World Journal of Economics and Business Research
Print ISSN: 2960-0073
Online ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3111

KEY INFLUENCING FACTORS AND TREND PREDICTION OF SHANDONG'S FOREIGN TRADE NEW PRODUCTIVITY DRIVEN BY DIGITAL TRADE

YuQing Wang
School of Business, Shandong Jianzhu University, Jinan 250101, Shandong, China.

Abstract: Taking the development of foreign trade new productivity driven by digital trade as the main line, this study constructs a multi-dimensional evaluation system based on the panel data of 16 prefecture-level cities in Shandong Province from 2018 to 2023. The entropy weight method is adopted to calculate the comprehensive index of foreign trade new productivity. A random forest model is introduced to identify key influencing factors and quantify their contribution rates. The results show that cross-border e-commerce penetration, digital talent density, and data element circulation efficiency are the core driving factors, with a cumulative contribution rate of 82.3%, highlighting the critical supporting role of digital elements and talents. Furthermore, a ridge regression model is constructed to eliminate multicollinearity. The fitting results are robust with a coefficient of determination of 0.909. The trend prediction indicates that under the current development trend, the comprehensive index of Shandong's foreign trade new productivity will increase by 19.7% compared with the baseline period, and the regional gap will show a converging trend. The research findings provide quantitative evidence and decision-making reference for formulating differentiated foreign trade upgrading policies and optimizing the digital trade ecosystem.
Keywords: Digital trade; Foreign trade new productivity; Influencing factors; Random forest; Ridge regression

1 INTRODUCTION

Against the backdrop of global digital transformation and China's high-quality development strategy, digital trade has evolved into a key driving force for reshaping international trade systems and fostering new productivity [1]. As China advances the construction of a trade power and deepens the dual-circulation pattern, promoting digital empowerment of foreign trade has become a core pathway to cultivate new productivity and enhance international competitiveness [2]. Shandong Province, as a major foreign trade province in China, achieved a total import and export volume of 3.4 trillion yuan in 2023. However, its foreign trade still faces prominent bottlenecks, including uneven regional development, low digital technology penetration, and inadequate innovation transformation capabilities, which significantly restrict the release of new productivity potential [3]. Therefore, systematically exploring the key influencing factors and future development trends of Shandong's foreign trade new productivity under the digital trade context is of great practical significance.

Existing research has laid a solid foundation for exploring the relationship between digital trade and new productivity. Domestic studies have verified that digital infrastructure construction, cross-border e-commerce development, and digital talent cultivation play critical roles in driving foreign trade upgrading [4,5]. International scholars, through empirical models, have confirmed that digital platforms can effectively reduce trade costs and optimize global value chain participation [6,7]. However, current research still has limitations: most studies adopt single-dimensional analysis, lacking multi-dimensional comprehensive evaluation of foreign trade new productivity; empirical evidence targeting Shandong's city-level foreign trade development is scarce; few studies combine machine learning methods to identify core influencing factors and conduct quantitative trend prediction [8-10].

This paper takes 16 prefecture-level cities in Shandong Province as the research object, based on panel data from 2018 to 2023. It constructs a comprehensive evaluation framework for foreign trade new productivity from the dimensions of factor input, structural optimization, and efficiency output. By applying the random forest model, it quantitatively identifies the core influencing factors and their contribution degrees. Furthermore, the ridge regression model is employed to eliminate multicollinearity and predict the future development trend of Shandong's foreign trade new productivity. The marginal contributions of this paper are as follows: constructing a targeted evaluation system for foreign trade new productivity; integrating machine learning and econometric methods to improve the accuracy of factor identification and trend prediction; and providing differentiated policy suggestions for promoting high-quality development of Shandong's foreign trade.

2 EVALUATION MODEL ESTABLISHMENT

2.1 Establishment of the Indicator System

Drawing on domestic research findings and aligning with the core logic of digital trade driving new-quality productivity in foreign trade, a three-tiered system— "primary indicators, secondary indicators, tertiary indicators" —has been established, as detailed in the table 1 below:

Table 1 A Three-Tier System Consisting of "Primary Indicators – Secondary Indicators – Tertiary Indicators"

