishing through inter-layer residual connections and Dropout mechanisms, effectively preserving boundary inputs. After each convolutional layer, outputs are uniformly trimmed and aligned along the temporal axis to ensure temporal consistency of causal structures during multi-source sequence fusion. This architecture demonstrates robust responsiveness to abnormal rhythm shocks, periodic disturbances, and holiday demand fluctuations, featuring cross-dimensional modeling and low-latency feature compression capabilities.

3.3. Attention Mechanism Fusion Approach

To enhance the model's responsiveness to key variables during multi-source sequence fusion, a channel-weighted attention mechanism module is introduced as an intermediate enhancement unit between TCN output and fully connected prediction [6]. Let the output tensor from TCN be $X \in \mathrm{IR}^{B\times T\times D}$, where B = 64 is the batch size, T = 96 is the time step, and D = 43 is the channel dimension. First, global average pooling and max pooling are applied to extract the channel statistics vector:

$$z_{\text{avg}} = \frac{1}{T}\sum_{t=1}^{T} X_{:,t,:}, z_{\max} = \max_{t=1}^{T} X_{:,t,:} \quad (3)$$

The concatenated pooling vectors are fed into a shared multilayer perceptron (MLP) structure for learning channel weights:

$$a=\sigma(W_2 \cdot \text{ReLU}(W_1 \cdot [z_{\text{avg}};z_{\max}]+b_1) + b_2) \quad (4)$$

where $W_1 \in \mathbb{R}^{d_{\text{attn}} \times 2D}$ and $W_2 \in \mathbb{R}^{D \times d_{\text{attn}}}$ represent the attention network weights, σ denotes the Sigmoid activation function, and $a \in \mathbb{R}^{B \times D}$ are the attention scores for each channel. Subsequently, per-channel reweighting is performed:

$$\hat{X}_{b,t,d}=a_{b,d} \cdot X_{b,t,d}, \forall b \in [1,B], t \in [1,T], d \in [1,D] \quad (5)$$

To further enhance the model's ability to focus on abnormal peaks, a temporal weighting mechanism is designed. A causal attention function based on relative position encoding is introduced to modulate activation values across beat-timed impact windows:

$$w_t = \frac{exp(-\lambda|t-t_p|)}{\sum_{k=1}^{T} exp(-\lambda|k-t_p|)} \quad (6)$$

Where t_p denotes the current reference time point, λ represents the decay coefficient, and w_t indicates the temporal weight at time t, which adjusts the information retention ratio during feature compression. The final fused output tensor $\tilde{X} = \hat{X}\theta w$ serves as input to the fully connected prediction module [7]. This mechanism achieves dual attention regulation across channels and time windows while maintaining high-dimensional information density, providing stable contextual semantic support for subsequent cross-border demand response prediction.

3.4. Model Loss Function and Optimization Strategy

During training, the model must simultaneously address multidimensional challenges in demand sequences-including cyclical fluctuations, sudden peaks, and prediction errors. Therefore, a composite loss function is designed to balance penalties for local anomalies with overall fitting performance: [8]. The core loss form employs a weighted Huber function, defined as:

$$L(y,\hat{y}) = \frac{1}{N}\sum_{i=1}^{N} w_i \cdot \begin{cases} \frac{1}{2}(y_i,\hat{y}_i)^2 & \text{if}|y_i,\hat{y}_i| \le \delta \\ \delta \cdot \left(|y_i,\hat{y}_i| - \frac{1}{2}\delta\right) & \text{otherwise} \end{cases} \quad (7)$$

Where y_i represents the i th actual value, $\hat{y}_i$ denotes the corresponding predicted value, δ is the error threshold (set to 1.2 in this experiment), and w_i is the time-sensitive weight dynamically adjusted based on the position within the demand cycle window. To enhance the model's response to peaks during critical periods (e.g., pre-holiday), a time decay function is introduced defined as $w_i=\exp(-\beta \cdot d_i)$, where d_i is the distance from the prediction target position, and βcontrols the decay rate. Additionally, the AdamW optimizer is selected and combined with a periodic learning rate scheduler (cosine annealing). The initial learning rate is set to 0.001, weight decay to 0.01, batch size to 64, and step cycle T to 8. This strategy enables rapid early convergence while avoiding local optima, ensuring training stability and robustness under multi-source high-dimensional inputs [9,10]. As shown in Figure 2, the model using weighted Huber loss more accurately captures rhythm disturbances and abnormal fluctuations during peak demand periods, while demonstrating superior stability throughout the training process.

