International Journal of Social Science and Education Research

ISSN Print: 2664-9845
ISSN Online: 2664-9853
Impact Factor: RJIF 8.42
IJSSER 2025; 7(2): 912-918
www.socialsciencejournals.net
Received: 22-09-2025
Accepted: 26-10-2025

Jikun Zheng
Business School, Lingnan Normal University, Zhanjiang 524048, Guangdong, China

Corresponding Author:
Jikun Zheng
Business School, Lingnan Normal University, Zhanjiang 524048, Guangdong, China

Research on Influencing Factors of Import and Export Trade

Jikun Zheng

DOI: https://doi.org/10.33545/26649845.2025.v7.i2k.458

Abstract
Guangdong Province, a major player in China's foreign trade, has consistently led the nation in total import and export volume, leveraging its geographical advantage of being adjacent to Hong Kong and Macao, a well-developed manufacturing sector, and a comprehensive industrial chain support. In 2023, Guangdong's total import and export value soared to RMB 8.3 trillion, representing 19.9% of the country's total import and export value. To delve into the dynamic shifts and principal drivers behind Guangdong's foreign trade performance, this study employs three machine learning (ML) models—Support Vector Machine (SVM), Neural Network (NN), and Random Forest (RF)—based on 16 key indicators from 2006 to 2022. Additionally, the SHAP method is utilized to assess and rank the importance and interpretability of these influencing factors, pinpointing the top five core indicators that significantly impact Guangdong's total import and export volume. The results of this research offer a quantitative basis for optimizing Guangdong's foreign trade policy system. By focusing on these key factors, policymakers can craft tailored strategies to foster regional industrial upgrading, deepen regional cooperation within the RCEP framework, enhance cross-border trade facilitation, and nurture new forms of foreign trade. This approach will bolster policy precision and execution, aiding Guangdong Province in building a higher level of open economy.

Keywords: Regional Economic Development; Import and Export Trade; Explanatory Power Analysis; Machine Learning Model

1. Introduction

As the core region of China's opening-up and a pivotal province for foreign trade, Guangdong Province has consistently led the nation in total import and export volume. Its geographical proximity to Hong Kong and Macao, coupled with a comprehensive manufacturing system and highly internationalized industrial clusters, has been instrumental in this achievement. In 2023, Guangdong's total import and export value soared to RMB 8.3 trillion, marking a 0.3% year-on-year increase, with exports reaching RMB 5.4 trillion (up 2.5%) and imports at RMB 2.9 trillion (down 3.6%). In recent years, Guangdong's trade network has covered more than 230 countries and regions worldwide, with a diversified trade structure dominated by electromechanical and high-tech products. Despite facing external challenges such as a sluggish global economic recovery, rising trade protectionism, and geopolitical conflicts, Guangdong Province has sustained steady growth in its import and export scale and structural optimization. This has been achieved through measures like deepening the "Five Externals Linkage" strategy, promoting trade facilitation reforms in the Guangdong-Hong Kong-Macao Greater Bay Area, and nurturing comprehensive foreign trade service enterprises, showcasing remarkable resilience in its foreign trade sector. However, with the global value chain restructuring, rising domestic factor costs, and the emergence of new trade barriers like carbon tariffs, Guangdong's foreign trade development is encountering deep-seated issues, including weakening traditional advantages, lagging cultivation of new growth points, and imbalanced regional development. To overcome these constraints, it is imperative to transcend traditional econometric analysis frameworks and leverage cutting-edge technological methods such as ML to identify core indicators influencing Guangdong's total import and export volume. This will provide precise data support and decision-making references for the Guangdong Provincial Government to formulate more targeted, scientific, and forward-looking foreign trade policies, aiding

Guangdong in achieving its strategic goal of high-quality foreign trade development in the context of profound global trade landscape adjustments.

