Leveraging Big Data For Real-Time Risk Assessment And Management In Cross-Border Business Administration

Pu Wang and Yao Gu*

College of Finance and Economics, Wanjiang University of Technology, Ma'anshan 243000, Anhui, China

*Corresponding author. E-mail: yaogu72ju@outlook.com

Received: Mar. 31, 2026; Accepted: June. 02, 2025

In globalized and digitally driven world, cross-border businesses face complex financial risks arising from differing regulations, geopolitical tensions, and unstable markets. Traditional risk management models struggle to process large-scale data in real time, limiting effective decision-making. This paper proposes an integrated big data framework for real-time risk analysis in cross-border operations. The framework incorporates Conditional Generative Adversarial Networks (cGANs) to address data imbalance, Long Short-Term Memory (LSTM) networks to forecast evolving risk patterns, and Extreme Value Theory (EVT) to quantify extreme losses through Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). Reinforcement Learning (RL) is also included to optimize decision strategies such as hedging and capital allocation. With real-time monitoring and adaptive learning, the system supports proactive and scalable risk management. Experiments using the DataCo Smart Supply Chain Dataset show excellent performance, achieving 99.83% accuracy, 99.77% precision, 99.18% recall, 99.88% F1-score, and strong AUC results, validated further through VaR, CVaR, and Kupiec backtests.

Keywords: Conditional Generative Adversarial Networks, Extreme Value Theory, Graph Neural Networks, K-Nearest Neighbour and Long Short-Term Memory

© The Author('s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202611_34.018

1. Introduction

The rapid boom in international trade, financial linkages, and digital payments has resulted in [1]. Despite technological advancements, integrating heterogeneous data sources into a unified risk analysis framework remains a significant challenge [2]. In the field of international business, danger lurks due to various causes [3, 4]. In an environment with an entangled market, danger tends to spread by the process of contagion, i.e., if the market is adversely affected in some sectors, the problem may spread immediately to other sectors as well [5].

This study addresses the limitations of existing research by proposing an integrated big data-driven framework for real-time risk. Unlike prior studies that primarily. In contrast to conventional hybrid models that combine LSTM and reinforcement learning, the proposed framework uniquely incorporates cGAN for data balancing and EVT for extreme risk modeling, enabling improved detection of rare high-impact events and more robust decision-making. The paper is structured into literature review, methodology, dataset preparation, and experimental analysis sections. It further discusses model evaluation results and concludes with key findings and future research directions.

2. Materials and methods

The rise in complexity in international trade has led researchers to investigate technology-based solutions. Adesanya et al. [6] studied the use of advanced digital technology in supporting US business. Another study by Adesuyi et al. [7] proposed an artificial intelligence-based risk scoring model for detecting high-risk cross-border trade payments. Adesanya et al. [6] and Li et al. [8] for cross-border data mobility risk. Wedraogo et al. [9] studied business strategies for international business expansions. Ylönen and Aven [10] discussed the use of intelligence system technology in risk management. Furthermore, Li et al. [8] studied blockchain technology in digital ecosystems in cross-border business. The proposed model systematically integrates multiple advanced big data techniques to enhance analytical performance, such as cGAN.

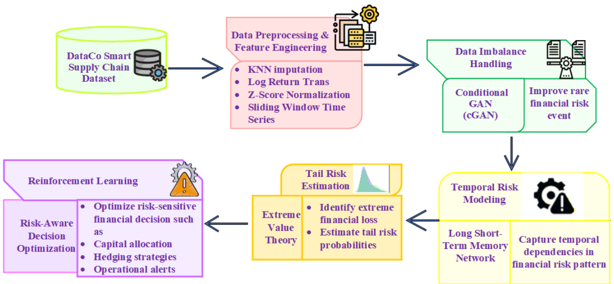

Fig. 1. Proposed Framework for Real-Time Cross-Border Financial Risk Assessment.