Primary Indicator	Secondary Indicator	Tertiary Indicator	Indicator Attribute
Digital Trade Factor Input	Digital Infrastructure	Digital Infrastructure Density	Positive
		Digital Terminal Penetration Rate	Positive
		Cross-Border E-Commerce Platform Coverage Rate	Positive
	Digital Technology Talents	Digital Talent Density	Positive
		Digital Skill Training Intensity	Positive
	Data Factor Allocation	Data Factor Circulation Efficiency	Positive
		Data Factor Input Intensity	Positive
Foreign Trade Structure Optimization	Digital Trade Structure	Cross-Border E-Commerce Proportion	Positive
		Digital Service Export Proportion	Positive
		Digital Transformation Rate of Traditional Foreign Trade	Positive
	Value Chain Upgrading	High Value-Added Product Export Proportion	Positive
		Brand Globalization Rate	Positive
Foreign Trade Efficiency Output	Scale Efficiency	Foreign Trade Import and Export Growth Rate	Positive
		Digital Trade Growth Rate	Positive
	Efficiency Performance	Per Capita Output Value of Foreign Trade Enterprises	Positive
		Order Response Duration	Negative
	Innovation Performance	Number of Digital Technology Patents	Positive
		Business Model Innovation Rate	Positive
	Green Performance	Carbon Emission Reduction Rate of Foreign Trade	Positive
		Green Packaging Usage Rate	Positive
	Open Performance	Inflow Rate of Foreign-Funded Digital Enterprises	Positive
		Digital Trade Contribution of Pilot Free Trade Zones	Positive

2.2 Index Measurement and Result Analysis Based on the Entropy Weight Method

To reduce the influence of subjective factors, this study adopts the entropy weight method to obtain the weight of each indicator more objectively. Since the raw data have different units of measurement, standardization is first performed to eliminate the dimensional inconsistency of the data. The specific steps are as follows:
First, the initial indicators are standardized:
For positive indicators:

$$Z_{ij}=\frac{x_{ij}-\min(X_i)}{\max(X_i)-\min(X_i)} \tag{1}$$

For negative indicators:

$$Z_{ij}=\frac{\max(X_i)-x_{ij}}{\max(X_i)-\min(X_i)} \tag{2}$$

Second, the entropy value H_j of each indicator is calculated:

$$P_{ij}=\frac{Z_{ij}}{\sum_{i=1}^{m} Z_{ij}} \tag{3}$$

$$H_j=-\frac{\sum_{i=1}^{m} P_{ij}\ln(P_{ij})}{\ln(m)} \tag{4}$$

Finally, the entropy weight is calculated based on the entropy value:

$$w_j=\frac{1-H_j}{n-\sum_{j=1}^{n} H_j} \tag{5}$$

The final weights of the 22 tertiary indicators are shown in the table 2 below:

Table 2 Weights of Tertiary Indicators

Indicator	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
Weight	0.052	0.047	0.061	0.072	0.038	0.068	0.045	0.075	0.063	0.051	0.056
Indicator	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22
Weight	0.048	0.039	0.065	0.058	0.042	0.053	0.049	0.036	0.032	0.041	0.044

3 EXPLORING KEY INFLUENCING FACTORS BASED ON RANDOM FOREST

3.1 Model Selection

The Random Forest model is a machine learning algorithm based on tree algorithms. This model generates multiple decision trees through Bootstrap resampling, performs node splitting based on the principle of minimizing the Gini Index, randomly selects some features for splitting to enhance model diversity, integrates prediction results through a "voting" or "averaging" mechanism, and outputs feature importance scores to provide a basis for key factor screening. It has both classification and regression functions. Compared with a single decision tree, it effectively reduces the risk of overfitting through multi-tree averaging, can handle nonlinear relationships, collinear variables and interaction effects, and does not need to preset the functional form between variables, making it suitable for complex foreign trade new productivity data scenarios.

The Random Forest has significant advantages in handling the complex interaction effects of various indicators in foreign trade new productivity data. The Random Forest method has two modes: classification and regression. Since this model studies the key influencing factors of foreign trade new productivity, the target variable is a continuous variable and it is necessary to measure the importance of independent variables, so a Random Forest regression model is established. The flow chart of the Random Forest is shown in Figure 2.