Current research on import and export trade and economic development primarily focuses on the influencing factors of import and export trade and the application of ML models in economic research. In terms of research on the influencing factors of import and export trade, Zhang Jiao studied the trade volume between China and RCEP member countries, revealing that China's GDP significantly impacts exports, while exchange rate effects vary by trading partner [1]. Zhenghuang Shi and Fangmiao Hou explored how fluctuations in imported timber prices affect the export margins of wood forest products across Chinese provinces. Their findings indicate that such price volatility heightens uncertainty, thereby suppressing the growth of China's wood forest product export share, extensive margin, and quantity margin, while conversely boosting the price margin [2]. Using panel data from 31 Chinese provinces and regions, Hongya Li *et al*. delved into the impact of digital inclusive finance on imports, exports as well as the overall foreign trade balance. Their research uncovered a notably positive effect of digital inclusive finance on China's trade performance [3]. Willem Thorbecke and Ahmet Sengonul investigated how exchange rates affect Turkish imports and exports. Nonlinear autoregressive distributed lag analysis revealed that appreciations during appreciation episodes stimulate both imports and exports, whereas exchange rate fluctuations during depreciation periods often exert no significant impact on trade [4]. Cheng YuHe analyzed the factors influencing Guangxi Zhuang Autonomous Region's total import and export volume by constructing a multiple regression model. The research results highlighted that local financial general budget expenditure and urban residents' per capita disposable income are the most pivotal factors affecting the total import and export trade of Guangxi from 2002 to 2021 [5]. Mingzhe Yu *et al*. empirically assessed the impact of US trade policy uncertainty on China's agricultural imports and exports, utilizing both aggregate and product-level data from China. Their findings indicate that heightened US trade policy uncertainty has diminished China's agricultural exports to the US and imports from the US, with effects surpassing those of traditional exchange rate factors [6].

In the realm of applying ML models to economic development research, Du ErLe and Ji Meng applied ML-based Data Envelopment Analysis (DEA) and Malmquist index measurement algorithms to analyze the dynamic and static characteristics of economic data from China's high-tech industrial development zones (IDZs). Their findings indicate that while the overall economic efficiency of all national high-tech IDZs in China is relatively high. However, there are huge differences among different regions [7]. Zhixiang Yin *et al*. took China's national emergency industry demonstration base (NEDB) as a quasi-natural experiment. They employed a double machine learning (DML) model to explore its impact on and the underlying mechanisms of regional economic resilience. The results reveal that the construction of NEDB significantly bolsters regional economic resilience [8]. Zhiqiang J combined ML models with regional economic scale prediction and management, discovering that ML substantially improves the accuracy of the regional economic forecasting and reduces the required forecasting time [9]. In order to understand the dynamics and influencing factors of green economy development across 287 Chinese cities, Rui Ding *et al*. introduced SHAP (SHapley Additive exPlanations)-interpreted RF modeling. Their research identified consumption, financial development, technological innovation, and business development as the primary driving factors [10]. Hongli Jiang *et al*. investigated the influence mechanism of green fiscal policies on carbon emission efficiency (CEE), the impact of the interplay between green fiscal policies and economic growth target (EGT) constraints on CEE, and the role of digital infrastructure. They used a staggered DID approach and a DML approach. The results revealed that green fiscal policies enhance CEE by promoting low-carbon technology innovation, improving energy efficiency, and facilitating industrial structure upgrading [11]. Mohamed F. Abd El Aal *et al*. utilized machine-learning algorithms to identify the economic sectors that most significantly influence Saudi Arabia's economic growth rate, focusing on agriculture, industry, and services. Their analysis showed that the RF algorithm provided the highest predictive accuracy in pinpointing the key sectors driving economic growth. The findings indicate that the service and industrial sectors contribute 39.3% and 37.7%, respectively, to Saudi Arabia's GDP growth [12].

Based on the aforementioned research, this study constructs three types of ML models: SVM, NN, and RF, and combines the SHAP method to analyze the explanatory power of 16 key indicators that influenced Guangdong Province's total import and export volume from 2006 to 2022. The objective is to furnish data-driven support for optimizing foreign trade policy strategies and fostering new competitive advantages internationally, thereby helping Guangdong Province reinforce its leading position in foreign trade as it establishes a new development paradigm.

2. Construction of ML Model

2.1 Selection of Model Indicators

In this study, the total import and export volume of Guangdong Province is designated as the explained variable, representing the overall development level of its import and export trade. Based on data from the Statistical Yearbook of Guangdong Province, 16 economic indicators from 2006 to 2022 are selected as explanatory variables, reflecting five aspects: macroeconomics, trade policy and openness, logistics and infrastructure, technology and innovation, as well as exchange rate and finance (Table 1).