Fig. 1 is a diagram representing the entire workflow of the proposed model. The proposed framework follows a sequential big data analytics pipeline for real-time cross-border financial risk assessment. The Graph Neural Network (GNN) module models cross-border financial institutions, supply-chain entities, and transactional relationships as interconnected graph structures consisting of nodes and edges as given in Eq. (1).

$$x_{syn} = G\left(z \mid y; \theta_g\right) \tag{1}$$

Where, z is the random noise vector, y is the risk condition label, θ_g as given in Eq. (2).

$$D\left(x \mid y; \theta_d\right) = P(\text{ real } \mid x, y) \tag{2}$$

Where, x is the financial transaction data and θ_d is the discriminator parameters as given in Eq. (3).

$$P(x \mid y) = \frac{P(y \mid x) P(x)}{P(y)} \tag{3}$$

Synthetic sample generation is as given in Eq. (4).

$$x_{\text{syn}} = G(z \mid y) \tag{4}$$

The distribution of the dataset is balanced by creating synthetic minority samples.

In Fig. 2 the cGAN architecture was trained using alternating generator-discriminator optimization with the

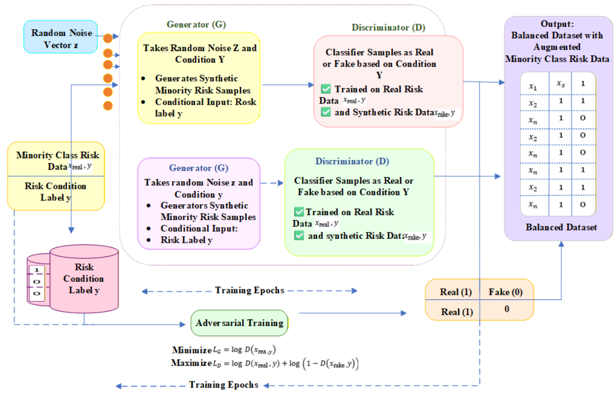

Fig. 2. cGAN based Data Imbalancing Handling for Cross-Border Financial Risk Detection.

Adam optimizer and a controlled learning rate strategy. Batch normalization and dropout regularization were incorporated to stabilize gradient propagation and reduce overfitting during training.

Lstm-based temporal dependency learning for financial risk prediction

Financial time-series data often exhibit volatility clustering, nonlinear temporal dependencies, and abrupt regime. The LSTM architecture effectively learns both short-term and long-term sequential risk given in Eq. (5).

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{5}$$

where, i_t is the input gate activation value, σ of real-time financial risk prediction.

In Fig. 3, The LSTM architecture is capable of adapting to sudden regime shifts in crossborder financial environments by continuously. This adaptive temporal learning capability improves the robustness and responsiveness of real-time financial risk prediction in highly dynamic international business environments.

The EVT framework applies the Peak Over Threshold (POT) approach, where the threshold value is selected using high quantile analysis to isolate extreme financial loss events. The Generalized Pareto Distribution parameters, including the shape parameter (ξ). Sensitivity analysis of these parameters was also considered to evaluate the stability of VaR and CVaR estimation under different threshold configurations, thereby improving the reliability of extreme financial risk quantification. The threshold demonstrated consistent tail-index behaviour and minimized estimation variance, thereby satisfying EVT assumptions for reliable tail-risk modelling. Furthermore, sensitivity analysis was conducted using alternative threshold levels including the

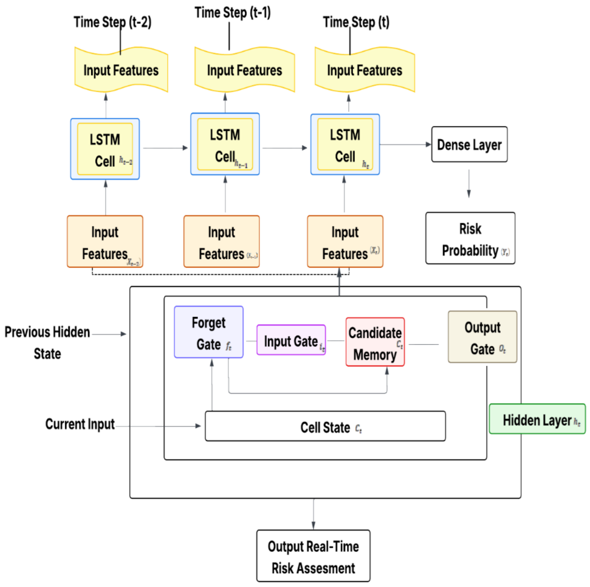

Fig. 3. Architecture of Temporal Risk Modelling using LSTM Network.