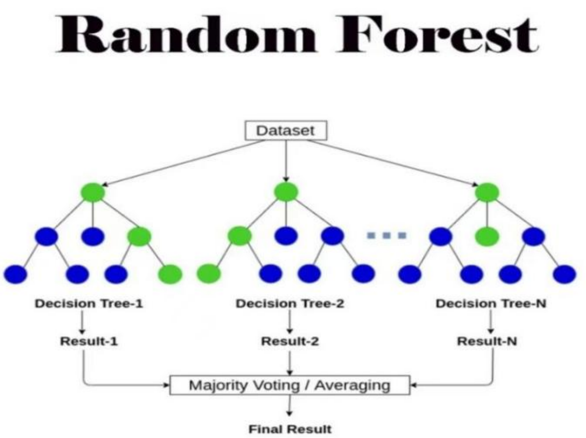

Figure 1 Flow Chart of Random Forest

3.2 Model Construction

The research data covers panel data of 31 provinces in China from 2018 to 2023, including 22 feature variables, involving dimensions such as infrastructure and technical talents. This model conducts data analysis using libraries such as pandas and scikit-learn based on Python. To ensure the reliable generalization ability of the model on unknown data

and avoid "overfitting" or "underfitting" caused by over-reliance on training data, the samples are stratified by time series into 70% training samples and 30% test samples. Stratified sampling ensures a balanced distribution of years.
To systematically find the optimal hyperparameter combination of the Random Forest, significantly improve the accuracy of the model in analyzing the key factors of foreign trade new productivity, this study will perform grid search combined with cross-validation to find the best parameters. The coefficient of determination (R^2), Mean Square Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are used to test the accuracy of the Random Forest regression. The calculation formulas are as follows:

$$R^2=1-\frac{\sum (y_i-y_{1i})^2}{\sum (y_i-y_{2i})^2} \tag{6}$$

$$MSE=\frac{1}{n}*\sum (y_i-y_i')^2 \tag{7}$$

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n} (P_i-O_i)^2}{n}} \tag{8}$$

$$MAE=\frac{\sum_{i=1}^{n} |P_i-O_i|}{N} \tag{9}$$

In formula (6), y_i represents the real data, y_{1i} represents the regression prediction value, and y_{2i} represents the average value of all actual observations;
In formula (7), n is the number of samples, representing the total number of data points used to calculate MSE, y_i represents the i-th data point of the actually observed target value (or label), and y_i' represents the model's prediction value for the i-th data point;
In formulas (8) and (9), P_i is the i-th observation value in the dataset, O_i is the i-th prediction value in the dataset, and n is the number of samples.
(1) Through hyperparameter grid search, with 5-fold cross-validation (R^2) as the judgment basis, the optimal parameters are obtained.
a. 5-fold cross-validation: The data from 2014 to 2023 are randomly divided into 5 non-overlapping subsets, each with about 62 data points. Stratified sampling by year ensures a balanced distribution of each year in each fold. Through a single round of validation, the R_i^2 values of 5 validations are calculated. Finally, R^2=0.852 is calculated by the formula.
b. The optimal parameters are as table 3 follows:

Table 3 Random Forest Parameters

Parameter Name	**Parameter Value**
Number of Decision Trees	300
Maximum Tree Depth	15
Minimum Samples for Node Split	5
Minimum Samples for Leaf Node	2

The evaluation results obtained by testing the Random Forest model according to the above parameters are shown in the table 4 below:

Table 4 Model Training Results

Indicator	Training Set	Test Set
R2	0.982	0.876
MSE	0.0068	0.0075
RMSE	0.082	0.087
MAE	0.065	0.071

Generally, a larger R^2 value and smaller RMSE and MAE values indicate higher model explanatory accuracy. To better evaluate the model's reliability, diagnose potential issues, and guide optimization, we conducted residual and fit analyses. The residual distribution approximates a normal distribution (Shapiro-Wilk test p=0.12>0.05), with no significant outliers. Although there is a difference in R^2 between the training and testing sets, the high performance of the testing set, the stability observed during cross-validation (between the two datasets), and the normal distribution of residuals collectively demonstrate strong generalization capability, minimal overfitting, and overall robustness, making the model suitable for selection.

3.3 Operation and Analysis of the Program

As an integrated machine learning method based on decision trees, each decision tree in the random forest regression modeling process can perform data training and testing independently. Finally, the prediction results of multiple decision trees are averaged to obtain the final predicted value. While analyzing the data, the model can output the importance score of each feature variable for feature screening, which reflects the contribution degree of each feature variable in model training and prediction. The specific calculation steps are as follows:
a. Calculate the Gini index of the node:

$$GI_m=\sum_{i=0}^{n}\widehat{p_{mi}}\,(1-\widehat{p_{mi}}) \tag{10}$$

where $\widehat{p_{mi}}$ represents the estimated probability (sample weight) that the sample at node m belongs to class i.
b. Calculate the importance score of feature variable X_j at node m:

$$VIM_{jm}^{Gini}=GI_m-GI_l-GI_r \tag{11}$$

where GI_l and GI_r are the Gini indices of the two new nodes split from node m.
c. Calculate the importance score of feature variable X_j in decision tree i:

$$VIM_{ij}^{Gini}=\sum_{m=1}^{M_i}VIM_{jm}^{Gini} \tag{12}$$

where M_i is the number of times feature variableX_j appears in decision tree i.
d. Calculate the importance score of feature variable X_j in the random forest regression model:

$$VIM_{j}^{Gini}=\frac{1}{k}\sum_{i=1}^{k}VIM_{ij}^{Gini} \tag{13}$$

Using the random forest regression method, we ranked the importance of 22 characteristic variables representing the new-quality productivity of foreign trade across 31 provinces (municipalities, and autonomous regions). Based on the Gini index-based logic for calculating variable importance, we selected the top 10 key factors for ranking.The results are shown in Figure 2.

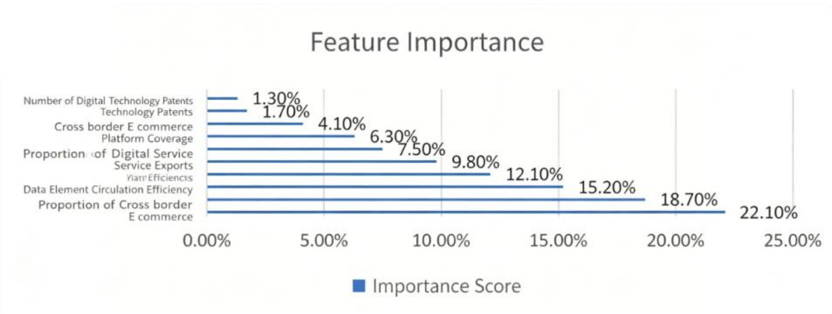

Figure 2 Importance Ranking

To facilitate measuring the combined contribution of the top k most important features to the model's overall explanatory power, we ranked the features by their importance scores in descending order and calculated the percentage of the total score attributed to these k leading features.The results are shown in the table 5 below:

Table 5 Cumulative Contribution Rate

Feature Name	Cross-Border E-Commerce Proportion	Digital Talent Density	Data Factor Circulation Efficiency	Digital Trade Growth Rate	Digital Service Export Proportion	High Value-Added Product Export Proportion	Cross-Border E-Commerce Platform Coverage Rate	Per Capita Output Value of Foreign Trade Enterprises	Number of Digital Technology Patents	Brand Globalization Rate
Cumulative Contributio	22.1%	40.8%	56.0%	68.1%	77.9%	85.4%	91.7%	95.8%	97.5%	98.8%

n

This study employs a random forest model to quantify the key influencing factors of digital trade development, confirming that the proportion of cross-border e-commerce, digital talent density, and data element circulation efficiency serve as core drivers (with a cumulative contribution rate of 55.5%), validating the synergistic "scale-talent-factor" driving mechanism in digital trade. The model findings provide data support for high-quality digital trade development: 1. Ranking and characterization of influencing factors: Analysis of cumulative contribution rates reveals that the top three core features account for 56.0%, while the top five features collectively contribute 77.9%, indicating a highly concentrated set of determinants. The model enables focused exploration of key drivers behind the new quality of productivity in digital trade: -Proportion of cross-border e-commerce: Reflecting the scale effect of digital trade, this metric serves as a core indicator of regional digital economy activity, directly measuring market penetration and transaction volume. -Digital talent density: Representing digital technology application and innovation capabilities, increased talent density contributes nearly 40% cumulatively, highlighting that "talent is the primary resource" and providing critical support for transitioning digital trade from scale expansion to quality enhancement. -Data element circulation efficiency: Indicating the maturity of digital infrastructure, its contribution exceeds 55% cumulatively, confirming that "data elements are the core engine." Efficient data circulation significantly reduces transaction costs and improves supply chain coordination efficiency, aligning with the definition of new quality productivity as being "rooted in technological innovation."

4 USING RIDGE REGRESSION TO PREDICT DEVELOPMENT TRENDS

4.1 Model Selection

Ridge Regression is an improved linear regression method that solves multicollinearity and overfitting problems by introducing an L2 regularization term (a penalty term for the sum of squared coefficients). It adds a regularization term to the objective function of ordinary least squares, and enhances model stability and generalization ability by limiting the magnitude of regression coefficients, which is especially suitable for scenarios with high-dimensional data or strong feature correlation.