Table 1: Model Indicators System

Category	Indicator
	Total import and export volume (RMB billion) I&E
Macroeconomics	Gross regional domestic product (RMB billion) RGDP
	Fixed assets investment (RMB billion) FAI
	Total retail sales of consumer goods (RMB billion) RSC
	The proportion of the total output value of the secondary industry (%) PSI
	GDP growth rate of the United States (%) GDPUS
	GDP growth rate of the European Union (%) GDPEU
	GDP growth rate of the five ASEAN countries(%) GDPASEAN
Trade policy and openness	China's tariff revenue (RMB billion) CTR
	Actual utilization of foreign direct investment (RMB billion) FDI
Logistics and infrastructure	Port cargo throughput (10,000 tons) PCT
	Cargo throughput of civil aviation terminals (10,000 tons) ACT
	Number of developed ports of the first category (unit: piece) DPFC
Technology and innovation	Research and Experimental Development personnel (in number of people) R&DP
	Internal expenditure on research and experimental development funding (10,000 RMB) R&DF
Exchange rate and finance	Exchange rate of RMB against USD (CNY/100USD) ER
	Scale of foreign exchange reserves (USD billion) FER

2.2 Sample Division and Data Preprocessing

The construction of an ML model requires dividing the overall sample into two parts: the training set and the test set. The training set is used to build the ML model, while the test set evaluates its performance. In this study, 80% of the overall sample data is allocated to the training set, with the remaining 20% assigned to the test set. For data standardization, the standard deviation normalization method is adopted, enabling the model to better learn data feature columns with different dimensions and scales, as shown in Formula 1 below:

$$X' = \frac{X - X_{mean}}{X_{std}} \quad (1)$$

In Formula 1, X' represents the standardized data; X denotes the original data; X_{mean} signifies the mean of the original data; and X_{std} stands for the standard deviation of the original data.

2.3 Selection of Model Evaluation Indicators

This study selects the RMSE as the evaluation metric for the model, as shown in Formula 2 below:

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

In Formula 2, N represents the sample size; y_i represents the true value of the dependent variable for the sample; and $\hat{y}_i$ denotes the predicted value of the model for the dependent variable of the sample.

The inclusion of a square root in the calculation formula effectively avoids the issue of square units appearing in the results. Compared to the MSE metric, the RMSE metric does not have a square term, thus preventing the magnification of the gap between abnormal predicted values and actual values and improving the evaluation of model's predictive ability. A smaller RMSE value indicates better prediction performance on the test set data, while a larger RMSE value indicates that the model's prediction performance on the test set data is worse.

2.4 Construction of SVM Model

This article primarily constructs two types of SVM models: the linear kernel model and the radial kernel model, comparing their predictive abilities based on the RMSE metric. To identify the optimal hyperparameter combinations for each model, grid search is employed to traverse the values of the three hyperparameters, "C", "epsilon", and "gamma", for both models. According to the screening results, the optimal hyperparameter combination for the linear kernel model is "C" set to 1, "epsilon" set to 0.01, and "gamma" set to 0.001; the optimal hyperparameter combination for the radial kernel model is "C" set to 100, "epsilon" set to 0.062, and "gamma" set to 0.001. Based on these results, the linear and the radial kernel models are trained, and their RMSE metrics are calculated. The RMSE metric for the linear kernel model is 0.082, while that for the radial kernel model is 0.079. The results above indicate that the radial kernel model exhibits better predictive performance on the test set data. Therefore, the radial kernel model is selected as the representative SVM model for analyzing the explanatory power of the indicators.

2.5 Construction of NN Model

This paper primarily constructs two types of NN models: single hidden-layer and double hidden-layer. The hyperparameter "max_iter" for both types of models is set to the default value of 10000, and the hyperparameter "random_state" for both types of models is set to the default value of 42. To select the optimal number of hidden layer neurons, i, for both types of models, a parameter traversal method is adopted. This method involves traversing all integers from 1 to 99 (inclusive) for the number of hidden layer neurons, i, and selecting the value of i that minimizes the model's RMSE metric as the optimal value for this hyperparameter. The results of parameter traversal indicate that the optimal number of neurons, i, for the single hidden layer model is 83, and the optimal number of neurons, i, for each layer of the double hidden layer model is 32. Based on these optimal hyperparameter combinations, this paper constructs two types of NN models: single hidden-layer and double hidden-layer. The RMSE metric for the single hidden layer model is 0.032, while that for the double hidden layer model is 0.027. Therefore, the double hidden layer model demonstrates better predictive performance on the test set data and is selected as the representative NN model for analyzing the explanatory power of the indicators.