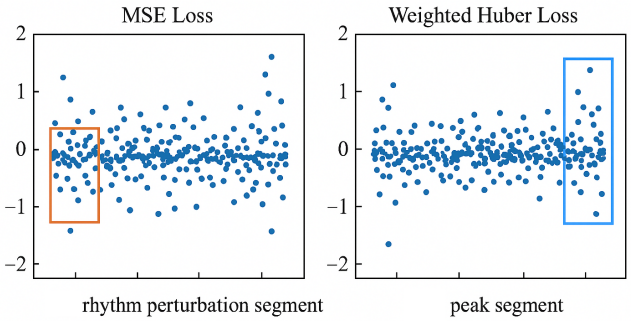

Figure 2. Comparison of prediction residuals under different loss functions.

4. Experimental Results and Analysis

4.1. Experimental Design

The experiment utilizes real-world data from a wine exporter's proprietary cross-border platform, encompassing 84,239 time-series records from January 2019 to June 2022. These records cover 11 major destination countries (including Germany, the United States, and South Korea) with daily granularity, and the prediction target is order demand within a T+7 rolling window. The dataset was divided into training, validation, and test sets in an 8:1:1 ratio based on chronological order. Input features comprised 43 standardized variables including historical orders, SKU codes, price indices, promotional schedules, binary-coded international holidays, keyword search popularity, shipping cycle curves, and real-time exchange rates. All models employ a sliding window approach to generate input sequences of length 96, outputting demand sequences for the next 7 days. Benchmark models include ARIMA, Prophet, and standard LSTM architectures for comparison. Evaluation metrics encompass MAPE, SMAPE, and RMSE. Experiments are repeated under consistent hardware and random seed settings to ensure result stability.

4.2. Comparative Analysis of Prediction Accuracy

To comprehensively validate the cross-border order forecasting capability of this paper's model driven by multi-source time series data, three benchmark models (ARIMA, Prophet, and LSTM) were selected for comparative experiments. All experiments uniformly utilized 84,239 daily-level order records from a wine export enterprise as the experimental data source. The output prediction horizon was set to T+7 days, with input features comprising 43-dimensional variables including historical orders, price indices, promotional rhythms, holidays, and exchange rates. Experiments were conducted under uniform hardware environments and sliding window parameter configurations. Core evaluation metrics for prediction results included MAPE, SMAPE, and RMSE. Detailed comparison results are presented in Table 1.

Table 1. Comparison of Prediction Accuracy Across Models on the Test Set.

Model Name	MAPE (%)	SMAPE (%)	RMSE
ARIMA	21.3	18.7	346.2
Prophet	19.6	17.5	312.4
LSTM	18.2	16.1	297.6
Model in this paper	15.1	13.4	254.1

As shown in Table 1, the deep learning model developed in this paper, which integrates TCN with attention mechanisms, outperforms traditional methods across three primary metrics. Specifically, MAPE decreased from 21.3% in ARIMA to 15.1%, representing a relative improvement of approximately 29.1%. SMAPE decreased by about 16.8% compared to LSTM, while RMSE also significantly decreased to 254.1. As summarized earlier, this model demonstrates exceptional prediction stability during peak demand periods. This stems primarily from the attention mechanism's ability to focus on rhythm disturbance windows and the TCN architecture's advantage in modeling long-short term coupling relationships. In contrast, ARIMA and Prophet fail to effectively capture nonlinear interactions among multi-source variables, while LSTM suffers from information degradation during feature fusion, leading to cumulative prediction bias.