90th, 95th, and 97.5th percentiles to evaluate the robustness of VaR and CVaR estimation under varying tail conditions. Tail distribution behaviour is as given in Eq. (6).

$$P(X > u + x \mid X > u) = \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} \tag{6}$$

Where, X is the financial loss variable, u is the threshold value for extreme events, x is the excess loss over shows the extreme loss distribution's probability density as given in Eq. (7).

$$f(x) = \frac{1}{\beta}\left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi - 1} \tag{7}$$

Where, $CVaR$ is the expected tail loss and VaR_α is the value-at-Risk at confidence level α. calculates the financial loss distribution's tail heaviness as given in Eq. (8).

$$\xi = \frac{1}{k}\sum_{i=1}^{k} \ln\left(\frac{X_i}{u}\right) \tag{8}$$

Where, $k \rightarrow$ Number of extreme observations, $X_i \rightarrow$ Extreme loss observations and ξ indicates heavier tail risk. Extreme Value Theory (EVT) has strong applicability in real-world financial risk management, particularly in estimating extreme losses in cross-border business environments. In real-time cross-border financial risk evaluation, VaR and CVaR offer complementary insights. VaR estimates the maximum expected loss within a 0confidence level and is useful for routine risk monitoring, but it may underestimate extreme losses beyond this threshold. The reinforcement learning framework utilizes policy-gradient-based optimization to continuously learn adaptive risk-aware decision policies under dynamic cross-border financial environments. An ε-greedy exploration strategy is incorporated to balance exploration and exploitation during policy learning, enabling the agent to discover optimal actions while maintaining stable long-term reward maximization as given in Eq. (9).

$$\max_{\pi} \mathbb{E}[U(\text{ Profit } - \text{ Risk })] \tag{9}$$

Where π is the policy (decision strategy), $\mathbb{E}$ is the expectation operator, $U(\cdot)$. DataCo Smart Supply Chain Dataset, which provides extensive information on international supply chain operations, was used in this research [11]. DataCo Global, a business that oversees global supply chain operations, provided the dataset. In this study, "risk" is defined as the likelihood of adverse events in crossborder business operations, particularly focusing on delayed deliveries and high-risk transactions within the supply chain. A binary risk label is constructed based on delivery status, where delayed or disrupted transactions are categorized as high-risk instances, while on-time and stable transactions are considered low-risk. The preprocessing statistical consistency analysis to improve feature representation quality and predictive performance. The K value in KNN imputation was chosen to balance local neighborhood similarity and noise reduction, while the sliding window size was optimized to effectively capture temporal financial dependencies without introducing excessive computational complexity. The distance between two samples x_i and x_j is computed as given in Eq. (10).

$$d\left(x_i, x_j\right) = \sqrt{\sum_{k=1}^{m}\left(x_{ik} - x_{jk}\right)^2} \tag{10}$$

The missing value is estimated as the average of its K nearest neighbours as given in Eq. (11).

$$\hat{x}_i = \frac{1}{K}\sum_{j \in \mathcal{N}_K(i)} x_j \tag{11}$$

Where, x_i is the feature route of the sample i, m is the amount of features, $\mathcal{N}_K(i)$ is the set of K nearest neighbours and $\hat{x}_i$ is the imputed value as given in Eq. (12).

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{12}$$

Where, P_t is the asset price at time t and r_t is the logarithmic return as given in Eq. (13).

$$z_i = \frac{x_i - \mu}{\sigma} \tag{13}$$

Where μ is the cruel of the feature, σ is the standard deviation and z_i is the normalized value as given in Eq. (14).

$$X_t = [r_{t-w+1}, r_{t-w+2}, \ldots, r_t] \tag{14}$$

Where, w is the window size and X_t is the effort sequence at time t. The effectiveness of real-time financial risk prediction in dynamic cross-border environments. Smaller window sizes improve responsiveness by enabling rapid adaptation to sudden market fluctuations, geopolitical disruptions, and transactional anomalies through the prioritization of recent temporal patterns as given in Eq. (15).

$$\Sigma = \frac{1}{n} X^T X \tag{15}$$

Where, Σ is the covariance matrix, n is the eigenvector, X^T is the eigenvalue and X is the matrix of top k eigenvectors as given in Eq. (16).

$$F \in \mathbb{R}^{n \times k} \tag{16}$$

Where, n is the amount of remarks, k is the amount of extracted principal components and F is the final feature matrix used for risk modelling. The percentage of accurate predictions with respect to all predictions is called accuracy. It is a broad indicator of the overall performance of the device.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

where TP is denoted as True Positives, TN is denoted as True Negatives, and FP.