2.6 Construction of RF Model

This paper primarily constructs two types of RF models with parameter max_features set to "sqrt" and "None", respectively. Furthermore, the parameter n_estimators for these two types of RF models is selected based on the OOB error rate metric. The parameter selection results indicate that for the RF model with max_features set to "None", the optimal value for parameter n_estimators is 15. Similarly, for the RF model with max_features set to "sqrt", the optimal value for parameter n_estimators is also 15. Based on these parameter selection results, the two types of RF models are constructed. The results show that the RMSE metric for the RF model with max_features set to "sqrt" is 0.119, while that for the RF model with max_features set to "None" is 0.176. Based on these results, the RF model with max_features set to "sqrt" is selected for subsequent analysis of the explanatory power of the metric.

3 Analysis of Explanatopy Power of Influencing Factor Indicators

Based on the construction of ML models, this paper employs the SHAP tool to analyze the explanatory power of indicators, focusing on the radial kernel model, the double hidden layer model, and the RF model with a parameter max_features value of "sqrt" and a parameter n_estimators value of 15, which exhibit good predictive ability.

3.1 Analysis of the Explanatory Power of SVM Model's Indicators

An analysis and ranking of the explanatory power of indicators influencing Guangdong Province's total import and export volume were conducted using the SVM radial kernel model. The specific results are depicted in Figure 1.