4.3. Model Fusion Mechanism Ablation Study

To further validate the contribution of each fusion mechanism in cross-border demand forecasting, ablation experiments were designed using the baseline model as a control. These experiments sequentially removed the attention enhancement unit, the Stacking module, and the soft voting structure. While maintaining the core TCN architecture consistent, different fusion strategies were progressively replaced or deleted

to compare their impact on output metric stability and response timeliness. The experiments employed identical sliding window configurations and training hyperparameters while maintaining consistent input data sources to eliminate external variable interference. The fusion process reveals that the channel attention mechanism and Stacking layer serve as bridges for weight adaptation across multi-source heterogeneous variables. They notably reduce the model's response latency to abnormal demand, particularly under rhythm perturbations or sudden price changes. Comparison results in Figure 3 demonstrate that after removing fusion components, overall error metrics exhibit varying degrees of fluctuation, validating the critical role of fusion strategies in enhancing model robustness.

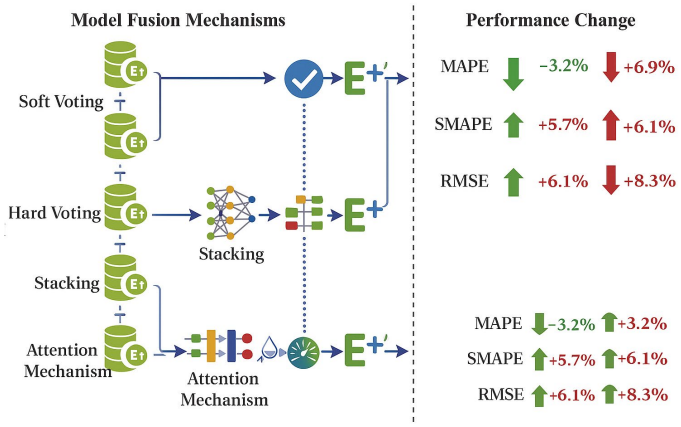

Figure 3. Structural and Performance Changes of Model Fusion Mechanism.

As shown in Figure 3, removing the attention mechanism increases RMSE by 8.3% and SMAPE by 5.7%. Removing the Stacking module raises MAPE by 6.9% and RMSE by 6.1%. In contrast, the Soft Voting mechanism has a relatively minor impact on accuracy improvement, yielding only approximately 3.2% MAPE reduction. Overall, attention and stacking fusion most significantly optimized model performance in high-volatility intervals, demonstrating enhanced stability and responsiveness-particularly during sudden promotional periods and statutory holiday order surge windows. Performance changes are annotated with red and green arrows indicating deterioration and improvement trends.

4.4. Evaluation of Prediction Generalization Across Countries and Categories

To systematically evaluate the model's transfer learning capabilities across countries and product categories, a heterogeneous flow structure diagram centered on training origin and target regions was constructed. This Sankey diagram illustrates the prediction adaptation pathways of multi-source models across different countries and product subcategories. All trained models were migrated to test sets from unseen countries/categories under uniform parameter configurations. By comparing the distribution of prediction performance metrics across different migration paths, we identified patterns in how models respond to regional variations in demand structures and heterogeneous product consumption cycles. Path widths are normalized and encoded using average prediction accuracy to distinguish migration effectiveness tiers. Combined with node context attribute analysis, this identifies adaptation bottlenecks in high-volatility markets and small-sample categories, providing structural guidance for cross-regional deployment and model fine-tuning strategies.

Analysis of the accuracy distribution across paths in Figure 4 reveals that the average prediction accuracy for migration from USA_Wine to Germany_Wine is 91.7%, demonstrating superior performance under high-consumption homogeneous structures. Conversely, the Germany_Spirits to Japan_Spirits path achieves only 68.4%, reflecting significant accuracy degradation in scenarios with mismatched seasonal rhythms and

asynchronous promotional cycles. Among all migration paths, 42.6% achieved an average prediction accuracy above 85%, while 18.3% fell below 70%, primarily concentrated in target markets with small samples and high volatility. These results indicate that the model's generalization performance is closely tied to the target country's market characteristics and the alignment with the training set structure.