The ratio of correctly predicted positive cases to all predicted positive cases

$$\text{Precision} = \frac{TP}{TP + FP} \tag{18}$$

The model's efficacy in identifying high-risk events is demonstrated by recall, which

$$\text{Recall} = \frac{TP}{TP + FN} \tag{19}$$

F1-score is the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

The F1-score is useful in cases where the data is imbalanced. when it is equal to 0.5, it indicates random guessing.

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})d(\text{FPR}) \tag{21}$$

Where TPR is denoted as True Positive and FPR is denoted as False Positive. The likelihood ratio is computed as:

$$LR_{uc} = -2 \ln \left[\frac{(1-\hat{p})^{n-x} \hat{p}^x}{(1-p)^{n-x} p^x} \right] \tag{22}$$

Where, n is denoted as the total number of observations

The proposed framework can be deployed within enterprise-level cross-border business infrastructures using cloud-based distributed computing and API-driven integration mechanisms. The architecture supports interoperability with ERP systems, financial transaction platforms, supply chain management systems, and regulatory compliance databases for real-time risk monitoring and analysis.

3. Results and discussion

The system configuration includes an Intel Core i7 12th Generation processor, 16 GB DDR4 RAM operating at 3200 MHz , and a 512 GB NVMe SSD for storage. The software environment runs on Windows 11 (64-bit) with Python 3.11 configured through Anaconda 2024. The software environment consists of Python 3.11.x and Jupyter Notebook 7.x as the core development platform. This performance improvement is largely attributed to the effective integration of big data, which facilitates real-time analytics and enhances the model's ability to respond to complex and dynamic cross-border risk scenarios.

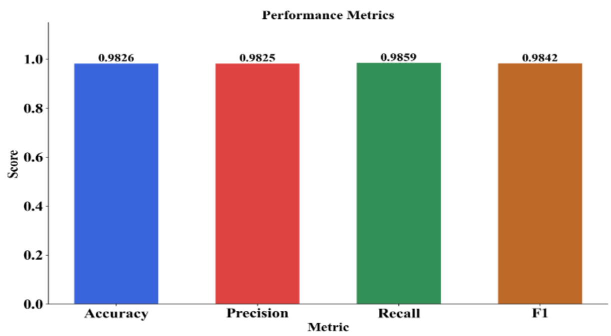

Fig. 4. Comparative performance of different risk assessment models.

From the data presented in Fig. 4, we can see that the hybrid approach of LSTM + EVT outperforms all the baselines.

Fig. 5 depicts false positives and false negatives. Considering the imbalanced nature of cross-border financial risk datasets, evaluation metrics Precision, recall, F1-score, AUC, and confusion matrix interpretation were utilized to assess minority risk-event detection.

As demonstrated in Fig. 6, our proposed LSTM-based solution excels relative to the other baseline models.

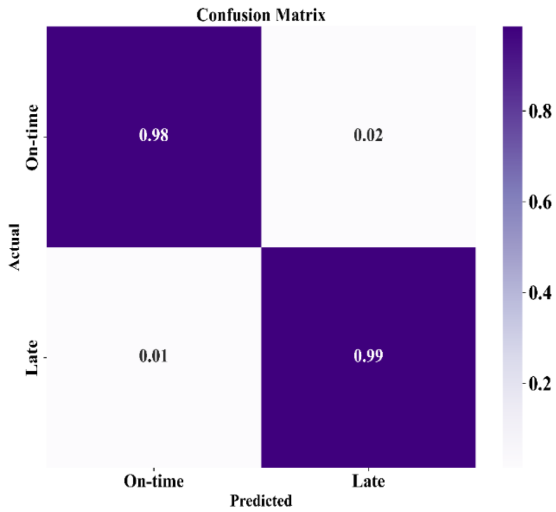

Fig. 5. Confusion matrix of the delivery model for On-time and Late events.

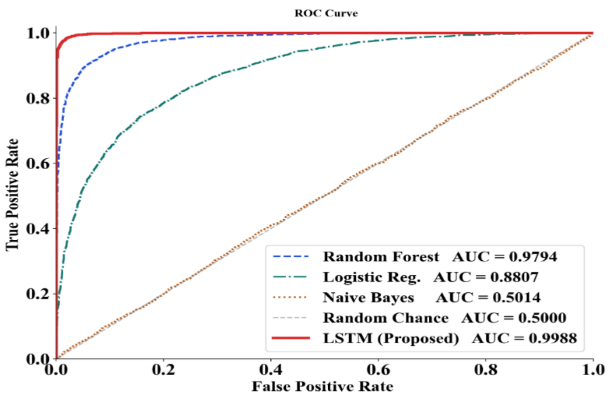

Fig. 6. ROC Curve Comparison of Proposed LSTM Model and Baseline Classifiers.