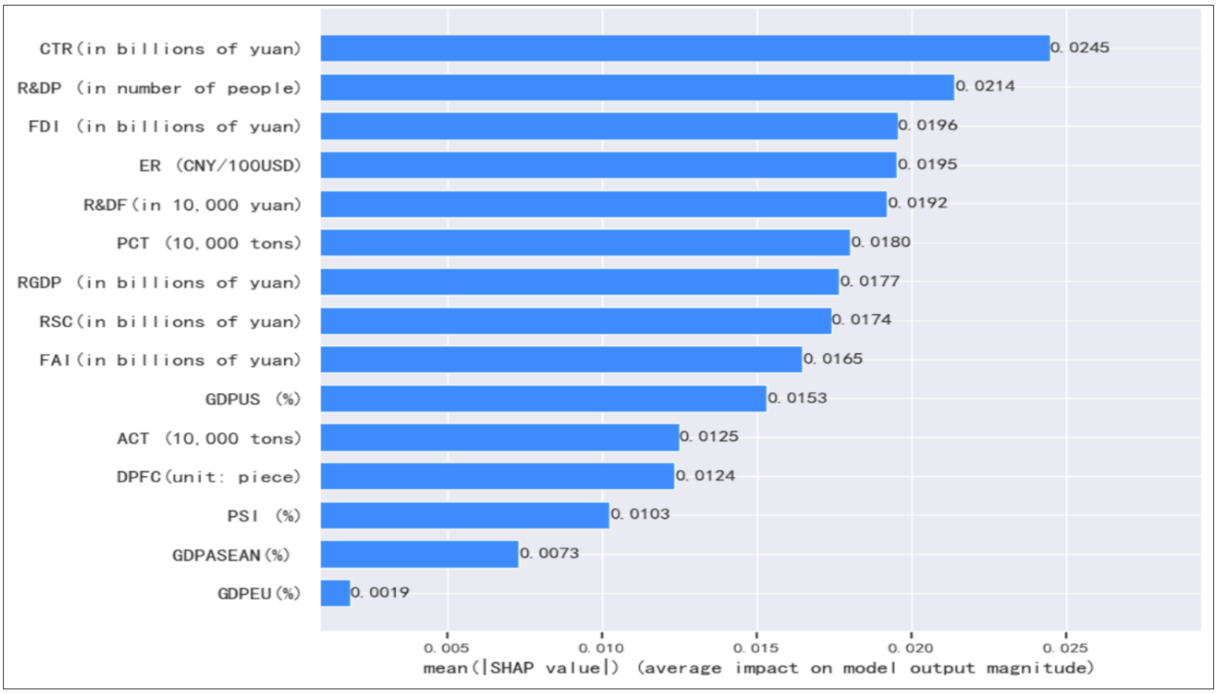

Fig 1: Bar Chart of SHAP for SVM Model

According to Figure 1, the top five indicators exerting the greatest impact on Guangdong province's total import and export volume in the SVM model are China's tariff revenue, research and experimental development (R&D) personnel, actual utilization of foreign direct investment, exchange rate of RMB against USD, and internal expenditure on R&D funding. Among these five indicators, two fall under the category of trade policy and openness, two under technology and innovation, and one under exchange rate and finance.

3.2 Analysis of the Explanatory Power of NN Model's Indicators

The ranking of the explanatory power of indicators for the NN double hidden layer model is presented in Figure 2. According to Figure 2, the top five factors significantly influencing Guangdong Province's total import and export volume in the NN model are the proportion of the total output value of the secondary industry, R&D personnel, port cargo throughput, cargo throughput of civil aviation terminals, and China's tariff revenue. Among these five indicators, two are categorized under logistics and infrastructure, one under macroeconomics, one under trade policy and openness, and one under technology and innovation.