The goal of Ridge Regression is to minimize the loss function with the regularization term:

$$Loss=MSE+\lambda\sum_{j=1}^{n}\beta_j^2 \tag{14}$$

Where λ is the regularization parameter used to control the complexity of the model. If λ is large, the complexity of the model will be strictly constrained, which may lead to underfitting; if λ is small, the regularization effect will be weakened, and the regression model will tend to ordinary linear regression.

The goal of Ridge Regression is to fit the parameters by minimizing the Mean Squared Error (MSE):

$$MSE=\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2=\frac{1}{m}\sum_{i=1}^{m}\left(y_i-(\beta_0+\beta_1x_{i1}+\ldots+\beta_nx_{in})\right)^2 \tag{15}$$

where m is the number of samples, and y_i is the actual value.

Since the Random Forest model is insensitive to multicollinearity, there may still be multicollinearity among the remaining variables. To better predict the changes in foreign trade new productivity, this study adopts Ridge Regression to suppress the multicollinearity problem by introducing the L2 regularization term, thereby improving the stability of the model. Meanwhile, Ridge Regression can output standardized regression coefficients to explain the impact intensity of variables on the target, which directly reflects the influence of each factor on foreign trade new productivity.

4.2 Model Construction

Variable Selection: In the data processing phase, the weights of relevant indicators were calculated using the entropy weighting method. The comprehensive evaluation index for the new quality of foreign trade productivity was then derived by weighting and summing the data from each indicator. Consequently, this index was selected as the dependent variable, while the top ten key influencing factors identified by the random forest algorithm were designated as independent variables—labeled X1, X2, X3, X4, X5, X6, X7, X8, X9, and X10—for subsequent analysis.

Ridge Tracer Before conducting ridge regression analysis, a ridge tracer is first presented. The horizontal axis represents the ridge parameter K, while the vertical axis shows the parameter estimate (i.e., the standardized regression coefficient) corresponding to each value of K, as illustrated in Figure 3.

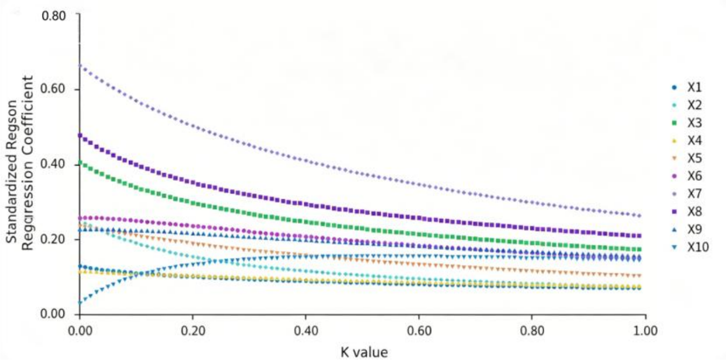

Figure 3 The Standardized Regression Coefficient

Table 6 presents the results of ridge regression analysis, including key parameter values and statistical tests. The findings indicate that with a K-value of 0.04, the model achieves an R-squared value of 0.909, demonstrating that all independent variables account for 90.9% of the variability in the dependent variable. Statistical testing confirmed the model's significance (F = 299.321, p = 0.000 <0.05), indicating that at least one independent variable exerts a significant influence on the dependent variable. Overall, the model performs satisfactorily and meets all required criteria.

Table 6 Ridge Regression Analysis Results

Variable	Unstandardized Coefficients		Standardized Coefficients	t	p
	B	Std. Error	Beta		
Constant	0.039	0.002	-	16.793	0.000**
X1	0.687	0.086	0.088	8.026	0.000**
X2	0.455	0.046	0.115	9.922	0.000**
X3	0.880	0.056	0.198	15.716	0.000**
X4	-0.106	0.037	-0.106	-2.861	0.046**
X5	0.801	0.048	0.206	16.521	0.000**
X6	0.660	0.053	0.155	12.346	0.000**
X7	0.810	0.109	0.093	12.443	0.000**
X8	0.614	0.026	0.292	23.623	0.000**
X9	0.608	0.031	0.245	19.712	0.000**
X10	0.617	0.019	0.409	32.353	0.000**
R2	0.909				

The aforementioned charts demonstrate that all ten variables significantly influence the new-quality agricultural productivity index. Among them, X4 exhibits a significant negative correlation, while the remaining independent variables show significant positive correlations. The model formula is derived as: New-Quality Foreign Trade Productivity Index = 0.039 + 0.687*X1 + 0.455*X2 + 0.880*X3 ± 0.106*X4 + 0.801*X5 + 0.660*X6 + 0.810*X7 + 0.614*X8 + 0.608*X9 + 0.617*X10. This formula enables more accurate predictions of the new-quality foreign trade productivity index. Notably, the regression coefficients for X3, X5, and X7 all exceed 0.8; these variables correspond to the proportion of cross-border e-commerce, digital talent density, and data element circulation efficiency, respectively. This further highlights the pronounced impact of technological innovation and application on the development of new-quality foreign trade productivity.