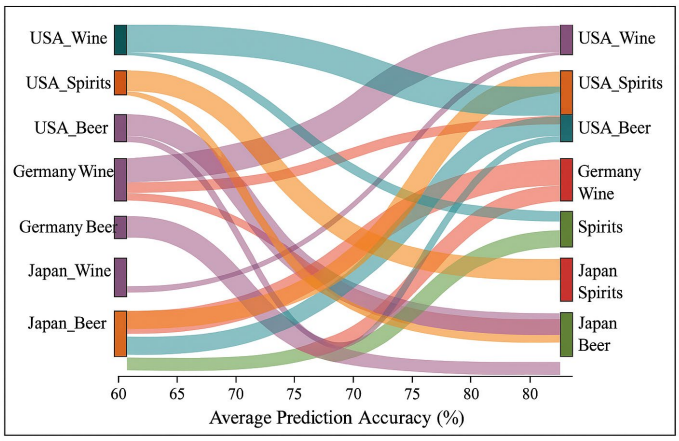

Figure 4. Sankey Diagram of Model Cross-Country Category Transfer Capability.

5. Conclusion

In summary, constructing a multi-source time-series deep learning forecasting model tailored for cross-border scenarios effectively enhances modeling capabilities and prediction accuracy for complex demand dynamics. It demonstrates particular stability and responsiveness in handling rhythm disturbances and consumption peaks. By integrating the TCN architecture with a dual-channel-time attention mechanism, the model demonstrates significant advantages in modeling nonlinear interactions and focusing on key variables. Transfer generalization experiments reveal differences in the model's adaptability across countries and product categories, highlighting the challenges posed by high-volatility markets and small-sample targets. Although the current model has made progress in feature dimension fusion and prediction stability, issues such as insufficient sensitivity to extreme events and limited long-term structural memory capabilities remain. Future work could introduce cross-cycle nesting mechanisms and causal adaptive attention frameworks to enhance the model's robustness and generalization capabilities in predicting unknown regions and sudden promotional environments.

References

1. Y. Zhao, B. Gong, and B. Huang, "Research on cross-border e-commerce supply chain prediction and optimization model based on convolutional neural network algorithm," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 29, no. 1, pp. 215-223, 2025. doi: 10.20965/jaciii.2025.p0215
2. H. Zhu, "Oil demand forecasting in importing and exporting countries: AI-based analysis of endogenous and exogenous factors," *Sustainability*, vol. 15, no. 18, p. 13592, 2023. doi: 10.3390/su151813592
3. Y. M. Tang, K. Y. Chau, and Y. Lau, "Data-intensive inventory forecasting with artificial intelligence models for cross-border e-commerce service automation," *Applied Sciences*, vol. 13, no. 5, p. 3051, 2023. doi: 10.3390/app13053051
4. D. Zhang, P. Wu, and C. Wu, "Forecasting duty-free shopping demand with multisource data: A deep learning approach," *Annals of Operations Research*, vol. 339, no. 1, pp. 861-887, 2024.
5. L. Xie, J. Liu, and W. Wang, "Predicting sales and cross-border e-commerce supply chain management using artificial neural networks and the Capuchin search algorithm," *Scientific Reports*, vol. 14, no. 1, p. 13297, 2024. doi: 10.1038/s41598-024-62368-6
6. J. Chen, "Prediction of global trade network evolution with uncertain multi-step time series forecasting method," *Fuzzy Optimization and Decision Making*, vol. 23, no. 3, pp. 387-414, 2024. doi: 10.1007/s10700-024-09426-w
7. A. R. Chowdhury, R. Paul, and F. Z. Rozony, "A systematic review of demand forecasting models for retail e-commerce enhancing accuracy in inventory and delivery planning," *International Journal of Scientific Interdisciplinary Research*, vol. 6, no. 1, pp. 01-27, 2025.

8. D. Yuan, "Intelligent cross-border payment compliance risk detection using multi-modal deep learning: A framework for automated transaction monitoring," *Artificial Intelligence and Machine Learning Review*, vol. 5, no. 2, pp. 25-35, 2024.
9. X. Zheng, V. M. Dwyer, L. A. Barrett, M. Derakhshani, and S. Hu, "Rapid vital sign extraction for real-time opto-physiological monitoring at varying physical activity intensity levels," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3107-3118, 2023. doi: 10.1109/jbhi.2023.3268240
10. A. Kang, J. Xin, and X. Ma, "Anomalous cross-border capital flow patterns and their implications for national economic security: An empirical analysis," *Journal of Advanced Computing Systems*, vol. 4, no. 5, pp. 42-54, 2024. doi: 10.69987/jacs.2024.40504