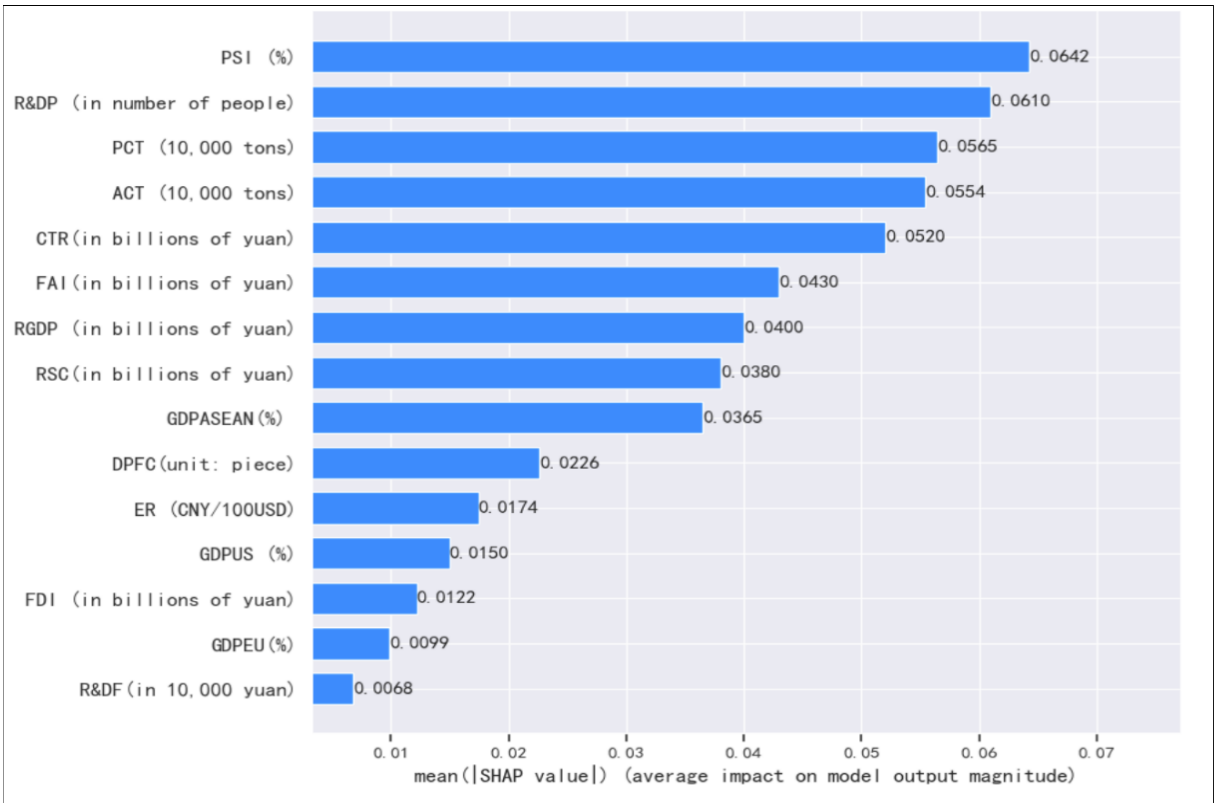

Fig 2: Bar Chart of SHAP for NN Model

3.3 Analysis of the Explanatory Power of RF Model's Indicators

The ranking of the explanatory power of indicators for the RF model is presented in Figure 3. According to Figure 3, the top five factors with the most significant impact on Guangdong's total import and export volume in the RF model are fixed assets investment, GDP growth rate of the five ASEAN countries, total retail sales of consumer goods, cargo throughput of civil aviation terminals, and the number of first-class developed ports. Among these five indicators, three belong to the macroeconomic category, and two fall under logistics and infrastructure category.