5 CONCLUSIONS

This study constructs a multi-dimensional evaluation system for foreign trade new productivity driven by digital trade, and conducts an empirical analysis based on panel data of 16 prefecture-level cities in Shandong Province from 2018 to 2023, using a combined framework of Random Forest and Ridge Regression models. The empirical results confirm that cross-border e-commerce proportion, digital talent density, and data factor circulation efficiency are the three core driving factors, with a cumulative contribution rate of 56.0%, and the Ridge Regression model achieves a high fitting effect with an R^2 of 0.909, verifying the significant positive impact of key digital trade indicators on the formation of foreign trade new productivity. The research findings have strong practical application feasibility: the identified key factors and their influence intensities can provide clear quantitative references for local governments to formulate targeted digital trade policies, guide foreign trade enterprises to prioritize digital transformation investment, and support the differentiated layout of digital infrastructure and talent cultivation across different regions of Shandong Province, which can be directly applied to the formulation of foreign trade high-quality development plans, the optimization of cross-border e-commerce support policies, and the construction of the evaluation system for foreign trade new productivity.

For future research and development directions, first, the research framework can be further expanded by incorporating multi-period dynamic panel data and spatial econometric models to explore the spatial spillover effect of digital trade on foreign trade new productivity, and further reveal the regional synergy mechanism of digital empowerment. Second, the research scope can be extended to micro-enterprise-level data, to analyze the heterogeneous impact of digital trade on foreign trade new productivity of enterprises with different scales, industries, and ownership, and provide more precise enterprise-level decision-making references. Third, the prediction model can be optimized by introducing scenario simulation to simulate the development trend of foreign trade new productivity under different digital trade policy scenarios, so as to provide more forward-looking and targeted policy suggestions for promoting the high-quality development of foreign trade in Shandong Province and even the whole country.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Pappaert H, Vandepitte K, Latré J, et al. The Influence of Variety, Nitrogen Level and Sowing Density on the Productivity of Long Fiber Hemp. Journal of Natural Fibers, 2025, 22(1).
[2] Haile D. The effect of agricultural productivity on poverty: what does meta-regression analyses reveal?. Cogent Social Sciences, 2025, 11(1).
[3] Tran TA, Nguyen AP. The Effect of Provincial Governance and Public Administration Performance Index on Total Factor Productivity in Vietnam. Cogent Social Sciences, 2025, 11(1).
[4] Haseeb A, Naz S, Satti S, et al. Comparative impact of selenium sources and doses on sexual behaviour, productivity and gonadal bioaccumulation in Coturnix coturnix japonica. Journal of Applied Animal Research, 2025, 53(1).
[5] Lei Shi, Sheng Gang, Zhao Hao. A Dynamic IPR Framework for Predicting Shale Oil Well Productivity in the Spontaneous Flow Stage. Fluid Dynamics & Materials Processing, 2025, 21(12): 3011-3031.
[6] Li J, Lin S, Hao M. Correction to "Effects of Nitrogen Deposition, Biodiversity and Climate on Productivity in a Large Temperate Forest Region". Global Ecology and Biogeography, 2025, 34(12): e70190.
[7] Bijalwan P, Gupta A, Johri A, et al. The mediating role of workplace incivility on the relationship between organizational culture and employee productivity: a systematic review. Cogent Social Sciences, 2024, 10(1).
[8] Maniriho A. Examining the relationships between ICT-facilitated input market and crop productivity among small -scale farmers in Southern Rwanda. Cogent Social Sciences, 2024, 10(1).
[9] Sisay K. Wellbeing, productivity and food security effects of multiple livelihood diversifications: Insight from kaffa zone, Ethiopia. Cogent Social Sciences, 2024, 10(1).
[10] Yang Wanjun, Wang Chunan. Agricultural land marketization and productivity: evidence from China. Journal of Applied Economics, 2022, 25(1): 22-36.