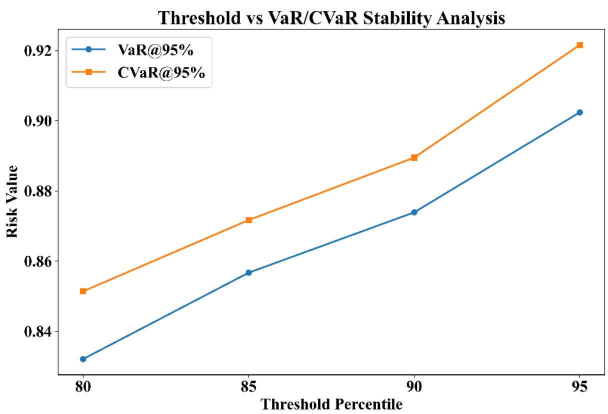

Fig. 7. Threshold Sensitivity Analysis for EVT-Based VaR and CVaR Estimation.

Fig. 7 illustrates the sensitivity analysis of EVT-based VaR and CVaR estimation under varying threshold percentiles. The results demonstrate gradual and stable changes in risk estimates as the threshold increases from the 80th to 95th percentile. The absence of abrupt fluctuations confirms the robustness of the selected EVT threshold configuration and validates the reliability of tail-risk estimation for cross-border financial risk assessment.

Table 1 ablation study shows stepwise performance improvement from baseline to full model. with the final model achieving highest accuracy (0.9826), precision (0.9825), recall (0.9859), and F1-score (0.9842), confirming overall effectiveness. Table 2 presents the comparative performance evaluation of the proposed framework against conventional machine learning. Table 3 presents the performance validation results of the proposed framework using training and future unseen test datasets. Table 4 illustrates the robustness analysis of the proposed framework under varying noise levels introduced into the financial dataset. Table 5 presents the computational efficiency analysis of the proposed framework compared with conventional machine learning and deep learning baseline models.

4. Conclusion

The proposed framework leverages big data analytics and advanced machine learning models such as cGAN, LSTM, and EVT-based risk estimation to deliver an efficient and accurate system for real-time risk assessment in cross-border business environments. It effectively addresses challenges including data imbalance, temporal patterns, and interconnected global networks. Experiments conducted using real-world international supply chain and carbon market datasets demonstrate that the framework outperforms existing techniques in accuracy, recall, F1-score, and risk estimation through VaR and CVaR. These results highlight its robustness and suitability for highly volatile global markets. Future enhancements may integrate additional data sources such as social media sentiment, news streams, and IoT-enabled supply chain information to strengthen predictive capabilities. The system can also be extended with explainable AI mechanisms and cloud-based deployment to improve transparency, scalability, and operational efficiency, making it highly adaptable for multinational corporations managing complex cross-border risks.

Declarations

Data Availability

The data that support the findings of this study are available from relevant financial databases and institutional

Table 1. Ablation Study for Component Contribution Analysis in the Proposed.

Model Variant	Accuracy	Precision	Recall	F1-score
Baseline	0.7420	0.7315	0.7284	0.7299
+ Feature Engineering	0.8643	0.8567	0.8482	0.8524
+ cGAN	0.9518	0.9463	0.9397	0.9430
+ LSTM	0.9654	0.9608	0.9552	0.9580
+ EVT	0.9723	0.9689	0.9701	0.9695
+ RL	0.9781	0.9756	0.9778	0.9767
Proposed Model (All)	0.9826	0.9825	0.9859	0.9842

Table 2. Comparative Performance Evaluation of the Proposed Framework and Baseline Models.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.89	0.87	0.85	0.86	0.88
Random Forest	0.94	0.93	0.92	0.92	0.95
XGBoost	0.96	0.95	0.94	0.94	0.97
LSTM (Baseline)	0.95	0.94	0.93	0.93	0.96
Proposed Model	0.9826	0.9825	0.9859	0.9842	0.986

Table 3. Dataset Split-Based Performance Validation of the Proposed Framework.