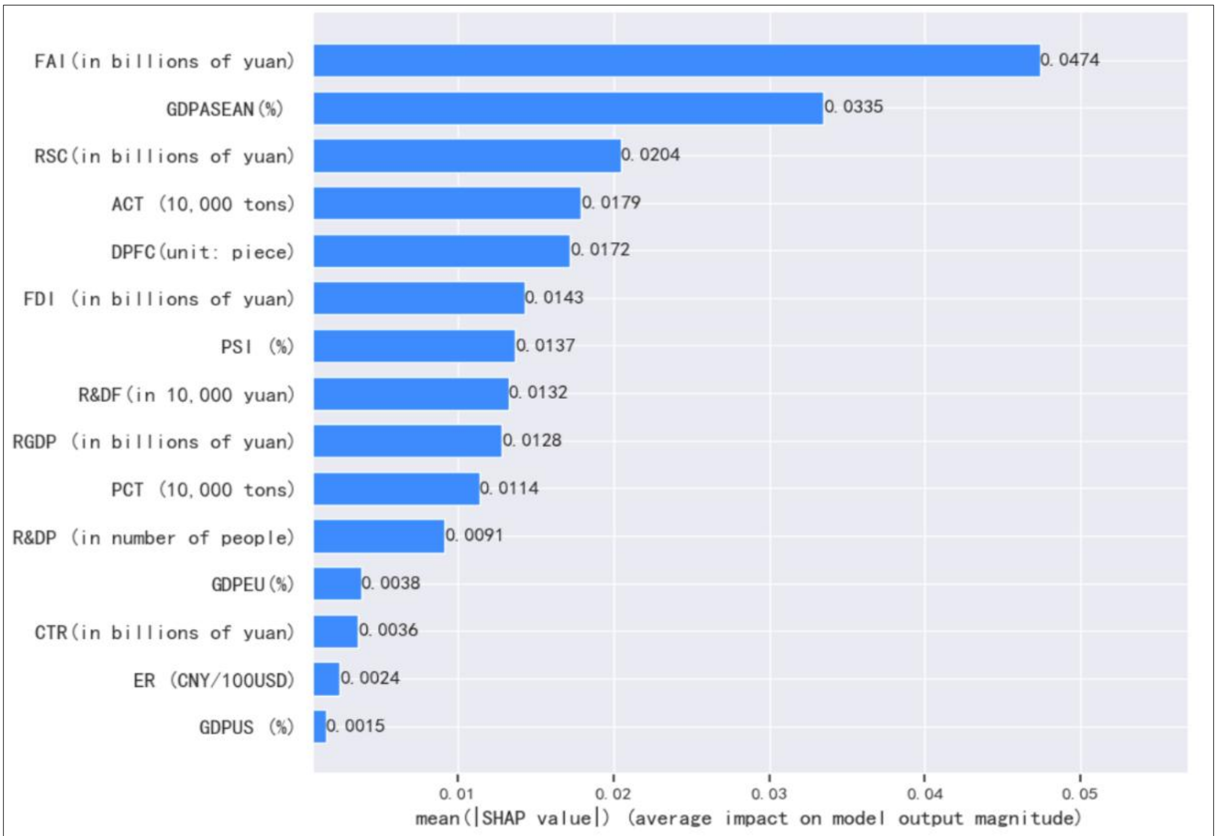

Fig 3: Bar Plot of SHAP for the RF Model

4 Conclusion and Suggestions

4.1 Conclusion

Based on the results of the explanatory power analysis of the above indicators, the indicators that have the greatest impact on Guangdong Province's total import and export volume are China's tariff revenue, R&D personnel, and cargo throughput of civil aviation terminals. These three indicators feature twice among the top five explanatory variables across the three types of ML models examined. Nine other indicators, including fixed assets investment, total retail sales of consumer goods, the proportion of the total output value of the secondary industry, GDP growth rate of the five ASEAN countries, actual utilization of foreign direct investment, port cargo throughput, number of first-class developed ports, internal expenditure on R&D funding, and exchange rate of RMB against USD, also significantly affect Guangdong Province's total import and export volume. Each of these appears among the top five influential indicators in only one type of ML model. The remaining four indicators have a relatively minor impact, as they do not rank among the top five in any of the ML models.

In terms of the impact of indicator categories on Guangdong Province's total import and export volume, logistics and infrastructure, as well as macroeconomic indicators, collectively appear four times in the top five influential indicators across the three types of ML models. Therefore, the construction and development of these two areas have the most significant impact on Guangdong Province's import and export trade. Secondly, indicators from the trade policy and openness, as well as technology and innovation categories, appeared a total of three times in the analysis of the explanatory power of indicators across the three types of ML models. Hence, the development of these two areas has a relatively high impact on Guangdong Province's import and export trade. Lastly, indicators from the exchange rate and finance category only appeared once in the analysis of the explanatory power of indicators across the three types of ML models. Therefore, this area has a relatively low impact on Guangdong Province's import and export trade.

4.2 Suggestions

Based on the research conclusions, in order to further optimize the import and export trade structure and enhance trade competitiveness in Guangdong Province, this paper puts forward the following suggestions from the perspectives of policy formulation, industrial layout, technological innovation, and regional collaboration.

Given the significant impact of China's tariff revenue on Guangdong's total import and export volume, the province should dynamically adjust its tariff structure. This adjustment should prioritize reducing import tariffs on high-tech intermediate goods and key equipment. Additionally, Guangdong Province can pilot tariff reduction and clearance facilitation reforms through platforms like pilot free trade zones and comprehensive bonded zones to lower trade costs for enterprises. Considering the crucial role of cargo throughput at civil aviation terminals, Guangdong Province needs to accelerate the coordinated development of the airport cluster in the Guangdong Hong Kong Macao Greater Bay Area. This involves increasing the density of international cargo routes, enhancing special cargo capabilities (such as cold chain and cross-border e-commerce) at hub airports like Guangzhou and Shenzhen, and forming a three-dimensional logistics system of "air rail intermodal transportation". The number of R&D personnel plays a pivotal role in Guangdong Province's import and export trade. Therefore, the province should implement the "Guangzhou Shenzhen Hong Kong Macao Science and Technology Innovation Corridor" talent special plan. This plan should attract top global technical talents through policies such as tax incentives, housing subsidies, and the distribution of research results conversion benefits. It should also promote joint training of applied R&D talents between universities and enterprises within the province.

In terms of macroeconomic construction, Guangdong Province can formulate differentiated investment promotion policies to guide foreign investment into high-end manufacturing, green energy and other fields. To stimulate domestic demand, the province can issue consumption vouchers, cultivate local brands, and expand the Southeast Asian market through cross-border e-commerce comprehensive pilot zones. In terms of logistics and infrastructure construction, Guangdong Province should take Guangzhou Port and Shenzhen Port as the core, promote the intelligent transformation of port clusters, build 5G smart ports, and increase the proportion of automated container loading and unloading. Moreover, Guangdong Province needs to improve its multimodal transport system, accelerate the construction of the "Guangdong Hong Kong Macao Greater Bay Area Combined Port", achieve interconnectivity of maritime, railway, and highway data, and reduce the cost of inland goods transit. In terms of designing and implementing trade policies, Guangdong Province should align its trade policies with the rules under the RCEP framework, deepen tariff reciprocity with ASEAN countries, and establish trade promotion mechanisms tailored to different countries. In the field of technological development and innovation, Guangdong Province can link internal R&D funding with industry demand, focusing on supporting the development of key technologies in areas like semiconductors and biomedicine, and improving the industrial transformation efficiency of R&D investment.

In order to enhance dynamic monitoring and early warning capabilities of risks, Guangdong Province has built a real-time monitoring platform and combined ML models to predict the short-term impact on imports and exports of indicators with strong volatility, such as the GDP growth rate of the five ASEAN countries and exchange rate, providing data support for policy adjustments. In promoting the coordinated development of regional industries and green transformation, Guangdong Province should advance the construction of the Guangdong Hong Kong Macao Greater Bay Area and jointly build an international trade platform leveraging Hong Kong and Macao's advantages in finance, law, and professional services. This will achieve seamless connections in customs clearance, logistics, payment, and other links.

By constructing an ML model and conducting indicator interpretation analysis based on the SHAP method, this study reveals the core driving effects of tariff policies, logistics efficiency, and R&D talents on Guangdong Province's import and export trade. It also clarifies the differential impact mechanisms in areas like logistics infrastructure, macroeconomic environment, and trade policy innovation. The research findings not only provide a quantitative basis for Guangdong Province to optimize its trade structure and enhance its position in the global value chain but also highlight the practical value of data-driven decision-making in complex economic systems. Looking ahead, Guangdong Province needs to respond to external uncertainties with dynamic policies, unleash domestic demand potential through regional synergy, and align with new global trade rules through green transformation to continue playing a role in building a new development pattern. This study also offers a replicable methodological framework for other outward-oriented economies. Further research can incorporate non-linear relationships and cross-regional spillover effects analysis to deepen the understanding of the dynamic evolution of open economies.

References

1. Zhang J. Economic influencing factors of trade volume between China and RCEP member states. Proceedings of Business and Economic Studies. 2023;6(3):14–19. doi:10.26689/PBES.V6I3.5082.
2. Shi H, Hou M. The impact of imported timber price fluctuations on the margins of wood forest product exports: empirical evidence from provincial-level data in China. Canadian Journal of Forest Research. 2025;55:1–18. doi:10.1139/cjfr-2024-0213.
3. Li Y, Lin X, Zhang Z. How does digital finance affect imports, exports and trade balance: Evidence from China. International Review of Economics and Finance. 2025;99:104054. doi:10.1016/j.iref.2025.104054.
4. Thorbecke W, Sengonul A. The impact of exchange rates on Turkish imports and exports. International Economics. 2023;174:231–249. doi:10.1016/j.inteco.2023.04.003.
5. Cheng Y. Study on the influence factors of Guangxi's total exports based on ridge regression and LASSO regression. Trends in Social Sciences and Humanities Research. 2024;2(4):28–32. doi:10.61784/TSSHR3007.
6. Yu M, Fan J, Wang H, Wang J. US trade policy uncertainty on Chinese agricultural imports and exports: An aggregate and product-level analysis. International Review of Economics and Finance. 2023;84:70–83. doi:10.1016/j.iref.2022.10.018.
7. Du E, Ji M. Analyzing the regional economic changes in a high-tech industrial development zone using machine learning algorithms. PLoS One. 2021;16(6):e0250802. doi:10.1371/journal.pone.0250802.
8. Yin Z, Yin Z, An Z, Hao L. Can emergency industry demonstration base promote regional economic resilience? Finance Research Letters. 2025;77:107019. doi:10.1016/j.frl.2025.107019.
9. Jiang Z. Prediction and management of regional economic scale based on machine learning model. Wireless Communications and Mobile Computing. 2024;2024:2083099. doi:10.1155/2022/2083099.
10. Ding R, Ran X, Jiang S, Zhang B. Interpretable machine learning for analyzing the spatial-temporal evolution and influencing factors of green economic development. Expert Systems with Applications. 2025;283:127863. doi:10.1016/j.eswa.2025.127863.
11. Jiang H, Hu W, Hou Y, Jiang P, Lu N. Green fiscal policies, economic growth target constraints and carbon emission efficiency: Evidence from DID and double machine learning. Energy. 2025;339:138961. doi:10.1016/j.energy.2025.138961.
12. Abd El-Aal MF, Shrahili M, Kayid M, Mohammad S. GDP growth drivers in Saudi Arabia based on machine learning algorithms. Journal of Radiation Research and Applied Sciences. 2025;18(2):101380. doi:10.1016/j.jrras.2025.101380.