Dataset Split	Accuracy	Precision	Recall	F1-Score	AUC
Training Data	0.985	0.984	0.986	0.985	0.988
Future Test Data	0.972	0.970	0.974	0.972	0.981

Table 4. Noise Robustness Analysis of the Proposed Financial Risk Prediction Framework.

Noise Level	Accuracy	F1-Score
0%	0.9826	0.9842
5%	0.9751	0.9770
10%	0.9684	0.9702
15%	0.9563	0.9581

Table 5. Computational Efficiency Analysis of Baseline Models and Proposed Framework.

Model	Training Time (s)	Prediction Time (ms/sample)
Random Forest	45	3.2
XGBoost	52	2.8
LSTM	120	5.5
Proposed Model	150	6.1

sources. Additional data may be obtained from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding Statement

This research received no external funding.

Author Contributions

Pu Wang: Conceptualization, methodology, data analysis, and writing—original draft preparation. Yao Gu: Supervision, validation, review and editing, and final approval of the manuscript.

Ethical Approval

This study does not involve human participants or animals and therefore does not require ethical approval.

Consent to Participate

Not applicable.

Consent to Publication

The authors consent to the publication of this manuscript.

Competing Interests

The authors declare that there are no competing interests.

References

[1] S. Mirzaye and M. Mohiuddin, (2025) *"Digital Transformation in International Trade: Opportunities, Challenges, and Policy Implications"* **Journal of Risk and Financial Management** *18*(8): 421. DOI: 10.3390/jrfm18080421.

[2] X. Zhang and C. Guo, (2024) *"Research on Multimodal Prediction of E-Commerce Customer Satisfaction Driven by Big Data"* **Applied Sciences *14***(18): 8181. DOI: 10.3390/app14188181.

[3] F. Wang and Y. Zhou, (2025) *"Cross-Border E-Commerce, Platform Economy, and Export Product Quality"* **International Business Review *34***(5): 102466. DOI: 10.1016/j.ibusrev.2025.102466.

[4] K. Wang, X. Guo, and D. Yang, (2022) *"Research on the Effectiveness of Cyber Security Awareness in ICS Risk Assessment Frameworks"* **Electronics *11***(10): 1659. DOI: 10.3390/electronics11101659.

[5] G. C. Landi, F. Iandolo, A. Renzi, and A. Rey, (2022) *"Embedding Sustainability in Risk Management: The Impact of Environmental, Social, and Governance Ratings on Corporate Financial Risk"* **Corporate Social Responsibility and Environmental Management *29***(4): 1096–1107. DOI: 10.1002/csr.2256.

[6] O. R. Adesanya, F. Anwansedo, O. T. Akinwande, S. Asiimire, C. O. Esechie, and M. Sekinobe, (2024) *"Advancing U.S. Business Interests Abroad: Leveraging Technology to Optimize Cross-Border Operations"* **International Journal of Latest Technology in Engineering Management & Applied Science *13***(7): 24–33. DOI: 10.51583/IJLTEMAS.2024.130704.

[7] M. O. Adesuyi, O. Akomolafe, B. O. Olaogun, V. U. Ndukwe, and J. K. Sakyi, (2024) *"AI-Driven Risk Scoring Model for Global Cross-Border Trade Payment Transactions"* **International Journal of Advanced Multidisciplinary Research and Studies *4***(1): 1569–1581. DOI: 10.62225/2583049X.2024.4.1.5281.

[8] J. Li, W. Dong, C. Zhang, and Z. Zhuo, (2022) *"Development of a Risk Index for Cross-Border Data Movement"* **Data Science and Management *5***(3): 97–104. DOI: 10.1016/j.dsm.2022.05.003.

[9] L. Wedraogo, S. Essandoh, J. K. Sakyi, et al., (2023) *"Analyzing Risk Management Practices in International Business Expansion"* **Journal of Frontiers in Multidisciplinary Research *4***(2): 300–313. DOI: 10.54660/JFMR.2023.4.2.300-313.

[10] M. Ylönen and T. Aven, (2023) *"A New Perspective for the Integration of Intelligence and Risk Management in a Customs and Border Control Context"* **Journal of Risk Research *26***(4): 433–449. DOI: 10.1080/13669877.2023.2176912.

[11] Kaggle Dataset. *DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS*. URL: https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis.