Volume 3, Issue 4, 2025

Print ISSN: 2960-0073

Online ISSN:2960-0081

# World Journal of Economics and Business Research

# World Journal of Economics and Business Research

Volume 3, Issue 4, 2025

Published by Upubscience Publisher

**World Journal of Economics and Business Research**
**Print ISSN: 2960-0073 Online ISSN: 2960-0081**
**Email: info@upubscience.com**
**Website: http://www.upubscience.com/**

# Table of Content

World Journal of Economics and Business Research
Print ISSN: 2960-0073
Online ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3059

# A HYBRID GA-PSO AND COUPLING COORDINATION FRAMEWORK FOR MULTI-OBJECTIVE OPTIMIZATION IN SUSTAINABLE TOURISM SYSTEMS

DingYi Ma[1*], KeYin Zhu[2], Lei Zang[1]
*[1]School of Materials Science and Engineering, Henan Polytechnic University, Jiaozuo 454003, Henan, China.*
*[2]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, Henan, China.*
*Corresponding Author: DingYi Ma, Email: madingyi1219@163.com*

**Abstract:** Rapid tourism growth has boosted local economies but also posed environmental and social challenges, highlighting the need for sustainable management. This study presents a hybrid optimization framework that integrates Genetic Algorithm with Particle Swarm Optimization (GA-PSO) with a Coupling Coordination Degree (CCD) mechanism to balance social, economic, and environmental subsystems. The model targets maximization of employment, tourism revenue, and glacier stability under statistical and coordination constraints. Regression analysis and the entropy weight–TOPSIS method were applied to quantify subsystem interactions. Results show that GA-PSO outperforms single GA and PSO, achieving an optimal coordination degree of 0.6254. Employment, tourism revenue (USD 147 million), and environmental sustainability all showed notable improvements. This research extends the use of hybrid evolutionary algorithms in sustainable tourism and provides quantitative support for policymakers seeking to balance economic growth, social well-being, and ecological resilience.
**Keywords:** Sustainable tourism; Multi-objective optimization; GA-PSO; Coupling coordination degree

## 1 INTRODUCTION

Juneau, Alaska, with a population of about 30,000, has faced mounting challenges in tourism carrying capacity due to the rapid growth of cruise tourism in recent years. On the busiest day in 2023, the city hosted seven cruise ships and more than 20,000 visitors. While this influx significantly bolstered the local tourism economy, it also generated negative externalities, including environmental pollution, traffic congestion, and pressure on local resources. These impacts have intensified community concerns over the sustainable management of tourism.

Research on tourism-related carbon emissions and resource regulation has developed along several lines. Previous studies developed an impact assessment model for cruise ship pollution and estimated the direct costs, demonstrating that environmental costs far outweighed the associated economic benefits[1,2]. However, these studies did not propose an effective dynamic coordination mechanism. Previous studies employed a coupling coordination degree model to examine the relationship between urban development and ecological civilization construction, but the analyses were limited to static evaluations and failed to provide dynamic cross-system optimization strategies[3,4]. Literature integrated the entropy weight-TOPSIS method with coupling coordination to assess the synergetic development of the economy, ecology, and tourism, revealing an overall upward trend but highlighting persistent regional disparities[5,6]. While these studies illuminated interrelationships across multiple dimensions, they did not account for more complex factors such as carbon emissions or social feedback mechanisms. Literature applied structural equation modeling to demonstrate that tourism participation significantly enhances residents' psychological and political empowerment, thereby reinforcing support for sustainable tourism[7]. Nonetheless, this work was largely confined to a social perspective and lacked integration with other dimensions. In contrast, studies proposed a variety of optimization algorithms—including GM-PSO, multi-objective optimization, PSO-based visitor distribution, and improved NSGA-II—that effectively improved efficiency and coordination in areas such as HPC workflow scheduling, tourism project funding allocation, balancing visitor experience with profitability, and managing the environmental carrying capacity of urban tourism[8-11]. Despite these advances, the methods remain restricted to single-dimensional optimization and have yet to realize comprehensive trade-offs across the socio-economic-environmental nexus.

Despite these advances, notable gaps remain in the field of sustainable tourism. First, research on multi-objective coordinated optimization tailored to complex tourism systems is still limited. Second, the systemic coupling among social feedback, environmental carrying capacity, and economic benefits has not been sufficiently examined, which constrains integrated coordination and policy-level decision-making.

To address these gaps, this study develops a three-subsystem multi-objective model that integrates the GA-PSO optimization algorithm with the Coupling Coordination Degree (CCD) mechanism. Focusing on the tourism context of Juneau, the model seeks to realize system optimization and dynamic scheduling across social, economic, and environmental subsystems, thereby providing both theoretical support and quantitative evidence for sustainable development.

## 2 METHOD

In the sustainable tourism development of Juneau, multiple factors such as visitor numbers, carbon emissions, and rainfall interact and jointly shape the coordinated development of social, economic, and environmental subsystems. To design more effective sustainability strategies, this study applies a multi-objective optimization approach, drawing on the Triple Bottom Line (TBL) framework and the United Nations 2030 Agenda for Sustainable Development. The objectives are classified into social, economic, and environmental dimensions, enabling a comprehensive assessment of interactions among subsystems and their impacts on economic growth, social well-being, and environmental protection. The ultimate aim is to maximize overall urban benefits while promoting coordinated development across the three subsystems.

### 2.1 Establishment of Multi-System Coordination Model Optimized by GA-PSO

#### *2.1.1 Establishment of objective function*

Against the backdrop of advancing global sustainable development goals, tourism in Juneau has become a key driver of local economic and social growth, exerting notable impacts across social, economic, and environmental dimensions. On the social side, tourism contributes to job creation, income growth, and overall economic expansion, thereby enhancing social well-being. However, the surge in visitor numbers has also intensified traffic congestion and placed pressure on public infrastructure. Optimizing the social subsystem therefore requires a balance between positive and negative effects. Economically, the core value of tourism lies in generating revenue. This process relies on increases in tourist arrivals, higher per capita spending, and synergistic growth within related industries such as hotel taxation, producing chain effects. Environmentally, large tourist inflows heighten ecological stress, particularly threatening glacier ecosystems. Rising carbon emissions and accelerated glacier melting may trigger irreversible environmental changes. Thus, achieving sustainable tourism development hinges on reconciling tourism growth with the protection of natural resources.

To comprehensively assess the multifaceted impacts of tourism on social, economic, and environmental dimensions, this study defines three optimization objectives: maximizing the annual average employment rate, maximizing tourism revenue, and maximizing the mean elevation of glaciers. Accordingly, multiple regression models are constructed to examine the influence of tourism activities on each dimension, which in turn serve as the basis for formulating the objective functions.

At the social dimension, to quantify the effects of tourism development on local residents' well-being, this study establishes an Ordinary Least Squares (OLS) model, with the annual average employment rate as the dependent variable and visitor numbers and residents' social satisfaction as independent variables:

$$f_1(m, m_1) = \alpha_0 + \alpha_1 m + \alpha_2 m_1 \tag{1}$$

At the social dimension, the annual average employment rate in Juneau is defined as the dependent variable $f_1(m, m_1)$, where $m$ denotes annual tourist arrivals and $m_1$ represents social satisfaction.

At the economic dimension, tourism revenue is used as the core indicator of economic benefits. A log-linear regression model is constructed to analyze the relationships among tourist arrivals, average per capita tourist expenditure, hotel tax revenue, and tourism revenue:

$$f_2(m, n_1, n_2) = \beta_0 + \beta_1 \cdot \log m + \beta_2 \cdot \log n_1 + \beta_3 \cdot \log n_2 \tag{2}$$

where $f_2(m, n_1, n_2)$ denotes tourism revenue, $m$ is the number of tourists, $n_1$ indicates per capita expenditure, and $n_2$ refers to hotel tax revenue.

At the environmental level, in order to explore the potential impacts of tourism activities on glacier ecosystems, the mean glacier elevation is taken as the dependent variable. A ridge regression model is constructed to capture the statistical relationships among tourist arrivals, annual precipitation, carbon emissions, and mean annual temperature:

$$f_3(m, L_1, T_1, E_1) = \gamma_0 + \gamma_1 \cdot L_1 + \gamma_2 \cdot m + \gamma_3 \cdot E_1 + \gamma_4 \cdot T_1 \tag{3}$$

where $f_3(m, L_1, T_1, E_1)$ represents mean glacier elevation, \$m\$ is the number of tourists, $L_1$ denotes carbon emissions, $T_1$ is mean annual temperature, and $E_1$ represents precipitation. The model achieves a coefficient of determination $R^2 = 0.865$, indicating a good fit.

Through these regression models, the impacts of tourism on the economic, social, and environmental subsystems can be effectively interpreted. On this basis, the optimization objectives for sustainable tourism development are defined as maximizing employment, tourism revenue, and glacier elevation. The multi-objective optimization function is therefore expressed as:

$$\max S = \max f_1(m, m_1) + \max f_2(m, n_1, n_2) + \max f_3(m, L_1, T_1, E_1) \tag{4}$$

#### *2.1.2 Determination of constraint conditions*

To ensure the feasibility and adaptability of the model in promoting sustainable development in Juneau, two constraints are introduced. The first is a statistical constraint, aimed at securing the robustness and reliability of the estimated relationships. Specifically, each decision variable is restricted to the 95% confidence interval implied by the ordinary least squares (OLS) regression results:

$$x_i^L \le x_i \le x_i^U, i = 1,2, \ldots, n. \tag{5}$$

Here, $x_i^L$ and $x_i^U$ denote the lower and upper bounds of variable $x_i$, obtained from the 95% confidence interval of the OLS regression. This interval reflects parameter uncertainty, accounts for model uncertainty, and ensures that variables remain within empirically grounded and statistically supported ranges.

The second is a coupling coordination constraint, introduced to address the interdependencies among the three subsystems in the multi-objective optimization process. To quantify and regulate their coordination, the coupling coordination degree (CCD) is employed as a constraint indicator:

$$D \geq D_0 \tag{6}$$

In this expression, $D$ denotes the coupling coordination degree, and $D_0$ is a preset minimum threshold that guarantees a sufficient level of coordinated development across the social, economic, and environmental subsystems. This constraint ensures that no single subsystem is excessively prioritized at the expense of overall system performance.

***2.1.3 Comprehensive evaluation by entropy Weight-TOPSIS method***

To objectively evaluate the composite score of each subsystem, this study adopts the entropy-weight method to determine the weights of indicators. By calculating the information entropy of each indicator, the method reflects the uncertainty and variability of indicators, and thereby derives their weights. The specific steps are as follows:

1. Construct the probability matrix by dividing each standardized value by the column sum:

$$p_{ij} = \frac{Z_{ij}}{\sum_{i=1}^{n} Z_{ij}} \tag{7}$$

where $Z_{ij}$ is the standardized data, and $p_{ij}$ is the probability of sample $i$ under indicator $j$.

2. Compute the entropy of each indicator:

$$E_j = -k \sum_{i=1}^{n} p_{ij} \ln\left(p_{ij}\right), \tag{8}$$

$$k = \frac{1}{\ln(n)}, \tag{9}$$

where $k$ is the normalization constant, $p_{ij}$ are the elements of the standardized matrix, and $n$ is the number of samples.

3. Calculate the information utility value of each indicator:

$$d_j = 1 - E_j \tag{10}$$

4. Determine the weight of each indicator based on its information utility:

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j}, \tag{11}$$

where $d_j$ is the information utility value, \$m\$ is the number of indicators, and $w_j$ is the weight of indicator $j$. The results of indicator weights are shown in Table 1.

**Table 1** Weights of Each Indicator

| Target Level | Targets | Causality | Weights |
|---|---|---|---|
| Society | Employment Rate | + | 25.16% |
| | Satisfaction | + | 33.15% |
| | Tourist Arrivals | + | 41.69% |
| Economy | Hotel tax | + | 26.5% |
| | Tourist Revenue | + | 12.51% |
| | Per Capita Expenditure | + | 50.17% |
| | Tourists arrivals | – | 10.83% |
| Environment | Glacier Average Elevation | + | 28.52% |
| | Annual Precipitation | + | 13.94% |
| | Tourists Arrivals | – | 20.56% |
| | Annual Average Temperature | – | 24.23% |
| | Carbon Emissions | – | 12.75% |

5. Then TOPSIS is used to comprehensively score each subsystem, and the positive and negative ideal solutions are obtained:

$$A^{+} = \{max\left(v_{ij}\right)\}, A^{-} = \{min\left(v_{ij}\right)\} \tag{12}$$

Where $v_{ij}$ is the element of the weighted normalized matrix.

6. Calculate the distance between the evaluation object and the positive and negative ideal solution:

$$D_i^{+} = \sqrt{\sum_{j=1}^{m}\left(v_{ij} - A_j^{+}\right)^2}, D_i^{-} = \sqrt{\sum_{j=1}^{m}\left(v_{ij} - A_j^{-}\right)^2} \tag{13}$$

7. Calculate the comprehensive score of each subsystem:

$$C_i = \frac{D_i^{-}}{D_i^{+} + D_i^{-}} \tag{14}$$

***2.1.4 Determination of the coupling coordination degree***

Firstly, based on the comprehensive scores $S_1$,, $S_2$, and $S_3$ of each subsystem obtained by the entropy weight method, the coupling degree between the three is calculated:

$$C = \frac{S_1 \cdot S_2 \cdot S_3}{(S_1 + S_2 + S_3)^3} \tag{15}$$

Secondly, construct the coordination index:

$$T = \epsilon_1 S_1 + \epsilon_2 S_2 + \epsilon_3 S_3 \tag{16}$$

where $\epsilon_1$,$\epsilon_2$,$\epsilon_3$ are the weights of each subsystem. This study assumes that the three weights are equal, that is :

$$\epsilon_1 = \epsilon_2 = \epsilon_3 = \frac{1}{3} \tag{17}$$

Finally, the coupling coordination degree is obtained:

$$D = \sqrt{C \cdot T} \tag{18}$$

In order to ensure the balanced development between the various dimensions, we introduce the following constraints in the optimization model:

$$D \geq D_0 \tag{19}$$

among them, $D_0$ represents the minimum acceptable level of coordination, which in this study is set at $D_0 = 0.6$. The threshold setting and the grading criteria for the coupling coordination degree are based on previous study and further adjusted according to the characteristics of the sample used here[12]. In order to facilitate comparison and interpretation, the coupling coordination degree D is categorized into six levels, as shown in Table 2.

**Table 2** Classification of Coupling Coordination Degree

| Degree of Coordination | Development Type | Degree of Coordination | Development Type |
|---|---|---|---|
| $0 \leq D \leq 0.2$ | Severe Imbalance | $0.5 < D \leq 0.6$ | Primary Coordination |
| $0.2 < D \leq 0.4$ | Moderate Imbalance | $0.6 < D \leq 0.8$ | Secondary Coordination |
| $0.4 < D \leq 0.5$ | Slight Imbalance | $0.8 < D \leq 1.0$ | Excellent Coordination |

## 3 MODEL OPTIMIZATION AND SOLUTION

### 3.1 Genetic Algorithm-Particle Swarm Optimization (GA-PSO)

#### *3.1.1 Genetic algorithm*

The genetic algorithm is a global optimization technique that mimics the evolutionary processes observed in nature. Its fundamental operations include selection, crossover, and mutation. The key steps are as follows.

1. Selection: Individuals with higher fitness are selected for reproduction based on the fitness function $f(x)$. The fitness function is defined as:

$$P(x) = \frac{f(x)}{\sum_{i=1}^{N} f(x_i)} \tag{20}$$

where, $P(x)$ represents the selection probability, which is based on the individual's fitness, and $N$ denotes the population size.

2. Crossover: New individuals are generated through the crossover operation, with the crossover formula given as follows:

$$x_{\mathbf{new}} = \alpha x_1 + (1 - \alpha) x_2 \tag{21}$$

where, $\alpha$ denotes the crossover weight, while $x_1$ and $x_2$ represent the two parent individuals.

3. Mutation: With a certain probability, individuals undergo mutation, defined by the following formula:

$$x_{new} = x_{old} + \delta \tag{22}$$

where, $\delta$ denotes the mutation amplitude.

#### *3.1.2 Particle swarm optimization*

The particle swarm optimization algorithm is an optimization method rooted in swarm intelligence, inspired by the foraging behavior of bird flocks. By facilitating information exchange among particles, the algorithm enhances the convergence speed of solutions. Its fundamental steps are as follows:

1. Particle velocity update:

$$v_i = w \cdot v_i + c_1 \cdot r_1 \cdot (pbest_i - x_i) + c_2 \cdot r_2 \cdot (gbest - x_i) \tag{23}$$

2. Particle position update:

$$x_i = x_i + v_i \tag{24}$$

where, $v_i$ denotes the velocity of particle, $x_i$ represents its position, $pbest_i$ refers to the individual best solution, and $gbest$ indicates the global best solution. The parameter $w$ is the inertia weight, while $c_1$ and $c_2$ are acceleration coefficients.

The GA-PSO algorithm integrates the global search capability of the genetic algorithm with the information-sharing mechanism of particle swarm optimization, thereby balancing global and local exploration, accelerating convergence, and enhancing optimization performance. The research process is illustrated in Figure 1.

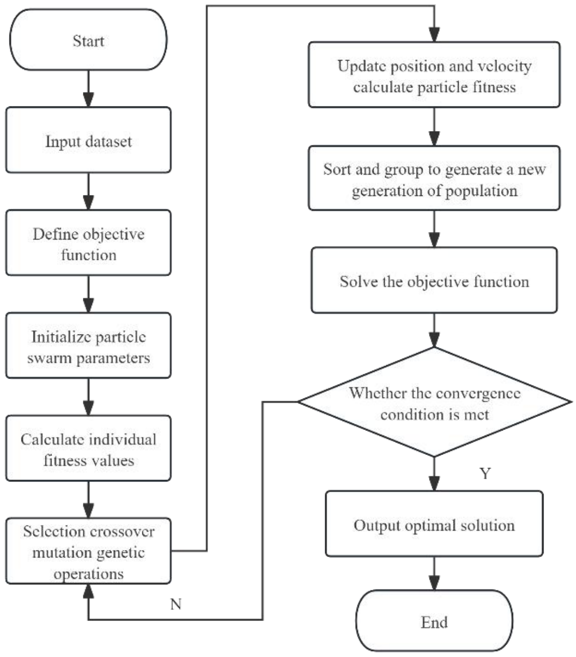


**Figure 1** Flow Chart of GA-PSO Algorithm

### 3.2 Tourism Optimization Strategy based on the GA-PSO Algorithm

After completing the fitting of the social, economic, and environmental subsystems, this study derives the corresponding fitting equations for each dimension.

Social dimension:

$$f_1(m, m_1) = 0.5063 - 2.82 \times 10^{-8} \cdot m - 0.0002 \cdot m_1 \tag{25}$$

Economic dimension:

$$f_2(m, n_1, n_2) = 16.7610 + 7.5152 \times 10^{-2} \cdot log\, m - 1.9589 \times 10^{-3} \cdot log\, n_1 + 2.8657 \times 10^{-7} \cdot log\, n_2 \tag{26}$$

Environmental dimension:

$$f_3(m, L_1, T_1, E_1) = 1400.0 + (-0.0055) \cdot L_1 + 0.0032 \cdot m + 0.02 \cdot E_1 + 1.5 \cdot T_1 \tag{27}$$

Building on this foundation, a multi-objective optimization function was developed to support the sustainable development of tourism in Juneau City, as presented below:

$$max\, S = max\, f_1\,(m, m_1) + max\, f_2\,(m, n_1, n_2) + max\, f_3\,(m, L_1, T_1, E_1) \tag{28}$$

$$s.t. \begin{cases} x_i^L \le x_i \le x_i^U, i = 1,2, \ldots, n \\ D \ge D_0 \end{cases} \tag{29}$$

Due to the nonlinear nature of the objective function, the complexity of the constraints, and the conflicting characteristics among the objectives, this study employs a hybrid optimization approach that integrates genetic algorithm with particle swarm optimization (GA-PSO). The performance of this hybrid method is compared with that of the individual genetic algorithm and particle swarm optimization algorithms. The corresponding convergence trajectories are presented in Figure 2.

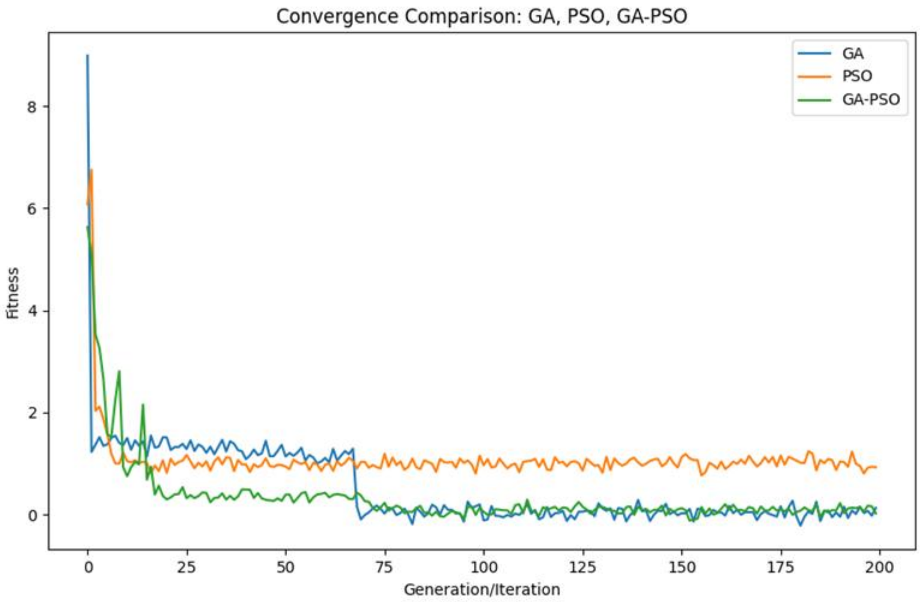


**Figure 2** Experimental Comparison Chart of Algorithm Optimization Effect

As illustrated by the convergence curve in Figure 2, the three algorithms show distinct differences in their optimization processes. GA initially converges slowly, with significant fluctuations in the early iterations. Although it gradually reduces the fitness value, its overall stability is limited. PSO, on the other hand, shows faster convergence in the early stages but stabilizes after approximately 30 iterations, with the fitness value remaining at a relatively high level. This suggests that PSO tends to become trapped in a local optimum in later iterations, limiting further improvement. In contrast, the GA-PSO hybrid demonstrates a similar fast descent in the initial phase. However, by leveraging the crossover and mutation mechanisms of the genetic algorithm, it maintains strong global search capabilities during the middle and later stages, effectively avoiding premature convergence. As a result, GA-PSO achieves a superior solution with fewer iterations, demonstrating greater convergence accuracy and stability in the later stages.

After applying the GA-PSO algorithm, each decision variable converged to its optimal solution, as presented in Table 3.

**Table 3** Solution Results of Decision Variables

| Decision Dimension | Decision Variable | Optimal Solution |
|---|---|---|
| Society | Employment Rate | 0.5663 |
| | Social Satisfaction | 34.45% |
| | Number of Tourists | 2,371,370 |
| Economy | Total Tourism Revenue | 147,024,940 USD |
| | Per Capita Consumption | 62.58 USD |
| | Hotel Tax | 4,290,392 USD |
| Environment | Air Quality Index | 84.64 |
| | Annual average elevation | 1,456.78 m |
| | Precipitation | 62.38 mm |
| | Carbon Emissions | 35,860 tons |
| | Average Annual Temperature | 44.91° F |

The results show that this solution achieves an optimal balance across the social, economic, and environmental dimensions, with a coordination equilibrium of 0.6254. Notably, both the employment rate and tourism revenue are high, while environmental indicators fall within acceptable ranges. This suggests that the GA-PSO algorithm effectively balances economic development, social benefits, and ecological protection. Based on these findings, we offer the following recommendations.

Economic Aspects: Juneau City should focus on developing the high-end tourism market by offering premium eco-tourism, cultural experiences, and customized travel services. This approach will help increase visitor spending, as well as boost hotel tax revenues and other related taxes.

Environmental Aspects: The city should enforce strict carbon emission controls while leveraging climate advantages, such as temperature, to promote low-carbon tourism initiatives. It is also important to develop green transportation options, encourage the use of renewable energy, and implement effective management of tourism-related carbon emissions.

Social Aspects: To improve overall urban livability, Juneau City should enhance infrastructure, optimize transportation networks, and improve public service facilities. Strengthening feedback mechanisms to address the needs of both tourists and residents will also be essential.

The GA-PSO hybrid algorithm effectively addresses the complex optimization challenges in tourism systems, which involve multiple objectives and constraints. Its global optimization capabilities and efficient constraint handling offer Juneau City a viable pathway toward sustainable development across tourism, the economy, society, and the environment.

## 4 CONCLUSION

In conclusion, this study proposes a GA-PSO-based multi-objective optimization framework that integrates social, economic, and environmental dimensions, offering a coordinated pathway for sustainable tourism development in Juneau City. The findings demonstrate that the hybrid approach achieves superior balance among employment, revenue, coordination, and ecological thresholds compared with single algorithms, thereby providing both theoretical insights into multi-system coupling and practical guidance for policy formulation.

Nevertheless, the model has limitations in terms of parameter generalizability and the exclusion of dynamic feedback mechanisms. Future research should expand the application of hybrid optimization to diverse tourism contexts, incorporate carbon emissions and social feedback into the framework, and explore real-time adaptive algorithms to enhance robustness and scalability.

## COMPETING INTERESTS

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

## REFERENCES

[1] Carić H. Challenges and prospects of valuation–cruise ship pollution case. Journal of cleaner production. 2016, 111: 487-498.

[2] Klein R A. Responsible cruise tourism: Issues of cruise tourism and sustainability. Journal of Hospitality and Tourism Management. 2011, 18(1): 107-116.

[3] Hu X, Dong C, Wang Y. Coupled and coordinated analysis of urban green development and ecological civilization construction in the Yangtze River Delta region. Sustainability. 2023, 15(7): 5955.

[4] Yin Z, Tang Y, Liu H, et al. Coupling coordination relationship between tourism economy-social welfare-ecological environment: Empirical analysis of Western Area, China. Ecological Indicators. 2023, 155: 110938.

[5] Tang C, Zeng R, Han Y. Coupling coordination evaluation of economic development, ecological environment, and the tourism industry in ski resorts. Journal of Cleaner Production. 2023, 426: 139076.

[6] Luo Z, Xu A, Gao S. Research on the coupling coordination relationship between tourism carrying capacity and the high-speed railway network: a case study in China. IEEE Access. 2023, 11: 20426-20444.

[7] Gautam V, Bhalla S. Exploring the relationships among tourism involvement, residents' empowerment, quality of life and their support for sustainable tourism development. Journal of Cleaner Production. 2024, 434: 139770.

[8] Hafsi H, Gharsellaoui H, Bouamama S. Genetically-modified multi-objective particle swarm optimization approach for high-performance computing workflow scheduling. Applied soft computing. 2022, 122: 108791.

[9] Arbolino R, Boffardi R, De Simone L, et al. Multi-objective optimization technique: A novel approach in tourism sustainability planning. Journal of Environmental Management. 2021, 285: 112016.

[10] Su F, Duan C, Wang R. Optimization model and algorithm design for rural leisure tourism passenger flow scheduling. Ieee Access. 2020, 8: 125295-125305.

[11] Tan Q. Research on sustainable carrying capacity of urban tourism environment based on multi objective optimization algorithm. Results in Engineering. 2023, 19: 101344.

[12] Shi T. Coupling coordination degree and spatial network effect of ecological protection and high-quality economic development in the Yellow River Basin. German Economic Review. 2020, 3: 25-34.

World Journal of Economics and Business Research
Print ISSN: 2960-0073
Online ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3060

# EMPIRICAL ANALYSIS ON THE EFFECT AND MECHANISM OF GREEN FINANCE RELATING TO REDUCTION AND CARBON EMISSION REDUCTION RELYING ON THE EMPIRICAL EVIDENCE FROM CHINESE LISTED COMPANIES

ZiMo Chen
*School of Accounting, Shanghai University of International Business and Economics, Shanghai 201620, China.*
*Corresponding Email: 23043019@suibe.edu.cn*

**Abstract:** In the process of implementing the "dual-carbon" strategy, research on whether green finance can truly translate into emission reduction effects and its conversion mechanism is crucial for testing green finance policies. The present study takes A-share listed non-financial enterprises in China from 2011 to 2022 as samples, matches enterprise emission data of sulfur dioxide ($SO_2$), nitrogen oxides ($NO_x$), and soot, and constructs a "firm-year" green finance index built on seven types of tools including green credit and green bonds to systematically examine the impact of green finance on corporate pollution reduction and carbon emission reduction. Baseline regression results show that for each 1-unit rise in the green finance index, the logarithm of total corporate pollutant emissions decreases by an average of 0.366, and this conclusion stays robust after instrumental variable and robustness examinations. Mechanism analysis reveals that green finance exerts its effects through a dual-pathway of alleviating financing constraints and promoting green technological innovation: on one hand, it significantly reduces the SA index, providing low-cost funds for corporate green investments; on the other hand, it significantly increases the quantity of applications submitted for green utility model patents, facilitating process upgrading and end-of-pipe treatment. Heterogeneity tests reveal that the emission reduction effect of green finance stands out more in non-state-owned enterprises (non-SOEs), eastern regions, and the manufacturing industry, presenting a pattern of "non-SOEs outperforming SOEs, eastern regions outperforming central and western regions, and manufacturing outperforming non-manufacturing". This study provides micro-level evidence for constructing a precise and efficient green finance support system, and also offers actionable policy references for accomplishing the "dual-carbon" goals.
**Keywords:** Green finance; Pollution control and carbon emission reduction; Funding constraints; Green tech innovation; Corporate heterogeneity

## 1 INTRODUCTION

Escalating greenhouse gas emissions and the resultant climate crisis—manifesting itself by extreme weather events, rising sea levels, and resource scarcity—highlight the urgent need for transformative solutions to ensure sustainable development. Against the backdrop of in-depth changes in the global governance system and accelerated advancement of the "dual-carbon" strategy, China, as the world's top carbon emitter, is facing the major era proposition of coordinated governance pertaining to pollution reduction and carbon emission reduction. Green finance, via market-oriented tools such as credit facilities and funds, serves as an important means to promote structural transformation after China's economic development stepped into the "new normal", and is regarded as a key fulcrum to drive enterprises' pollution control and carbon mitigation [1]. As specified in the Official Report of the 19th National Congress of China, green finance is an inevitable path to promote green economic transformation, a key measure to address unbalanced and inadequate regional economic development, and an important impetus for high-quality economic development; thus, further development of green finance should be promoted [2].

Existing scholars have carried out a series of studies on the effect of green finance on corporate production and operation activities. Studies shows that green finance promotes the expansion of green total factor productivity by improving the efficiency in the allocation of financial resources [3]. At the same time, society's demand for pollution control and carbon mitigation is increasing. With the continuous improvement of green GDP assessment methods and the implementation of strict environmental regulations, local governments attach greater importance to the ecological benefits of economic growth. Therefore, promoting corporate pollution reduction and carbon emission reduction is an inevitable requirement for implementing the "dual-carbon" strategy and a practical need to respond to the social concern of pollution reduction and carbon emission reduction.

In practice, when the focus of green finance policies shifts from "credit scale" to "corporate pollution and carbon reduction levels", whether the growth of green finance development can promote the advancement of corporate pollution and carbon emissions reduction levels has become the first criterion for measuring the effectiveness of the "dual-carbon" strategy. The core of this issue lies in the transmission mechanism of green finance on the coordination of corporate pollution reduction and carbon emission reduction.

### 1.1 Literature Review

Relevant studies on this research topic can be summarized into two levels:
Macro-level studies: Existing research measures green finance using green finance indexes and tests its impact on the abatement of pollution and carbon.For example,using panel data of 30 provinces in China spanning 2011 to 2022, Wang et al. introduced green finance indexes as mechanism variables and found that the combination of digital and real economies directly promotes the synergistic function of pollutant abatement and carbon emission reduction . In addition, some scholars have taken green finance reform pilot zones as research objects to evaluate their impact. Among them, Zhang & Hu constructed a general equilibrium model to systematically explain the theoretical association between green finance and pollution/carbon reduction, and used panel data of 281 Chinese cities from 2011 to 2022 to empirically examine the impact along with the working mechanism of green finance reform pilot zone policies concerning urban pollution and carbon emission reduction using a staggered difference-in-differences (DID) model [3]. Based on the current development status of China's pilot zones for green finance reform, Wang et al. proposed that local governments should strengthen the establishment of guarantee mechanisms and innovate green financial products based on local economic advantages to better promote green finance practice[4].
Micro-level studies: Researchers have probed into the effect of green finance on corporate behavior. In terms of factor productivity, green finance significantly facilitates the enhancement of corporate factor productivity. For example, Liu et al. undertaken a quasi-natural experiment drawing on the Green Credit Guidelines and found that green credit significantly improves the factor productivity of high-pollution enterprises[5].Regarding financing constraints,green finance can drive corporate pollution reduction and carbon reduction by reducing financing constraints. Using the MLF collateral expansion event, Yan et al. confirmed that relevant policies ultimately reduce corporate emissions by lowering financing costs and improving environmental information disclosure quality, with more significant effects in enterprises with high financing constraints and in the growth stage [6]. Regarding the "dual-carbon" goal, enterprises need to promote digital transformation, and the operation of carbon emission trading mechanisms has an important impact on corporate production, strategy, and technological innovation. Relying on data of listed firms in China's A - share market from 2007 to 2020, Liu et al. took the carbon emission trading policy regarded as a quasi-experiment and found that this policy promotes the digital transformation of enterprises in pilot areas, thereby better integrating the concept of green development with digital economy theory [7].

### 1.2 Research Gaps and Marginal Contributions

Although the above literature confirms that green finance promotes enterprise pollution abatement and carbon emission reduction from multiple dimensions, there are still two gaps: First, existing studies mainly measure the dependent variable (y) using city or industry averages, failing to fully verify the micro-level direct influence of the "green finance development level(x) → corporate pollution-carbon reduction coordination (y)", which makes it impossible to reliably verify policy effects. Second, although some studies have proposed two potential mediating mechanisms—"alleviation of financing constraints" and "green technological innovation"—most studies only focus on the city level or conduct single-mechanism tests, lacking the identification and comparison of the two pathways in the same micro sample, resulting in more macro-level studies and less micro-level evidence.
To fill these research voids, this study first analyzes the direct impact of green finance on corporate pollution reduction and carbon emission cut, and then examines the channels through which green finance drives corporate pollution and carbon reduction by alleviating financing constraints and boosting technological innovation—this constitutes the marginal contribution of this study.

## 2 THEORETICAL DISCUSSION AND RESEARCH PROPOSITIONS

### 2.1 Direct Effect of Green Finance on Corporate Pollution-Carbon Reduction Coordination

Green finance policies can guide capital factors to flow into the sector of energy conservation and environmental protection, promote regional industrial transition and upgrading by encouraging the advancement of clean industries and restricting high-polluting enterprises [8], and advance the economy's green-oriented and low-carbon transformation [9]. At the same time, through diversified types of financial tools such as green credit facilities and green bonds, green finance reallocates scarce resources from "high-energy-consuming, high-polluting, and overcapacity" industries to green and low-carbon projects, thereby directly curbing corporate pollutant and carbon emissions and improving the standardization of enterprises' use of carbon emission rights, energy use rights, and pollution discharge rights. On this basis, the following hypothesis is put forward:
H1: Green finance significantly improves the level of corporate pollution-carbon reduction coordination; that is, the higher the intensity of green finance, the lower the pollutant and carbon emissions per unit output of enterprises.

### 2.2 Mechanism Routes Through Which Green Finance Boosts Corporate Pollution-Carbon Reduction Coordination

The synergetic function of green finance in corporate pollution and carbon reduction not only manifests as a direct effect but also exerts its action through the two mediating pathways that follow:

#### *2.2.1 Financing constraint alleviation mechanism*

Green finance forms an effective financing constraint mechanism through differentiated interest rate policies and green supervision measures, promoting corporate green transformation [3]. Specifically, at the economic performance level, enterprises can obtain financing through green finance, sending a signal to the government and the public that they actively fulfill green social responsibilities. This helps enterprises gain a good reputation, enhance the confidence and trust of stakeholders, attract more investors, reduce financing constraints, and ultimately improve corporate competitiveness and economic benefits [10].

Meanwhile, Chen took Chinese commercial banks and listed enterprises from 2014 to 2023 as research objects, constructed a bank fintech index and corporate sustainability indicators using machine learning algorithms, and found that fintech promotes corporate sustainability through financing and governance effects [11]. By selecting Chinese A-share listed companies that issued green bonds and those that issued ordinary bonds but not green bonds from 2016 to 2021, Ren further found that green bond issuance improves corporate ESG performance by alleviating financing constraints and reducing agency costs [12].

For green enterprises, green finance reduces the financing costs and thresholds of green projects through green credit quotas and green bond special channels, thereby alleviating financing constraints for their pollution and carbon reduction investments. In contrast, for polluting enterprises, green finance implements punitive monetary policies to strictly restrict the financing channels for their high-polluting projects. This differentiated financing constraint mechanism encourages green enterprises to update clean equipment, expand environmental protection investments, and achieve both end-of-pipe treatment and source prevention of pollution, thereby reducing pollutant and carbon emissions simultaneously and promoting enterprises to choose a green development path. Using this as a basis, the following hypothesis is put forward:

H2: Green finance indirectly improves the level of corporate pollution-carbon reduction coordination through the mitigation of financing constraints for enterprises.

***2.2.2 Green technological innovation promotion mechanism***

The ability of technological innovation is affected by many factors, among which R&D investment is the main driving factor. The capital flow of green finance encourages enterprises to develop green technologies, strengthen technological innovation, and create a competitive environment favorable for enhancing the level of green technological innovation. Green technological innovation also promotes the channeling of economic resources to green enterprises, optimizes the industrial structure, helps enterprises create new products and services, and drives high-quality economic development [13].

Using provincial panel data across China from 2018 to 2022, Ren constructed a green low-carbon development index and found through panel regression and mediating effect models that green finance exhibits a significant positive impact on green low-carbon economic development process, and indirectly empowers corporate green development by accelerating technological innovation and the upgrading of industrial structure[14]. From the dual viewpoints of technological innovation and environmental attention, Mao et al. utilized data from 30 provinces (municipalities, autonomous regions) in China from 2011 to 2022 and uncovered that green finance promotes the "quantity and quality improvement" of technological innovation, thereby driving the development of new-quality productive forces [15]. With the green finance reform and innovation policy as a reference, Zhang et al. constructed multi-period DID and DDD models and found that green finance significantly improves corporate total factor productivity using data of Chinese A-share listed companies from 2010 to 2021 [16]. In addition, Meng et al. utilized the entropy method to examine the mechanism of green finance accelerating high-quality economic development, from the dual dimensions of technological innovation and industrial structure upgrading, and their empirical results showed that technological innovation and industrial structure upgrading play key roles in green finance propelling high-quality economic development[17].

Green finance provides stable and low-cost capital for corporate green technology R&D through market-oriented means such as preferential interest rates and government subsidies, thereby increasing the enthusiasm and success rate of corporate technological innovation. With the support of green finance, enterprises significantly strengthen their green technological innovation capabilities, and further augment the synergistic effect of pollution and carbon reduction through process upgrading, equipment renewal, and product restructuring. Drawing upon this, the following hypothesis is put forward:

H3: Green finance indirectly improves the level of corporate pollution-carbon reduction coordination by promoting corporate green technological innovation.

## 3 MODEL SPECIFICATION AND VARIABLE SELECTION

Green finance curbs pollutant emissions and achieves corporate - level pollution reduction by optimizing financial resource allocation, reducing corporate financing costs, and encouraging green technological innovation. Meanwhile, pollution levels are also affected by internal corporate characteristics such as scale, profitability, growth, and age. Based on existing studies, this study constructs the following dynamic panel econometric model:

$$Pollutant_{i,t} = \alpha_0 + \alpha_1 Pollutant_{i,t-1} + \beta GreenFinance_{i,t} + \gamma \sum Control_{i,t} + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (1)$$

In this model, $Pollutant_{i,t}$ represents the pollutant discharge level of an enterprise in a certain year; $Pollutant_{i,t-1}$ is the lagged one - period pollutant emission of enterprise $i$; $GreenFinance_{i,t}$ is the green finance index, reflecting the degree of green financial resources obtained by enterprise $i$ (or its location) in year $t$; $Control_{i,t}$ stands for firm size (size) and equity multiplier (em) are included as control variables, firm age (firmage), and firm growth (growth); $\mu_i$ denotes firm fixed effects; $\lambda_t$ represents year fixed effects; $\alpha_0,\alpha_1,\beta,\gamma$ are estimated coefficients; and $\varepsilon_{i,t}$ is the idiosyncratic error term.

### 3.1 Variable Selection

***3.2.1 Dependent variable***

Pollutant Emission (pollution): Referring to common practices in existing studies on industrial pollutants, this study employs the natural logarithm of the total emissions of three major air pollutants ($SO_2$, $NO_x$, and soot) as the core dependent variable to reflect the total pollution emission intensity of enterprises.

***3.2.2 Core independent variable***

Green Finance Index (GreenFin): Green finance refers to financial services offered to sectors like environmental protection, energy conservation, and clean energy, aiming to optimize resource utilization efficiency, strengthen environmental governance, and steer resources away from high-polluting and high-energy-consuming industries toward clean industrial sectors that feature advanced technologies [2]. Internationally recognized definitions further clarify its connotation: the World Economic Forum defines it as "organized financial activity created to ensure a better environmental outcome," while UN ESCAP describes it as "support for environment-oriented technologies, projects, industries or businesses."[18]

Green finance in China is primarily composed of green credit, green securities, green insurance, and green investment. Based on the definition of green finance's multi-dimensional connotation in existing studies [2], this study constructs a green finance development index from 7 core dimensions—green credit, green investment, green insurance, green bonds, green support, green funds, and green equity—using the entropy method to measure the regional green finance development level at the "firm-year" level. A higher green finance index indicates a higher level of regional green finance. Considering the lag effect of green finance, this study uses the lagged one-period green finance index to more accurately verify its impact on corporate the mitigation of pollution and carbon.

***3.2.3 Control variables***

To manage other factors that may affect corporate pollutant emissions, this study selects five indicators as control variables, grouped into three categories:

1. Scale and Capital Structure: firm size (size), equity multiplier (em);
2. Profitability and Growth: return on assets (roa), firm growth (growth);
3. Life Cycle: firm age (firmage).
4. 3.2.4 Mechanism Variables

To comprehensively analyze the supportive role of green finance in promoting corporate pollution and carbon reduction, it is necessary to systematically test the internal mechanism of green finance on corporate pollution and carbon reduction:

1. Financing Constraint: Measured by the SA Index (Hadlock & Pierce SA Index), where a smaller value indicates weaker financing constraints (and thus a higher magnitude of pollution control and carbon reduction);
2. Green Technological Innovation: Measured by the natural logarithm of the total number of green utility model patent applications of enterprises in the current year.

The specific definitions of all variables are shown in Table 1.

**Table 1** Operational Definition and Description of Variables

| Variable Name | Variable Symbol | Variable Definition |
|---|---|---|
| Pollution Emission | pollution | Natural logarithm of total emissions of air pollutants ($SO_2$ + $NO_x$ + soot) |
| Green Finance Index | GreenFin | Comprehensive green finance index synthesized by the entropy method (one-period lagged version) |
| Firm Size | size | Natural logarithm of total assets at the end of the year |
| Equity Multiplier | em | Total Assets at the End of the Year / Owner's Equity at the End of the Year |
| Return on Assets | roa | Net profit / Total assets at the fiscal year-end |
| Firm Age | firmage | ln (Current year - Founding year + 1) |
| Firm Growth | growth | [Operating income (current year) - Operating income (previous year)] / Operating income (previous year) |
| Ownership Nature | soe | 1 = State-owned enterprise (SOE), 0 = Non-SOE |
| Eastern Region | east | 1 = Enterprise located in eastern China, 0 = Otherwise |
| Western Region | west | 1 = Enterprise located in western China, 0 = Otherwise |
| Central Region | mid | 1 = Enterprise located in central China, 0 = Otherwise |

| | | |
|---|---|---|
| Manufacturing Industry | manufacturing | 1 = Enterprise in manufacturing industry, 0 = Otherwise |
| Financing Constraint | SA Index | Hadlock & Pierce SA Index (smaller value = weaker constraint) |
| Green Innovation | ln_green_patent | Natural logarithm of the number of green utility model patents applied for in the current year |
| Year Fixed Effects | year_fe | Year dummy variable |

## 4 RESULTS

### 4.1 Descriptive Statistics

Table 2 reports the descriptive statistics of the primary variables. For pollutant emission (pollution), the mean value is 1.203, the population standard deviation is 1.066, the minimum figure is 0, and the maximum figure is 6.227, indicating significant differences in statistics in pollutant emissions among enterprises. The green finance index (GreenFin) has a mean value of 0.390, a standard deviation of 0.123, and a value range of 0.013-0.719, showing that the support of green finance for corporate pollution and carbon reduction is generally moderate and relatively concentrated. The mean value of firm size (size) is 22.526 with a standard deviation of 1.411; similarly, other variables also show varying degrees of differences. Except for the equity multiplier (em) and firm growth (growth), the standard deviations across all variables are less than 1.5, indicating a low degree of data dispersion.

**Table 2** Descriptive Statistics for Key Variables

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| pollution | 37174 | 1.203 | 1.066 | 0 | 6.227 |
| GreenFin | 18640 | 0.390 | 0.123 | 0.013 | 0.719 |
| size | 18640 | 22.526 | 1.411 | 17.641 | 30.081 |
| em | 18640 | 2.502 | 9.292 | -865.898 | 417.253 |
| roa | 18640 | 0.020 | 0.147 | -14.302 | 0.786 |
| firmage | 18640 | 3.022 | 0.317 | 1.099 | 4.290 |
| growth | 18622 | 1.029 | 24.526 | -48.417 | 2354.549 |

### 4.2 Baseline Regression Findings

Table 3 presents the baseline regression results of the green finance index on corporate pollutant emissions (pollution). The regression controls for firm size (size), equity multiplier (em), return on assets (roa), firm age (firmage), firm growth (growth), and year fixed effects.

The coefficient of the core independent variable (GreenFin) is -0.366, which is significant at the 5% level. This indicates that for each 1-unit increase in the green finance index, corporate pollutant emissions decrease by an average of 0.366 units, verifying the inhibitory effect of green finance on pollution mitigation and verifying Hypothesis H1.

Regarding control variables: The coefficient of firm size (size) is 0.066, significant at the 1% level, indicating that larger enterprises have higher pollution emission levels; the coefficients of equity multiplier (em), return on assets (roa), and firm age (firmage) are not significant, suggesting that their marginal impact on pollutant emissions is limited under the baseline setting.

**Table 3** Baseline Regression Results (Dependent Variable: pollution)

| Variable | Coefficient | Std. Error |
|---|---|---|
| GreenFin | -0.366** | 0.183 |
| size | 0.066*** | 0.012 |

| | | |
|---|---|---|
| em | 0.001 | 0.001 |
| roa | 0.055 | 0.036 |
| firmage | -0.093 | 0.102 |
| growth | 0.000 | 0.000 |
| Constant | -0.381 | 0.373 |
| Observations | 15481 | — |
| R-squared | 0.429 | — |
| Firm FE | YES | — |
| Year FE | YES | — |

*Note: *** p<0.01, ** p<0.05, * p<0.1; Standard errors are reported in parentheses.

### 4.3 Mechanism Test Results

***4.3.1 Financing constraint alleviation mechanism***

Table 4's Column (1) shows that the estimated coefficient of the green finance index is -0.029 ($p < 0.1$), which demonstrates statistical significance negative, indicating the fact that green finance significantly reduces the corporate SA index—i.e., financing constraints are alleviated. This result is consistent with the "capital threshold" logic in green transformation: under traditional credit frameworks, green projects struggle to secure sufficient funds due to long payback periods and high risk premiums. However, green finance tools (e.g.green credit windows, discounted reloans) expand financial institutions' service scope, enabling banks to lend to green enterprises with lower risk weights and enterprises to obtain low-cost, long-term funds via green certification—shifting capital from "passive emission reduction" to "active pollution control". Thus, Hypothesis H2 is verified.

***4.3.2 Green innovation-driven mechanism***

Column (2) of Table 4 (ln_green_patent) shows that the coefficient of the green finance index is 0.453 ($p < 0.1$), and this is statistically significantly positive, indicating that green finance stimulates corporate green technological innovation. This is because green finance reduces R&D costs via subsidies and risk compensation, and boosts innovation returns through "green technology certification-credit quota linkage". With financing constraints eased, enterprises can afford R&D trial-and-error costs, increasing investment in clean processes and end-of-pipe equipment. Supporting policies (e.g., green patent rewards) further amplify innovation incentives. Thus, Hypothesis H3 is verified.

**Table 4** Mechanism Test Results

| Variable | (1) SA Index (Financing Constraint) | (2) ln_green_patent (Green Innovation) |
|---|---|---|
| GreenFin | -0.029* | 0.453* |
| | (0.016) | (0.250) |
| Constant | -3.703*** | 5.533*** |
| | (0.035) | (0.495) |
| Observations | 18622 | 17113 |
| R-squared | 0.808 | 0.212 |
| Controls | YES | YES |
| Firm FE | YES | YES |

| | | |
|---|---|---|
| Year FE | YES | YES |

*Note: *** p<0.01, ** p<0.05, * p<0.1; Standard errors are reported in parentheses.

### 4.4 Heterogeneity Analysis Results

#### *4.4.1 Ownership heterogeneity*

The first and second columns of Table 5 show that the green finance index has no significant impact on SOEs' pollutant emissions (coefficient = -0.065, $p > 0.1$) but significantly reduces non-SOEs' emissions (coefficient = -0.458, $p < 0.05$). This difference arises because non-SOEs face stricter financing constraints and are more sensitive to green finance's cost advantages, while SOEs have soft budget constraints and implicit government guarantees, reducing their sensitivity to green finance signals[19].

#### *4.4.2 Regional heterogeneity*

Column (3) within Table 5 shows that the green finance index inhibits pollutant emissions in eastern China (coefficient = -0.571, $p < 0.05$) but has no significant effect in central and western regions. Eastern China's mature financial markets, abundant green tools, and sound information disclosure enable enterprises to quickly convert green funds into emission reduction investments. In contrast, central and western regions lack financial infrastructure and policy implementation capacity, blocking policy transmission. Additionally, eastern China benefits from overlapping environmental policies (e.g., carbon trading, pollution permits), amplifying green finance's emission reduction effects.

#### *4.4.3 Industry heterogeneity*

Columns (6)-(7) of Table 5 show that green finance demonstrates a more significant emission reduction effect in manufacturing (coefficient = -0.654, $p < 0.01$) than in non-manufacturing (coefficient = -0.341, $p > 0.1$). Manufacturing enterprises are major polluters with strong green investment demand—green finance directly supports their process upgrading and equipment renewal. Non-manufacturing enterprises have lower emission intensity and scattered green investment needs, diluting green finance's incentive effects.

**Table 5** Heterogeneity Analysis Results (Dependent Variable: pollution)

| Variable | (1) SOEs | (2) Non-SOEs | (3) Eastern China | (4) Central China | (5) Western China | (6) Manufacturing | (7) Non-Manufacturing |
|---|---|---|---|---|---|---|---|
| GreenFin | -0.065 | -0.458** | -0.571** | -0.434 | -0.065 | -0.654*** | -0.341 |
| | (0.316) | (0.229) | (0.276) | (0.506) | (0.316) | (0.237) | (0.297) |
| Constant | 0.371 | -1.170** | -0.397 | -0.326 | -0.371 | -0.356 | -1.237* |
| | (0.751) | (0.461) | (0.461) | (1.055) | (0.751) | (0.461) | (0.721) |
| Observations | 6229 | 8880 | 9377 | 2588 | 6229 | 9833 | 5648 |
| R-squared | 0.479 | 0.375 | 0.406 | 0.451 | 0.479 | 0.444 | 0.424 |
| Controls | YES | YES | YES | YES | YES | YES | YES |
| Firm FE | YES | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES |

*Note: *** p<0.01, ** p<0.05, * p<0.1; Standard errors are reported in parentheses; "Controls refers to firm size (size), equity multiplier (em), return on assets (roa), firm age (firmage), and firm growth (growth)."

## 5 CONCLUSIONS

Green bonds and impact investing have become indispensable drivers for promoting sustainable finance. By 2024, they had attracted over USD 800 billion in funding annually, fully demonstrating that capital can not only generate competitive financial returns but also create significant environmental value.[20]This study takes Chinese A-share non-financial listed companies spanning 2011 to 2022 as samples, constructs a "firm-year" entropy method-constructed green finance index, and measures corporate pollution abatement and carbon emission reduction levels by the logarithm of total emissions of $SO_2$, $NO_x$, and soot. The primary research findings are as follows:

First, green finance shows a significant direct inhibitory effect on corporate pollutant emissions. After controlling for firm characteristics and year fixed effects, each 1-unit increase in the lagged green finance index reduces the logarithm of corporate pollutant emissions by an average of 0.366.

Second, green finance achieves synergistic emission reduction through the dual pathway of "alleviating financing constraints—promoting green technological innovation". It reduces the SA index to ease financing pressure and increases green utility model patent applications to enhance technological innovation.

Third, green finance's emission reduction effect exhibits significant heterogeneity: it is more prominent in non-SOEs, eastern regions, and manufacturing enterprises, but insignificant in SOEs, central/western regions, and non-manufacturing sectors.

### 5.1 Policy Recommendation

At the level of optimizing green finance practices, first of all, it is necessary to strengthen the intensity of differentiated support. This can be achieved by increasing the scale of green re loans, loan interest subsidies, and risk compensation, with a focus on tilting these supports toward the central and western parts of the country, as well as state-owned enterprises to decrease the financing costs associated with green projects. Meanwhile, efforts should be made to synchronously improve the construction of green project databases and the environmental information disclosure mechanism of enterprises, so as to provide a fundamental guarantee for precise support.

Secondly, it is essential to promote the coordinated development of the industrial chain system. On one hand, special funds should be established to support green technology R&D and equipment upgrading in the manufacturing industry; on the other hand, financial institutions should be guided to develop exclusive green credit products selected based on the characteristics of the non-manufacturing industry, so as to realize the coverage of green financial services across the entire industrial chain.

On this basis, it is required to establish and improve a dynamic monitoring system for emission reduction performance, build a dynamic management platform that links the use of green financial funds with the real-time emission data of enterprises, and connect the emission reduction performance with loan interest rates and credit limits. This will use market-oriented means to force the implementation of emission reduction responsibilities.

Finally, we need to strengthen the construction of policy synergy and data sharing mechanisms, promote the connection and integration of green finance policies with environmental policies such as carbon trading and pollution permits, and at the same time break down data barriers. By strengthening the data intercommunication between enterprise environmental information disclosure and financial supervision departments, a joint force for the development of green financial services can be formed.

### 5.2 Limitations and Future Research

This study has limitations: first, the sample focuses on large listed companies, and future research could include SMEs to expand generalizability; second, only two mechanism pathways are examined, and future work could incorporate government supervision, media attention, and supply chain pressure to explore "multi-stakeholder governance" scenarios.

Further, the studies may also conduct a cross-country analysis of how ESG and ENT practices could help to address the national economic policies, cultural attitudes, and local market conditions to overcome investment volatility.

In conclusion, this study provides micro-level evidence for green finance policy evaluation and offers references for building a precise, efficient green finance system, which is of great practical significance for securing the achievement of the "dual-carbon" goals.

## COMPETING INTERESTS

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

## REFERENCES

[1] Wang Y, Magnan M, Qiu Y. Going Green: An Analysis of Mutual Fund Investment and Green Innovation in the Context of China. Environmental and Resource Economics, 2025, 1-38. DOI: https://doi.org/10.1007/s10640-025-01035-4.

[2] Liu H K, He C. The mechanism and test of green finance promoting high-quality urban economic development: Empirical evidence from 272 prefecture-level cities in China. Investment Research, 2021, 40(07): 37-52.

[3] Zhang C, Hu Z G. How does green finance promote the coordinated advancement of urban pollution reduction and carbon emission reduction? Empirical evidence from green finance reform and innovation pilot zones. Commercial Research, 2025(02): 51-62.

[4] Wang K, Bai S Z, Tan J X, et al. The Impact of the Integration of Digital and Real Economies on the Synergistic Effect of Pollution Reduction and Carbon Emission Reduction in China. Research of Environmental Sciences, 2025, 38(08): 1668-1677.

[5] Liu J K, Liu J X, Chao Y. Green credit and low-carbon transformation: Capital integration or technological innovation? Evidence from a quasi-natural experiment. Journal of Quantitative & Technical Economics, 2024, 41(06): 151-171.

[6] Yan Y T, Shi G F, Xue J X. Research on the pollution and carbon reduction effect of structural monetary policy: A discussion on the structural effect of MLF collateral expansion. Modern Finance Research, 2025, 30(02): 61-71.

[7] Liu J Q, Yin G Y, Wu J H. Does the carbon emission trading policy promote corporate digital transformation?. Economic Perspectives, 2024(08): 75-92.

[8] Guo K S, Tian X X. Green technology and the green transformation of industrial development models. Tianjin Social Sciences, 2024(02): 99-107.

[9] Pan D Y, Chen C Q, Grubb M. Financial policies and low-carbon economic transformation: A study from the growth perspective. Journal of Financial Research, 2021(12): 1-19.

[10] Cheng X J, Ye J H. The driving effect of green finance on the sustainable development performance of energy enterprises. Industrial Economic Review (Shandong University), 2024, 23(01): 57-84.

[11] Chen D H. Bank fintech credit decision-making, and the sustainable development of borrowing enterprises. Financial Theory and Practice, 2024(07): 31-46.

[12] Ren X S. Green bond issuance and corporate ESG performance. Contemporary Economic Management, 2024, 46(04): 71-83.

[13] Zhang Y, Liu J Y. Green finance, green technological innovation, and high-quality economic development. Science Technology and Industry, 2025, 25(10): 313-318.

[14] Ren Z R. How does green finance promote the green low-carbon development of the economy?. Financial Development Research, 2025(07): 58-67.

[15] Mao X M, Wang R Z. Green finance and new-quality productive forces: Promotion or inhibition? From the perspective of technological innovation and environmental attention. Journal of Shanghai University of Finance and Economics, 2024, 26(05): 30-45.

[16] Zhang X Y, Guo A J, Lei Z H. Does green finance improve corporate total factor productivity? A quasi-natural experiment based on green finance reform and innovation pilot zones. Journal of University of Jinan (Social Science Edition), 2024, 34(02): 86-100.

[17] Meng W F, Liu J H. The Effect and Heterogeneity of Green Finance Promoting High-Quality Economic Development: From the Perspective of Technological Innovation and Industrial Structure Upgrading. Economic Review, 2023, (07): 100-110.

[18] Zhao Y. Green finance and urban planning: accelerating China's wind energy for environmental sustainability. GeoJournal, 2025, 90(5): 224.

[19] Habib A, Khan A M. Bank financing and green financing: The influence of sustainable practices on investment sensitivity. Sustainable Futures, 2025(10): 100775. DOI: https://doi.org/10.1016/j.sftr.2025.100775.

[20] Yu Y. Green finance as a catalyst for sustainable development: exploring the role of green bonds and impact investment in advancing environmental goals. Journal of Fintech and Business Analysis, 2025, 2(2): 27-31.

**World Journal of Economics and Business Research**
**Print ISSN: 2960-0073**
**Online ISSN: 2960-0081**
**DOI: https://doi.org/10.61784/wjebr3062**

# CHALLENGES AND FUTURE TRENDS OF MACHINE LEARNING IN DIGITAL FINANCE: AN ANALYSIS OF INTERPRETABILITY, REGULATION, AND DATA GOVERNANCE

RongHua Li*, ShuRui Xiao
*School of Finance and Economics, Hainan Vocational University of Science and Technology, Haikou 571126, Hainan, China.*
*Corresponding Author: RongHua Li, Email: 1034064667@qq.com*

**Abstract:** Machine learning (ML) technologies are transforming digital finance through applications in credit assessment, fraud detection, and algorithmic trading. However, their deployment faces three critical challenges: model interpretability, robust data governance, and complex regulatory compliance. This paper analyzes these challenges through a systematic examination of recent literature and regulatory developments. We find that the "black-box" nature of complex models conflicts with transparency requirements mandated by financial regulations such as the EU AI Act and GDPR. Data quality issues, including class imbalance and inconsistency, coupled with privacy concerns, further constrain model reliability. Privacy-preserving approaches, particularly federated learning, offer promising solutions but require wider adoption. We identify that current model governance frameworks lack standardization across institutions and jurisdictions. Our analysis suggests that addressing these challenges requires coordinated efforts across three dimensions: advancing explainable AI (XAI) techniques, establishing unified model governance standards, and implementing privacy-preserving technologies. This study contributes to the understanding of socio-technical barriers in financial ML adoption and provides guidance for practitioners and policymakers.
**Keywords:** Machine learning; Digital finance; Explainable AI; Model interpretability; Regulatory compliance

## 1 INTRODUCTION

In recent years, the widespread application of machine learning (ML) and artificial intelligence (AI) in digital finance has significantly propelled innovation and transformation within the financial industry. Leveraging their capabilities for efficient mining of massive datasets, strong non-linear modeling advantages, and adaptive intelligent decision-making features, ML methods have progressively permeated nearly all critical financial segments. These include credit approval, asset valuation, fraud detection, algorithmic trading, and personalized financial services, establishing ML as a core driver of the intelligent enhancement of financial services [1,2]. Statistics indicate that the global AI in FinTech market is projected to grow significantly, reflecting the sustained impetus of this technological wave on industry development [3].

Although advanced ML methods, such as deep learning and ensemble learning, have demonstrated exceptional performance in financial data analysis and decision-making, their inherent "black-box" structures, ambiguous causal linkages, and lack of decision interpretability pose unprecedented challenges to financial regulation, risk management, and consumer rights protection [3]. Particularly in business scenarios characterized by stringent regulation and high legal transparency requirements, the non-interpretable nature of ML models not only impacts compliance but may also lead to erroneous risk identification and reputational crises [4]. Furthermore, issues such as data quality, privacy protection, model bias, and the absence of industry standards have erected significant barriers to the intelligent transformation of digital finance [2,5,6].

In response to these complex and multifaceted challenges, the global financial industry and academia are continuously driving technological innovation, optimizing governance structures, and fostering cross-sector collaboration, attempting to find an equilibrium between enhancing financial service efficacy and managing technological risks [2,4]. This paper systematically reviews the primary difficulties facing ML in digital finance—including interpretability, regulatory compliance, data governance, and internal model governance. It assesses the industry's current explorations in standardization and multi-method integration, and forecasts future trends such as algorithmic innovation, cross-regional knowledge sharing, and sustainable development.

This paper systematically examines three primary challenge domains: model interpretability and transparency (Section 2.1), regulatory and ethical considerations (Section 2.2), and data quality and availability (Section 2.3). Section 3 discusses emerging responses, including standardization initiatives and privacy-preserving technologies. We conclude by identifying priority areas for coordinated action among financial institutions, technology providers, and regulatory bodies.

## 2 MAIN CHALLENGES AND CORE LIMITATIONS

### 2.1 Interpretability and Model Transparency

Financial regulations impose stringent requirements on model interpretability, particularly in scenarios such as credit approval and high-risk asset management [7,8]. Although complex models like deep learning enhance predictive accuracy, their "black-box" nature renders the decision-making process difficult to explain, creating a core obstacle to compliance and trust-building [3]. This opacity magnifies regulatory and reputational risks, as regulators and stakeholders demand sufficient control over model outputs and decision logic [4].

To mitigate the "black-box" problem, eXplainable Artificial Intelligence (XAI) techniques have emerged as a focal point. Methods such as SHAP (Shapley Additive exPlanations) and PDP (Partial Dependence Plots) are employed to reveal the marginal contributions of model features and their non-linear relationships. In credit rating, for example, SHAP analysis indicates that financial indicators like total revenue, asset turnover ratio, and ICR (Interest Coverage Ratio) have a decisive impact on ratings, and credit quality deteriorates when the ICR falls below 2.0. However, XAI methods themselves present new limitations, including applicability constraints and inconsistencies in explanations [3].

Technology alone is insufficient to fully resolve transparency concerns; financial institutions must also establish robust model governance frameworks. These frameworks should encompass result attribution, comprehensive documentation for development and deployment, periodic validation, and supervision across the entire project lifecycle [4]. Regulatory standards vary significantly across global jurisdictions, yet there is a universal trend toward strengthening model risk management and compliance reviews [9].

### 2.2 Regulatory and Ethical Considerations

The financial application of AI/ML is currently situated in a phase of continuously evolving regulation. Various jurisdictions are establishing systemic requirements for data management, model governance, and third-party audits to ensure model transparency, accountability, and fairness [2,9]. Furthermore, regulations such as the EU's GDPR and the US's CCPA have imposed higher standards for data privacy protection, impacting sample selection and modeling processes [2].

Bias and fairness are critical ethical issues. Systemic biases within historical data can lead to discriminatory model outputs, impeding equitable access to financial services for certain groups. While ML can mitigate some human biases, it can also amplify unfairness if data, feature selection, or model design is improper [2,5]. In response, the industry is beginning to implement countermeasures, such as introducing fairness constraints, enhancing transparent documentation, and assembling diverse development teams.

Systemic risks are gradually emerging. Model convergence—the widespread adoption of services from the same vendor—may trigger a financial "monoculture" and herd behavior risks, potentially exacerbating market instability in extreme scenarios [2].

### 2.3 Model Governance and Risk Management

Beyond external regulatory pressures, significant shortfalls exist in internal governance capacity building. Most institutions still lack comprehensive responsible AI development processes, clear accountability structures, and continuous monitoring measures [4]. Establishing robust Model Risk Management (MRM) frameworks, as highlighted by industry analysis, is essential for identifying, assessing, and mitigating risks throughout the entire model lifecycle [4]. This internal governance is critical for ensuring that models, including third-party vendor solutions, align with the institution's risk appetite and ethical standards [3,9]. Future regulatory requirements are expected to place greater emphasis on this "dynamic governance," promoting cross-departmental and inter-agency collaboration [10].

### 2.4 Data Quality and Availability

Data barriers directly impact model performance. Scenarios such as fraud detection and credit risk assessment often suffer from class imbalance, where the disparity between positive and negative samples hinders the model's ability to detect high-value, rare risk behaviors [5,11]. Research recommends sampling techniques like SMOTE and ensemble enhancement algorithms to improve minority class recognition capabilities [5].

Another challenge is data consistency. Data in large-scale institutions often originate from diverse sources with disparate standards, lacking unified semantics and synchronized update mechanisms. This introduces noise into system modeling and can even exacerbate model drift [6]. High-quality preprocessing pipelines, anomaly detection, and cleansing mechanisms are thus critical components for constructing an intelligent financial infrastructure [3,6].

Privacy regulations are becoming increasingly stringent. The GDPR, for instance, requires financial institutions to hold only minimal necessary data and to guarantee users' rights to information and erasure [2]. Privacy-preserving machine learning (PPML) techniques, such as federated learning, are instrumental in enhancing inter-institutional modeling capabilities while reducing the risks associated with centralized storage of sensitive data [12].

Traditional finance has primarily relied on structured data. Now, various forms of unstructured and alternative data (e.g., text, social media, images) are being incorporated into ML analysis. While this vastly expands the breadth of model prediction, it also introduces significant challenges related to data reliability and validation.

Disparities in data accessibility and quality contribute to regional imbalances in digital finance development. Emerging markets, constrained by limited infrastructure, struggle to support complex ML solutions, which in turn raises issues of financial inclusion and policy equity [13]. Financial institutions equipped with mature data governance platforms and sophisticated feature engineering capabilities possess a distinct advantage in the pursuit of intelligent transformation [3].

## 3 FUTURE DIRECTIONS AND EMERGING TRENDS

### 3.1 Model Integration and Standardization

The adoption of machine learning in the financial industry has often occurred in a fragmented manner, with institutions developing specialized applications for discrete challenges rather than implementing comprehensive, integrated approaches [3]. This pattern has led to a proliferation of models with divergent methodologies, governance structures, and performance characteristics—creating significant challenges for standardization, interoperability, and institutional oversight.

Research indicates a persistent gap in the field: the lack of standardized frameworks for implementing ML across the financial sector [3]. Although individual applications have demonstrated notable success, the absence of common standards impedes broader adoption, complicates regulatory oversight, and creates potential inefficiencies in development and deployment. As ML becomes increasingly central to financial operations, establishing standardized methodologies represents a key priority for both practitioners and researchers.

Several key areas for standardization have emerged as particularly crucial for advancing ML integration in finance. First, model development methodologies require standardization to ensure consistent quality, appropriate validation, and responsible implementation practices [3]. This includes standardized approaches to data preprocessing, feature engineering, model selection, hyperparameter optimization, and validation procedures that can be applied across diverse financial applications. Second, model documentation standards are essential for ensuring transparency, facilitating audit processes, and supporting regulatory compliance [4]. Comprehensive documentation should cover model objectives, methodological choices, data sources, performance metrics, validation procedures, limitations, and governance structures. Standardized "model cards" or documentation templates can promote consistent information capture while supporting effective review and oversight. Third, interpretability frameworks need standardization to ensure model decisions can be adequately explained and justified [14]. Different XAI techniques may offer varying insights; standardized methods for selecting and applying these techniques can help ensure explanations are meaningful, accurate, and consistent. This standardization is particularly vital for regulated applications where decision transparency is required to ensure compliance and customer trust. Fourth, model risk management (MRM) frameworks require standardization to ensure appropriate governance and oversight of ML implementations [4]. This involves standardized approaches to risk identification, assessment, mitigation, and monitoring throughout the entire model lifecycle. Given the potential consequences of model failure in the financial context, robust and consistent MRM practices are essential for responsible ML implementation.

As financial institutions continue to expand their ML implementations, developing standardized, integrated approaches will become increasingly critical to ensure effective governance, consistent quality, and responsible deployment.

### 3.2 Future Challenges and Opportunities

Continued algorithmic innovation offers substantial opportunities for enhanced performance across financial applications. Emerging research directions include: transformer architectures and large language models for processing financial text and time series; reinforcement learning for dynamic portfolio optimization and adaptive trading strategies; graph neural networks for modeling complex relationship networks in fraud detection and credit risk; and hybrid neuro-symbolic systems combining neural networks' pattern recognition with symbolic reasoning's interpretability [14,15]. These advances may enable more accurate predictions, more nuanced risk assessment, and more personalized services while potentially improving explainability.

Natural Language Processing (NLP) applications present particularly promising opportunities. Advanced models can extract insights from earnings call transcripts, regulatory filings, news sentiment, and social media to enhance market forecasting, event prediction, and risk assessment [12,14]. Research demonstrates that text-based models analyzing annual report language can identify early warning signals of financial distress not captured by traditional financial ratios, potentially improving credit risk models and investment strategies.

Federated learning and related privacy-preserving techniques hold significant potential for enabling collaborative learning without compromising data confidentiality [6]. For fraud detection, cross-institutional FL could leverage broader transaction patterns to improve detection rates while maintaining customer privacy and regulatory compliance. However, realizing this potential requires addressing technical challenges (communication efficiency, Byzantine robustness) and establishing governance frameworks for multi-party collaboration including data sharing agreements, liability allocation, and benefit distribution [16].

Organizational capacity building remains essential for effective ML implementation. The specialized skills required—spanning statistical modeling, software engineering, domain expertise, and ethical awareness—are scarce across industries [3]. Financial institutions must develop strategies for talent acquisition and retention, invest in continuous learning programs, and cultivate organizational cultures that balance innovation with risk management. Leadership commitment and cross-functional collaboration between technology, risk, compliance, and business units are critical success factors.

Cross-regional collaboration and knowledge sharing present important opportunities. Current ML research and implementation exhibit significant geographical concentration in developed markets and major financial centers [13]. Expanding initiatives to diverse contexts can enhance understanding of how ML applications perform under different institutional environments, regulatory regimes, and market structures. This expansion is particularly valuable for advancing financial inclusion objectives and ensuring equitable access to AI-driven financial services in emerging economies.

Finally, interdisciplinary research is essential for addressing the complex socio-technical challenges at finance, technology, ethics, and regulation intersections. Sustained progress requires collaboration among financial economists, computer scientists, legal scholars, ethicists, and policymakers to develop frameworks that promote innovation while addressing regulatory compliance, ethical concerns, and societal impacts [3,10]. This collaborative approach can inform the development of "responsible by design" practices embedding transparency, fairness, and accountability into ML systems from inception.

## 4 CONCLUSION

Machine learning has comprehensively permeated the financial services ecosystem, driving significant efficiency gains in areas such as credit risk, fraud detection, and algorithmic trading. However, challenges related to model interpretability, data quality, internal model governance, external regulatory compliance, and ethical risks are increasingly prominent. The industry urgently needs to strengthen transparent governance frameworks, promote data standardization and privacy protection, and strike a balance between innovation and risk. Driven by multiple forces—including algorithmic innovation, standardized governance, and sustainability considerations—the wave of intelligent transformation in digital finance is expected to continue to deepen in the coming years.

## COMPETING INTERESTS

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

## FUNDING

This article is the Scientific Research Project of Higher Education Institutions in Hainan Province in 2025: Research on the Driving Mechanism and Realisation Path of the High-quality Development of Hainan's Advanced Manufacturing Industry under the New Pattern of Double-cycle Development (Project No. Hnky2025-63).

## REFERENCES

[1] Jada Irshaad, Mayayise TO. The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. Data and Information Management, 2024, 8(2): 100063.

[2] Blatch-Jones A, Church H, Crane K. Exploring the potential benefits and challenges of artificial intelligence for research funding organisations: a scoping review. F1000Research, 2025, 14: 126.

[3] Černevičienė J, Malaškevičius K, Gazdzicki K. Explainable artificial intelligence (XAI) in finance: A systematic literature review. Artificial Intelligence Review, 2024, 58: 1-28.

[4] Chartis Research. Mitigating model risk in AI: Advancing an MRM framework for AI/ML models at financial institutions. 2024. https://www.chartis-research.com/custom-insights/7947296/mitigating-model-risk-in-ai-advancing-an-mrm-framework-for-aiml-models-at-financial-institutions.

[5] Murikah W, Nthenge JK, Musyoka FM. Bias and ethics of AI systems applied in auditing-A systematic review. Scientific African, 2024, 25: e02281.

[6] Li S, Ye F, Fang M, et al. Synergizing Foundation Models and Federated Learning: A Survey. arXiv, 2024. DOI: https://doi.org/10.48550/arXiv.2406.12844.

[7] Weber P, Carl KV, Hinz O. Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature. Management Review Quarterly, 2024, 74: 867-907.

[8] Nallakaruppan MK, Narayanan C, Rajasekaran KM. An explainable AI framework for credit evaluation and analysis. Applied Soft Computing, 2024, 153: 111307.

[9] Ridzuan NN, Masri M, Anshari M, et al. AI in the Financial Sector: The Line between Innovation, Regulation and Ethical Responsibility. Information, 2024, 15(8): 432.

[10] OECD. Futures of Global AI Governance: Background Note. 2024. https://www.oecd.org/content/dam/oecd/en/about/programmes/strategic-foresight/GSG%20Background%20Note_GSG(2024)1en.pdf.

[11] Gabellini M, Mak S, Schoepf S, et al. A continuous training approach for risk informed supplier selection and order allocation. Production & Manufacturing Research, 2025, 13(1): 2447035.

[12] Stanford University. Artificial intelligence (AI) adoption in financial services worldwide from 2023 to 2024, by function. Statista, 2025, 04. https://www.statista.com/statistics/1560120/ai-adoption-financial-services-by-function/.

[13] Chishti MZ, Dogan E, Binsaeed RH. Can artificial intelligence and green finance affect economic cycles?. Technological Forecasting and Social Change, 2024, 209: 123740.

[14] Chen X, Tian W, Fang H. Bibliometric analysis of natural language processing using CiteSpace and VOSviewer. Natural Language Processing Journal, 2025, 10: 100123.

[15] Gupta P, Ding B, Guan C, et al. Generative AI: A systematic review using topic modeling techniques. Data and Information Management, 2024, 8(2): 100066.

[16] Arora S, Beams A, Chatzigiannis P, et al. Privacy-Preserving financial anomaly detection via federated learning & multi-party computation. in 2024 Annual Computer Security Applications Conference Workshops (ACSAC Workshops), Honolulu, HI, USA, 2024, 270-279.

**World Journal of Economics and Business Research**
**Print ISSN: 2960-0073**
**Online ISSN: 2960-0081**
**DOI: https://doi.org/10.61784/wjebr3063**

# MACHINE LEARNING IN DIGITAL FINANCE: APPLICATIONS, METHODS AND CHALLENGES

Yun Li*, RongHua Li
*School of Finance and Economics, Hainan Vocational University of Science and Technology, Haikou 571126, Hainan, China.*
*Corresponding Email: 704823609@qq.com*

**Abstract:** Machine learning (ML) technology is profoundly reshaping digital financial services and risk management systems. This paper systematically reviews the current applications of ML in four core scenarios: credit scoring, fraud detection, algorithmic trading, and customer segmentation. Literature analysis reveals that deep learning and ensemble methods demonstrate superior performance compared to traditional statistical approaches in financial risk prediction tasks, with supervised learning techniques predominating in fraud detection systems and algorithmic trading becoming increasingly prevalent in capital markets. However, the trade-off between model interpretability and predictive performance remains a critical challenge in regulated financial environments. Furthermore, data quality limitations and regulatory compliance requirements impose substantial constraints on model deployment. This review identifies key research gaps and suggests that future developments must prioritize explainable AI techniques, privacy-preserving methods, and regulatory-compliant frameworks to enable the sustainable adoption of ML in the financial sector.
**Keywords:** Machine learning; Digital finance; Credit scoring; Fraud detection; Model interpretability

## 1 INTRODUCTION

Machine learning (ML) is driving a profound transformation within the financial industry. Investment in artificial intelligence (AI) by financial institutions is steadily increasing, with annual expenditure in this domain projected to reach $100 billion in the United States by 2025, and the global figure approaching $200 billion [1]. The core capabilities of ML—including high-dimensional data processing, non-linear pattern recognition, and adaptive learning mechanisms—enable it to demonstrate significant potential in algorithmic trading, credit risk assessment, fraud prevention, and personalized financial services [2-3].

In recent years, the financial industry has increasingly adopted advanced ML architectures, encompassing deep learning networks, ensemble techniques, and reinforcement learning algorithms. These technologies are fundamentally reshaping critical business processes including risk management, customer analytics, and automated decision-making [1, 4]. Deep learning models, particularly neural networks with multiple hidden layers, have shown remarkable capabilities in capturing complex relationships within financial data. Ensemble methods such as random forests and gradient boosting have become standard tools for improving prediction accuracy through the aggregation of multiple base models [5]. Meanwhile, reinforcement learning techniques are enabling dynamic strategy optimization in trading systems and portfolio management [2].

However, alongside these technological advances, significant challenges have emerged that require systematic investigation. Model interpretability remains a central concern, as financial regulators increasingly demand transparent decision-making processes, particularly in credit allocation and automated underwriting [3]. Data governance issues, including quality assurance, privacy protection, and ethical use of personal information, impose substantial constraints on model development and deployment [4]. Furthermore, the generalization performance of ML models under distributional shifts and adversarial conditions raises questions about their reliability in real-world financial environments.

This paper provides a comprehensive review of ML applications across four essential financial scenarios: credit scoring, fraud detection, algorithmic trading, and customer segmentation. We examine the mainstream methodologies employed in each domain, analyze their comparative strengths and limitations, and identify critical challenges that impede broader adoption. Our review synthesizes recent developments in the literature to provide both academic researchers and industry practitioners with a reference framework for understanding the current state and future directions of intelligent transformation in digital finance.

## 2 MACHINE LEARNING APPLICATIONS IN FINANCIAL SERVICES

### 2.1 Credit Scoring and Risk Assessment

Credit scoring represents one of the most mature applications of ML in the financial industry. Traditional credit assessment models, which typically rely on limited structured data sources and linear statistical methods, often result in the "credit invisibility" phenomenon where individuals with thin credit files are systematically excluded from formal credit markets—a challenge particularly pronounced in emerging economies [6]. Modern ML approaches address these limitations by incorporating diverse data sources including transactional records, digital payment histories, mobile phone usage patterns, and online behavioral footprints. Advanced techniques such as deep neural networks (DNNs) and ensemble learning algorithms have substantially enhanced risk stratification capabilities and predictive accuracy [7].

Empirical studies demonstrate that deep learning architectures generally achieve superior performance compared to traditional methods such as logistic regression and basic decision trees in credit risk classification tasks. Ensemble algorithms, which combine predictions from multiple base models, further enhance robustness and generalization across different borrower populations [7]. Logistic regression nevertheless remains valuable in credit assessment due to its computational efficiency and inherent interpretability—qualities that facilitate regulatory compliance and auditability [8]. Random forest algorithms, owing to their ability to capture non-linear feature interactions and rank variable importance, have gained widespread adoption in credit scoring systems [8, 9]. Furthermore, specialized deep learning architectures including Multilayer Perceptrons (MLP) and Long Short-Term Memory (LSTM) networks demonstrate exceptional performance in handling high-dimensional feature spaces and temporal sequences, with several commercial banking institutions having integrated these models into production credit evaluation systems [7].

Despite these advances, several challenges persist in ML-based credit scoring. The interpretability-accuracy trade-off presents a fundamental dilemma: while complex models may achieve higher predictive performance, their opaque decision processes complicate regulatory approval and customer communication. Additionally, concerns regarding algorithmic fairness and potential discrimination against protected demographic groups require careful attention to model design and validation procedures.

### 2.2 Fraud Detection and Security

As financial services increasingly migrate to digital platforms, fraud patterns have evolved in sophistication and diversity, rendering traditional rule-based detection systems inadequate for identifying novel attack vectors in real-time operational environments. Machine learning, with its capacity for processing large-scale transactional data streams and capturing evolving behavioral patterns, has emerged as the predominant approach to fraud detection [9, 10]. Research synthesis indicates that supervised learning methods constitute the majority of fraud detection implementations, with ensemble techniques and deep learning architectures proving particularly effective in addressing the severe class imbalance characteristic of fraud datasets [7, 9].

Classical algorithms including random forest (RF), support vector machines (SVM), logistic regression (LR), and decision trees (DT) are frequently integrated into ensemble configurations to create robust detection systems that leverage the complementary strengths of different model families [9]. Recent developments in neural network architectures—encompassing Multilayer Perceptrons (MLP), Back-Propagation Neural Networks (BPNN), and Long Short-Term Memory (LSTM) networks—have further enhanced detection capabilities. Hybrid models that combine supervised learning with anomaly detection techniques have shown particular promise, with some implementations achieving detection accuracy exceeding 99% while maintaining low false positive rates [7, 9]. Beyond transactional analysis, complementary technologies such as biometric authentication, behavioral analytics, and network analysis are increasingly deployed in online anti-money laundering (AML) systems and transaction security frameworks [3].

Critical challenges in fraud detection include the need for real-time inference with minimal latency, the management of highly imbalanced datasets where fraudulent transactions represent a tiny minority of cases, and the continuous adaptation to adversarial tactics as fraudsters develop countermeasures to detection systems. The interpretability of fraud detection models is also important for investigation workflows and regulatory reporting requirements.

### 2.3 Algorithmic Trading and Investment Strategies

In the investment and trading domain, ML-driven algorithmic systems have fundamentally transformed market forecasting capabilities and strategy adaptation mechanisms. Algorithmic trading now accounts for approximately 70% of equity trading volume in U.S. markets, reflecting the widespread adoption of automated decision-making systems [2]. Recent advances in deep reinforcement learning (DRL) techniques, including Q-learning and policy gradient methods, combined with alternative data sources such as sentiment analysis from social media platforms and satellite imagery, have significantly expanded both the feature space and predictive power of trading models [2].

Machine learning applications in this domain extend beyond traditional technical indicators such as price and volume to encompass complex tasks including derivatives pricing, high-frequency trading (HFT) strategy optimization, and dynamic asset allocation [5]. Deep reinforcement learning methods offer distinctive advantages by enabling systems to learn optimal trading policies through interaction with market environments, continuously refining strategies based on reward signals that reflect trading performance and risk metrics. This adaptive, self-learning capability provides resilience in responding to market regime changes and evolving microstructure dynamics [2, 5].

However, algorithmic trading systems face several inherent challenges. Overfitting to historical data patterns can lead to poor out-of-sample performance when market conditions shift. Transaction costs, market impact, and liquidity constraints often significantly erode theoretical profitability in live trading. Additionally, the potential for algorithmic instability and flash crashes raises systemic risk concerns that necessitate robust risk management protocols and regulatory oversight.

### 2.4 Customer Segmentation and Personalization

Customer segmentation constitutes an important application area where financial institutions leverage ML to achieve service differentiation and personalization strategies. Through unsupervised learning techniques such as clustering algorithms and dimensionality reduction methods, ML systems can identify latent behavioral structures and preference patterns from heterogeneous data sources, enabling precise customer classification and granular profiling [3, 11]. K-means clustering has emerged as a widely adopted segmentation approach, capable of efficiently processing large-scale high-dimensional behavioral datasets and rapidly identifying distinct market segments and emerging customer cohorts [11].

A fundamental advantage of ML in customer segmentation lies in its capacity to synthesize multi-layered information streams—including transaction histories, social network data, online interaction patterns, and unstructured text feedback—to support personalized product recommendations, customized service offerings, and targeted marketing campaigns [12]. These capabilities enhance customer satisfaction metrics, improve conversion rates, and create opportunities for financial institutions to optimize resource allocation across customer segments and implement sophisticated cross-selling strategies.

Advanced segmentation approaches now incorporate temporal dynamics, enabling the identification of customer lifecycle stages and behavioral transitions. However, the increasing use of personal data in segmentation models raises significant privacy concerns and necessitates compliance with data protection regulations. Financial institutions must balance the pursuit of personalization benefits against the requirements for data minimization and customer consent.

## 3 MAINSTREAM MODEL METHODOLOGIES

### 3.1 Supervised Learning in Finance

Supervised learning techniques constitute the predominant category of ML methods deployed across financial applications, with empirical surveys indicating their prevalence in the majority of implemented systems spanning credit evaluation, fraud detection, and market forecasting domains [5, 9]. This prevalence reflects the availability of labeled historical data in many financial contexts—a prerequisite for training supervised models to recognize patterns that generalize to new observations.

Logistic Regression (LR) remains among the most frequently deployed supervised learning algorithms in financial applications, particularly for binary classification problems such as default prediction and fraud identification. Despite its relative simplicity compared to modern deep learning architectures, LR consistently delivers robust performance while offering the critical advantage of high interpretability through readily understandable coefficients and odds ratios. This transparency proves especially valuable in regulated financial environments where institutions must explain automated decisions to customers and regulators [7, 8]. Comparative studies report that well-calibrated LR models can achieve strong performance in credit assessment scenarios, sometimes rivaling more complex methods when feature engineering is appropriately conducted.

Decision Trees (DT) represent another common supervised learning method, providing intuitive interpretability through their hierarchical structure that partitions data based on feature thresholds. The resulting tree-like decision model can be readily visualized and explained to non-technical stakeholders, making DTs valuable in regulatory compliance applications where model transparency is essential [7]. However, individual decision trees are prone to overfitting and high variance, which motivates the use of ensemble methods.

The Random Forest (RF) algorithm has gained substantial adoption in financial applications due to its superior performance and inherent robustness against overfitting [7, 9]. By aggregating predictions from multiple decision trees trained on bootstrap samples with random feature subsets, random forests capture complex non-linear relationships while maintaining reasonable interpretability through feature importance metrics. Ensemble methods of this type often outperform single algorithms and prove particularly effective for handling the class imbalance problems endemic to financial datasets [9].

Support Vector Machines (SVM) have demonstrated competitive performance in various financial classification tasks, especially when dealing with high-dimensional feature spaces. By identifying an optimal separating hyperplane that maximizes the margin between classes in a transformed feature space, SVMs effectively handle complex data patterns. However, their reduced interpretability relative to LR or DTs, combined with computational challenges at large scale, limits their adoption in some production environments.

Neural Networks (NNs), including Multilayer Perceptrons (MLPs), have shown exceptional capabilities in financial modeling scenarios involving complex, non-linear relationships that simpler models cannot adequately capture. These feedforward architectures with multiple hidden layers can approximate arbitrary continuous functions, making them suitable for diverse financial prediction tasks. However, the performance gains of neural networks often come at the cost of reduced interpretability—a significant consideration in applications subject to regulatory scrutiny.

### 3.2 Deep Learning Applications

The Multilayer Perceptron (MLP) represents one of the most extensively utilized deep learning architectures in financial modeling. These feedforward neural networks consist of multiple layers of interconnected nodes, where each node applies a non-linear activation function to weighted inputs from the preceding layer. MLPs have demonstrated considerable efficacy across various financial applications, with studies reporting strong performance in fraud identification tasks [7, 9]. Their capacity to approximate complex non-linear functions makes them particularly suitable for modeling intricate financial relationships that simpler parametric models cannot adequately represent.

Convolutional Neural Networks (CNNs), originally developed for computer vision tasks, have found innovative applications in finance through their application to structured sequential data such as time-series financial information. By applying convolutional operations that detect local patterns across temporal windows, these networks can identify hierarchical features indicative of market trends, price movements, or anomalous transaction sequences [2]. The spatial invariance property of CNNs makes them valuable for detecting patterns regardless of their temporal position within input sequences.

Recurrent Neural Networks (RNNs), and particularly Long Short-Term Memory (LSTM) networks, have demonstrated exceptional capabilities in analyzing sequential financial data including transaction histories, price trajectories, and temporal behavioral patterns. These architectures incorporate memory cells that retain information over extended sequences, enabling them to capture long-term dependencies in financial time series. LSTM networks have been successfully deployed in fraud detection systems, with some implementations achieving very high detection rates while maintaining operational efficiency [7]. Bidirectional LSTM (Bi-LSTM) models represent a further advancement, processing sequential data in both forward and backward directions to capture contextual information from both past and future states. Hybrid architectures combining Bi-LSTM with autoencoder and anomaly detection techniques have shown promising results in identifying fraudulent patterns [7].

Deep Reinforcement Learning (DRL) has emerged as a powerful methodology for optimizing sequential decision-making in trading and portfolio management. By combining deep neural networks with reinforcement learning principles, these models learn optimal action policies through iterative interaction with simulated or real market environments. DRL methods demonstrate particular promise in algorithmic trading contexts where they adapt to evolving market conditions and optimize trading parameters based on reward signals reflecting profitability and risk metrics [2].

### 3.3 Ensemble and Hybrid Methods

Ensemble and hybrid methods have emerged as particularly effective strategies for addressing complex challenges in financial applications. By combining multiple algorithms or modeling paradigms, these approaches leverage the complementary strengths of different models while mitigating their individual weaknesses. Empirical evidence consistently demonstrates that ensemble methods generally achieve superior performance compared to single models across diverse financial prediction tasks [5, 9].

Random Forest represents one of the most widely implemented ensemble techniques in finance. By aggregating predictions from multiple decision trees trained on different bootstrap samples and random feature subsets, random forests reduce overfitting while enhancing both accuracy and stability. Their effectiveness has been validated across numerous financial applications, including credit scoring, fraud detection, and market forecasting, where they often achieve competitive performance with relatively modest computational requirements [7].

Gradient Boosting methods, including XGBoost and LightGBM, have garnered significant attention in financial modeling for their exceptional predictive performance and computational efficiency [5]. These algorithms construct an ensemble of weak predictive models—typically shallow decision trees—in a sequential manner, where each new model focuses on correcting the errors of its predecessors through gradient descent optimization. Systematic reviews of ML applications in credit risk assessment identify boosting techniques as consistently ranking among the top-performing methods for complex tasks involving numerous features and intricate non-linear relationships [7]. Beyond credit scoring, gradient boosting has demonstrated strong results in fraud detection, with XGBoost-based systems achieving high accuracy while maintaining interpretability through feature importance analysis.

Stacking, also known as stacked generalization, represents a more sophisticated ensemble approach that combines predictions from multiple diverse base models through a meta-learner that weights their contributions [5, 7]. This hierarchical structure allows the meta-model to learn optimal combinations of base model predictions, potentially achieving performance superior to any individual component. Hybrid models that integrate different ML paradigms show particular promise in complex financial scenarios. For example, combinations of supervised and unsupervised learning techniques can address distinct aspects of a problem—using unsupervised methods for feature extraction or anomaly detection before applying supervised classification. The integration of deep learning architectures with traditional statistical methods represents another valuable direction, as it can enhance predictive accuracy while preserving some degree of interpretability required in regulated environments [5].

## 4 KEY CHALLENGES AND FUTURE DIRECTIONS

Machine learning adoption in finance faces several critical challenges. The interpretability-performance trade-off represents a fundamental tension: complex models such as deep neural networks achieve superior accuracy but operate as "black boxes" that resist explanation. In regulated contexts requiring justification of credit decisions and fraud alerts, this opacity creates significant barriers. While explainable AI (XAI) techniques offer post-hoc explanations, their integration into production systems remains challenging [3, 8].

Data constraints pose another major obstacle. Financial ML models require large volumes of high-quality labeled data, yet institutions face data fragmentation, inconsistent formats, and labeling errors. Class imbalance—where events like fraud represent tiny dataset minorities—complicates training and requires specialized techniques. Privacy regulations including GDPR further restrict data access and usage [3, 4].

Model robustness under distributional shifts presents ongoing concerns. Financial patterns evolve continuously, yet models trained on historical data may fail when conditions change. Adversarial actors actively adapt strategies to evade detection, necessitating continuous updates. Algorithmic bias against protected demographic groups requires careful attention to fairness metrics throughout the model lifecycle [3, 8].

Future research should prioritize interpretable model architectures that maintain competitive performance, privacy-preserving techniques such as federated learning enabling secure collaboration, causal inference methods distinguishing correlation from causation, and adversarial robustness enhancing model resilience in dynamic environments [5].

## 5 CONCLUSION

Machine learning is accelerating the transformation of financial services toward greater intelligence, efficiency, and precision. This review has synthesized current research on ML applications across four critical domains: credit scoring and risk assessment, fraud detection and security, algorithmic trading and investment strategies, and customer segmentation and personalization. Our analysis reveals that while supervised learning techniques remain predominant, deep learning architectures and ensemble methods increasingly demonstrate superior performance in complex prediction tasks. These technological advances are creating substantial value through improved risk management, enhanced operational efficiency, and more personalized customer experiences.

However, realizing the full potential of ML in finance requires addressing several persistent challenges. The interpretability-accuracy trade-off remains a central tension, particularly in regulated contexts where transparent decision-making is mandated. Data quality limitations, class imbalance, and privacy protection requirements constrain model development and deployment. Model robustness under distributional shifts and adversarial conditions presents ongoing risks. Future research must prioritize explainable AI techniques that provide meaningful transparency without sacrificing predictive power, privacy-preserving methods such as federated learning that enable secure collaboration, and robust learning frameworks that maintain performance under evolving conditions. By addressing these challenges through interdisciplinary collaboration among researchers, practitioners, and regulators, the financial industry can advance toward secure, ethical, and sustainable adoption of machine learning technologies.

## COMPETING INTERESTS

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

## FUNDING

This article is the Scientific Research Project of Higher Education Institutions in Hainan Province in 2025: Research on the Driving Mechanism and Realisation Path of the High-quality Development of Hainan's Advanced Manufacturing Industry under the New Pattern of Double-cycle Development (Project No. Hnky2025-63).

## REFERENCES

[1] Vuković DB, Dekpo-Adza S, Matović S. AI integration in financial services: a systematic review of trends and regulatory challenges. Humanit Soc Sci Commun, 2025, 12(1): 1-29.

[2] Salehpour Arash, Samadzamini Karim. Machine Learning Applications in Algorithmic Trading: A Comprehensive Systematic Review. International Journal of Education and Management Engineering (IJEME), 2023, 13(6): 41.

[3] Congress.gov. Artificial Intelligence and Machine Learning in Financial Services. 2025, 03. https://www.congress.gov/crs-product/R47997.

[4] International Association of Insurance. Regulation and supervision of artificial intelligence and machine learning (AI/ML) in insurance: a thematic review. 2023, 12. https://www.iais.org/uploads/2023/12/Regulation-and-supervision-of-AI-ML-a-thematic-review.pdf.

[5] Nazareth Noella, Reddy Yeruva Venkata Ramana. Financial applications of machine learning: A literature review. Expert Systems with Applications, 2023, 219: 119640.

[6] Svitla Team. Machine Learning for Credit Scoring: Benefits, Models, and Implementation Challenges. 2024. https://svitla.com/blog/machine-learning-for-credit-scoring.

[7] Shi S, Tse R, Luo W, et al. Machine learning-driven credit risk: a systemic review. Neural Computing and Applications, 2022, 34(17): 14327-14339.

[8] Hashimoto Ryuichiro, Miura Kakeru, Yoshizaki Yasunori. Application of Machine Learning to a Credit Rating Classification Model: Techniques for Improving the Explainability of Machine Learning. Bank of Japan, 2023: 23-E-6.

[9] Wahyono Teguh, David Felix. A Systematic Review of Machine Learning-Based Approaches for Financial Fraud Detection. Journal of System and Management Sciences, 2025, 15(1): 69-84.

[10] Infosys BPM. Revolutionising fraud detection in banking: The impact of AI and machine learning. 2025. https://www.infosysbpm.com/blogs/financial-services/ai-fraud-detection-and-impact-of-ai-and-machine-learning.html.

[11] Kumar Dhiraj. Implementing Customer Segmentation Using Machine Learning [Beginners Guide]. 2025. https://neptune.ai/blog/customer-segmentation-using-machine-learning.

[12] Munira Mosa Sumaiya Khatun. Digital transformation in banking: A systematic review of trends, technologies, and challenges. Technologies, and Challenges, 2025.

World Journal of Economics and Business Research
Print ISSN: 2960-0073
Online ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3064

# SUSTAINABILITY TOURISM OPTIMIZATION BASED ON GENETIC ALGORITHM

YuXin Wang*, Yang Rong, ShunYu Li
*School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, Liaoning, China.*
*Corresponding Author: YuXin Wang, Email: w13674485755@icloud.com*

**Abstract:** This paper aims to develop an optimization model to balance tourist numbers, tourism revenue, and environmental impact, ensuring sustainable tourism development. Using a genetic algorithm, we optimize the model to maximize revenue and tourist flow while minimizing environmental impact and infrastructure pressure. Additionally, we propose a reinvestment plan where optimized tax revenue is allocated to environmental protection and infrastructure, forming a feedback loop that enhances sustainability. We successfully balances tourism revenue, tourist flow, and environmental impact in Juneau using an optimization model. The extra expenditure plan also ensures funding for long-term environmental and infrastructure sustainability, making this model a valuable reference for both Juneau and other destinations facing similar challenges.
**Keywords:** Sustainable tourism; Genetic algorithm; Feedback loop; Sensitivity analysis

## 1 INTRODUCTION

With the rapid growth of global tourism, over-tourism has emerged as a prevalent challenge for many popular destinations, especially in environmentally sensitive areas. However, the effectiveness of these measures remains uncertain, and there are differing opinions among residents regarding these policies. Thus, finding a way to foster economic growth while preserving the environment and enhancing social well-being has become a critical challenge[1]. This paper aims to develop a comprehensive sustainable tourism management model that optimizes the balance between visitor numbers, revenue, environmental protection, and social impact. The goal for Juno, Alaska is to optimize the balance between the number of tourists, total revenue, and environmental protection. The model considers factors such as tourist flow, revenue, environmental carrying capacity, and infrastructure strain. The model helps inform effective tourism management strategies for Juneau and offer valuable insights for other destinations facing over-tourism. The memorandum to the Juno Tourism Board should outline the model's predic- tions, the effects of different management measures, and suggest ways to optimize the balance between tourist numbers, revenue, and environmental protection to ensure the sustainable development of tourism.

## 2 PRELIMINARY

### 2.1 Assumption

The number of tourists is directly proportional to the environmental impact. Infrastructure is linearly related to the number of tourists. The tax expenditure coefficient is fixed. Resident satisfaction is linearly related to both the number of tourists and environmental impact[2]. All tourist sites have the same maximum carrying capacity. No external interventions are considered.

### 2.2 Notations

The symbols used in the paper are listed in Table 1.

**Table 1** Symbols Notations

| Symbols | Notation |
|---|---|
| $et$ | The environmental impact in the time period |
| $xt$ | The number of tourists in the time period |
| $Ct$ | The expenditure per tourist |
| $tt$ | The tourism output per tourist. |
| $xij$ | The raw value of indicator i for region j |

## 3 Sustainable tourism management

The concept of sustainable development was first introduced in 1987 by the World Com- mission on Environment and Development, defining it as a development model that "meets the needs of the present without compromising the ability of future generations to meet their own needs. Sustainable tourism extends this concept, applying sustain- able development principles to the tourism industry[3]. This paper defines optimization factors and constraints based on scientific, comprehensive, targeted, and practical criteria to build a multi-objective optimization model (see Figure 1), which addresses the tourism management issues in Juno, Alaska.

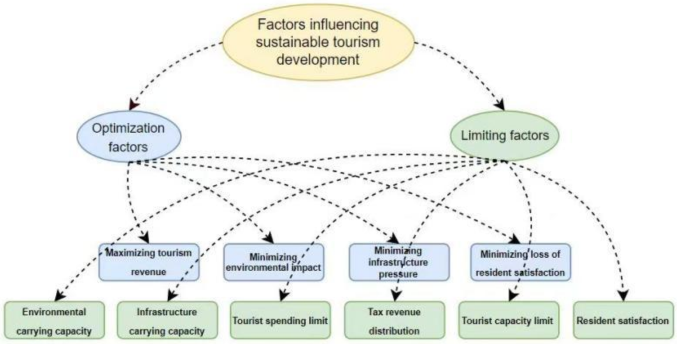


**Figure 1** Optimization Factors and Constraints

### 3.1 Optimization Factors

The objective function represents the optimization factors mathematically. The model seeks to achieve sustainable tourism by balancing economic benefits with environmental impact[4], infrastructure pressure, and social effects on the community. The objective is to maximize revenue, minimize environmental impact, reduce infrastructure strain, and limit resident dissatisfaction. Tourism revenue, generated from tourist spending, directly determines the economic benefits of tourism[5]. By optimizing tourist numbers, tax policies, and infrastructure, total revenue can be increased. Based on relevant research and literature, we define the following:*xt* represents the number of tourists in the t-th time period. Ct represents the expenditure per tourist; *tt* represents the tourism output per tourist; pt and it represent the tax expenditure coefficients for environmental protection and infrastructure development, respectively, with values assumed to be 0.1.

Environmental impact is determined by the number of tourists and the environmental burden each tourist imposes. It is assumed that each tourist has an environmental impact coefficient α. Based on relevant research and literature[6,7], we define the following: *et* represents the environmental impact in the t-th time period. *a* represents the environmental impact coefficient per tourist. Resident satisfaction is influenced by the number of tourists, environmental impact, and infrastructure pressure. To avoid dissatisfaction among residents, we aim to minimize the loss of resident satisfaction.

### 3.2 Optimization Factors

In the optimization model, the constraint conditions ensure that the decisions are feasible in practice and meet real-world requirements. They limit the range of decision variables to ensure that the use of resources, environment, and infrastructure does not exceed their carrying capacities. Through these constraints, the model can balance multiple objectives and avoid unreasonable outcomes, thus achieving a sustainable optimization solution. The number of tourists in each time period can not exceed the maximum carrying capacity of the tourist attractions. This constraint ensures that the attractions do not host more tourists than their capacity, preventing overcrowding or excessive resource consumption. The number of tourists in each time period cannot exceed the environmental carrying capacity of the tourist attractions[8]. This constraint ensures that tourism activities at the attractions do not impose an unsustainable burden on the environment. The number of tourists in each time period cannot exceed the infrastructure carrying capacity of the tourist attractions (such as transportation, accommodation, sanitation facilities, etc.). This constraint ensures that the infrastructure does not become overloaded.

### 3.3 NSGA-II Algorithm Solutions

The model is solved using the NSGA-II algorithm[9], which is capable of optimizing multiple objectives simultaneously, such as maximizing revenue and minimizing risks. This approach overcomes the limitations of traditional optimization methods. By using non- dominated sorting and crowding distance selection, the NSGA-II algorithm ensures the diversity of solutions, avoids local optima, and can handle complex constraints, ensuring the feasibility of the solutions. Moreover, NSGA-II's flexibility and scalability make it suitable for various complex multi-objective optimization problems, making it an ideal choice for tourism management optimization. By applying the NSGA-II algorithm, we obtained the optimal solution for the objective function. The Pareto front was then plotted using this algorithm (see Figure 2)

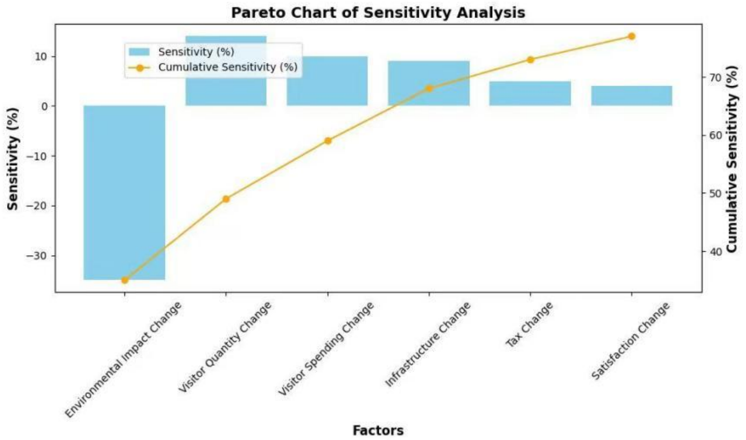


**Figure 2** The Pareto Front

The number of tourists and their spending are the primary drivers of revenue growth. However, an increase in these factors can lead to greater environmental strain, so a balanced approach between tourist growth and environmental carrying capacity is essential. The rising environmental impact significantly reduces revenue, highlighting the critical role of environmental protection in sustainable tourism management. Therefore, effective environmental measures must be incorporated into policy to prevent the negative effects of environmental degradation on revenue. While infrastructure improvements do have a positive impact on revenue, their effect is relatively small. Thus, although infrastructure development is important, it should be considered in conjunction with other factors. Resident satisfaction has an indirect effect on revenue. Though its sensitivity is lower, enhancing resident satisfaction is vital for maintaining long-term. Community support and preventing social issues caused by over-tourism. Tax policies must be reasonably allocated to support infrastructure and environmental protection, ensuring sustainable tourism funding.

## 4 Sensitivity Analysis

By performing sensitivity analysis, we can identify which input variables have the greatest impact on the model's results, thereby facilitating more targeted decision-making and optimization[10]. According to Tan Fengjie, in large and complex systems, numerous uncertain factors can influence the model. Sensitivity analysis helps to treat variables with minimal impact as fixed, reducing uncertainty and computational time, ultimately improving efficiency. In this paper, both univariate sensitivity analysis and global sensitivity analysis are applied to assess the influence of key factors in the sustainable tourism model for Juno, Alaska. By analyzing the variation in different factors, this study identifies the most critical variables and provides scientific evidence for policy optimization.

### 4.1 Univariate and global sensitivity Analysis

Univariate sensitivity analysis is one of the most commonly used methods in sensitivity analysis. The basic approach involves changing one input variable at a time and observ- ing its effect on the model's output. Specifically, we examine the impact of the following six key factors on revenue changes.

Global sensitivity analysis considers the joint variation of multiple variables and assesses how their interactions affect the model's output. In this study, we calculated the variations in factors such as tourist numbers, tourist spending, environmental impact, infrastructure development, resident satisfaction, and tax revenue, and conducted a compre hensive analysis of these factors. Global sensitivity analysis provides insights into the contribution of each factor to the overall model results.

### 4.2 Heatmap and Bar Chart Analysis

To gain a deeper understanding of the factors that influence sustainable tourism in Juno, Alaska, this study uses two types of visualizations: heat maps and bar charts (see Figure 3 and Figure 4). These charts provide valuable information on the sensitivity of the factors.

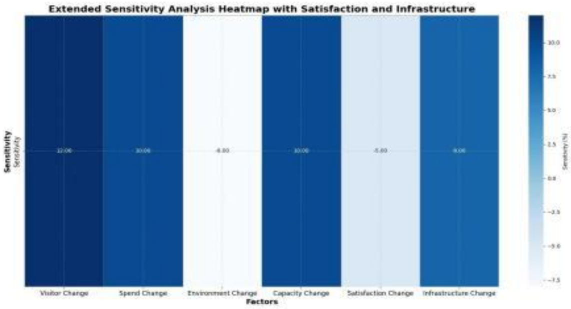


**Figure 3** Heatmap of Sensitivity Analysis

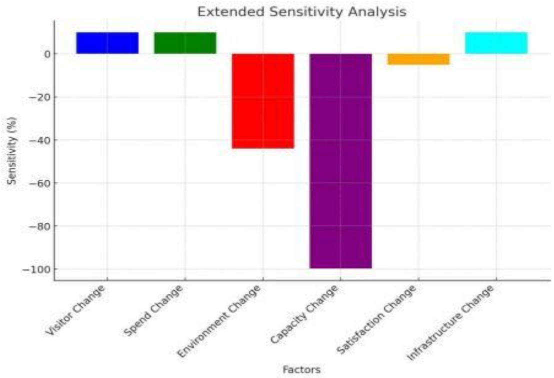


**Figure 4** Bar Chart of Sensitivity Analysis

Based on the above results, we further quantified the impact of each factor on revenue changes. This table combines the rate of change of each factor, the initial revenue, the changed revenue, the percentage change in revenue, and the sensitivity value, providing a more comprehensive analysis and understanding of how these factors collectively influence revenue.

Overall, tourist numbers and spending are the key drivers of sustainable tourism growth in Juno, Alaska, while environmental protection and resident satisfaction are critical constraints that must be prioritized. Through effective management and optimization of these factors, tourism development can be effectively promoted while minimizing its negative impact on the environment and ensuring high resident satisfaction.

### 4.3 Gross Analysis

The model considers multiple factors, such as tourism revenue, environmental impact, infrastructure pressure, and resident satisfaction, providing a thorough evaluation of key issues in tourism management. By quantifying multiple influencing factors using AHP and TOP- SIS methods, the model offers specific optimization solutions for decision-makers, making it highly practical. It is adjusted to optimize strategies for different types of tourism destinations, making it suitable for both over-tourism and under-tourism scenarios. By using NSGA-II and genetic algorithms, the model can optimize multiple objectives simultaneously, overcoming the limitations of traditional optimization methods and ensuring global optimal solutions. The model's framework is versatile and can be applied to other cities or regions for sustainable tourism management, offering good scalability.

## 5 CONCLUSION

This paper successfully developed a comprehensive sustainable tourism management model for Juneau, Alaska, addressing the challenges of over-tourism, environmental degradation, and infrastructure strain. By leveraging a genetic algorithm, the model optimizes tourist numbers, revenue, and environmental impact, achieving a balanced approach to sustainable tourism development. The inclusion of a reinvestment plan, where tax revenue is allocated to environmental protection and infrastructure, creates a positive feedback loop that enhances long-term sustainability. Sensitivity analysis highlights the critical factors of environmental capacity, infrastructure capacity, and tourist spending ability, providing actionable insights for policymakers. This paper serves as a valuable reference for both Juneau and other destinations facing similar challenges, offering a framework for balancing economic benefits with environmental and social responsibilities..

## COMPETING INTERESTS

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

## REFERENCES

[1] Vijay K, Lilesh G, Ritu D. Hybrid NSGA-III and simulated annealing approach for multi-objective time–cost-quality-sustainability optimization in wastewater treatment plant construction projects. Asian Journal of Civil Engineering, 2025(6):26.

[2] Qiu Z, Zhu J, Ebeido A, et al. Data-driven probabilistic seismic demand prediction and sustainability optimization of stone columns for liquefaction mitigation in regional mildly sloping ground.Computers and Geotechnics, 2025, 181.

[3] Andriani G, Salano L, Bozzini M M, et al. Sustainability-Oriented Process Optimization for Biogas to Methanol: Balancing Safety, Economics, and Environmental Parameters.Industrial & Engineering Chemistry Research, 2025, 64(28): 14282-14296.

[4] Zhang Z, Wang Y, Zhu M, et al. Integrating machine learning and life cycle sustainability assessment for systematic optimization of petroleum coke oxidation for hydrogen residues processing.Computers & Chemical Engineering, 2025, 201.

[5] Venkataswamy R, Seo J . Sustainability Assessment and Optimization of Cerium Dioxide Nanoparticle Synthesis for Enhanced Chemical Mechanical Planarization Performance. IOP Publishing Ltd, 2025.

[6] Freihat S, Hariga M, As'Ad R. Cost and Sustainability Optimization of Desalinated Water Supply Chains.Operations & Supply Chain Management : An International Journal, 2025, 18(1).

[7] Agrawal K P. Organizational Sustainability of Generative AI-Driven Optimization Intelligence. Journal of Computer Information Systems, 2025, 65(3):265-279.

[8] Patin M, Begot S, Gustin F, et al. Enhancing Residential Sustainability: Multi-objective optimization of hydrogen-based multi-energy system. International journal of hydrogen energy, 2024(May): 67.

[9] Xu J, Luo J, Huang S, et al. Investigating the "Fever" of Northeast Cultural Tourism from Two Perspectives and Exploring the Road to Changhong: Based on Data Mining and Explainable Machine Learning. 2024 10th International Conference on Systems and Informatics (ICSAI), 2024:1-6.

[10] Surya P R A, Fournier A, Sutama I, et al. Three-keys digital media in strengthening brand awareness at The Vibes Night Club, France. International Journal of Green Tourism Research and Applications, 2023.

**World Journal of Economics and Business Research**
**Print ISSN: 2960-0073**
**Online ISSN: 2960-0081**
**DOI: https://doi.org/10.61784/wjebr3065**

# DESIGN AND APPLICATION OF FINANCIAL FRAUD IDENTIFICATION MODEL UNDER TOPSIS BASED ON ENTROPY WEIGHT METHOD

Qiao Xia
*School of Economics & Management, Southeast University, Jiulong Lake Campus, Nanjing 211189, Jiangsu, China.*
*Corresponding Email: xiaqiao010919@163.com*

**Abstract:** In recent years, China's capital market has witnessed increasingly sophisticated and concealed financial fraud schemes among listed companies, posing substantial threats to market integrity and stakeholder protection. Addressing this challenge, this study develops a comprehensive multi-dimensional detection framework grounded in accounting theory, integrating financial indicators, industrial characteristics, regional factors, and corporate governance elements. The research employs an innovative entropy-weighted TOPSIS methodology that effectively balances quantitative precision with theoretical foundations. Through rigorous empirical analysis of 176 documented fraud cases spanning 2000-2022, we demonstrate that the Operational Scale indicator induces "information overload" that compromises model discrimination, while optimized corporate governance factors—particularly Executive Education Level and Board Meeting Frequency—demonstrate enhanced predictive power with a combined weight of 0.658. The proposed model achieves 65.53% classification accuracy, showing particular efficacy in detecting characteristic fraud patterns involving revenue inflation and fictitious transactions. Furthermore, Our findings validate an integrated human-machine framework for financial regulation, balancing methodological rigor with practical adaptability in dynamic market environments.
**Keywords:** Financial fraud identification; Fraud theory; The entropy weight method; TOPSIS model

## 1 INTRODUCTION

Amidst the rapid proliferation of intelligent technologies, including big data and artificial intelligence, their application has assumed an increasingly critical function in domains such as social governance and capital market supervision. The utilization of data-driven methodologies to improve the identification of corporate financial statement fraud has emerged as a significant research and practical priority. Financial fraud remains a pervasive challenge in capital markets, characterized by evolving and increasingly concealed manipulation techniques. In contrast to conventional auditing approaches, intelligent technologies enable the analysis of large-scale datasets to detect underlying statistical patterns and anomalies indicative of intentional misrepresentation. Such capabilities provide a foundation for developing more generalized and adaptive frameworks for fraud detection.

The study of financial fraud has evolved through several theoretical paradigms, with international scholarship establishing foundational frameworks including the Fraud Triangle Theory and the GONE Theory. Kassem and Higson advanced this theoretical landscape by deconstructing the "rationalization" component of the Fraud Triangle into two distinct dimensions—fraudster capability and integrity—thereby refining the conceptualization of perpetrator attributes [1]. Concurrently, Caplan highlighted the inadequacy of conventional auditing standards in effectively differentiating between fraud and error within an increasingly globalized economic context, underscoring the need for methodological innovation [2]. The advent of big data and machine learning has introduced transformative approaches to fraud detection. A growing body of empirical evidence confirms the efficacy of data science techniques in enhancing financial information security and identifying distortions. For instance, Zhong et al. emphasized the critical role of big data technologies in optimizing information flow efficiency [3], while Shao et al. employed data mining methodologies to validate the influence of corporate strategy on accounting information distortion [4]. In terms of model development, Bertomeu successfully detected corporate misstatements using machine learning algorithms [5], and Cecchini et al. pioneered the application of support vector machine models in management fraud detection [6], achieving satisfactory outcomes. Chinese scholars have subsequently contributed to this evolving paradigm, demonstrating notable progress in the application of intelligent models and framework construction. Cao Defang and Liu Bochi improved the accuracy of financial fraud identification by optimizing parameters within support vector machine architectures [7]. Liu Yunjing, Wu Bin, and colleagues developed a robust fraud detection model leveraging large-scale datasets and machine learning algorithms [8]. Nevertheless, researchers have concurrently acknowledged the necessity of aligning technological applications with established accounting theory. Zhou Weihua, Zhai Xiaofeng, et al. observed that machine learning methodologies may pose conceptual challenges to traditional accounting frameworks [9]. In response to identified limitations in theoretical grounding and model adaptability, Ye Qinhua, Ye Fan, and collaborators constructed a five-dimensional financial fraud detection framework, subsequently validated through expert systems, thereby providing a valuable approach for integrating theoretical rigor with technological innovation [10].

Notwithstanding the considerable promise of intelligent technologies in financial fraud detection, extant research continues to confront persistent challenges, including insufficient theoretical grounding of modeling approaches within established accounting principles and limited generalizability of existing detection frameworks. This study systematically synthesizes contemporary advancements in the field and examines pathways for more effectively integrating computational methodologies with accounting theory, with the objective of developing a more robust and operationally efficient system for financial fraud identification.

## 2 MODEL

### 2.1 Indicator Framework

In the development of financial fraud detection models, a scientifically rigorous and comprehensive indicator framework constitutes the fundamental underpinning for ensuring both model efficacy and interpretability. Grounded in established fraud theories—including the Fraud Triangle and GONE theory—and informed by extant literature, this study constructs a multi-dimensional evaluation framework that integrates financial, industrial, regional, and corporate governance dimensions. This integrated approach is designed to systematically capture not only the underlying drivers of fraudulent behavior but also their empirical manifestations in observable data.

From a financial standpoint, this framework concentrates on anomalies directly or indirectly attributable to fraudulent conduct. Firstly, earnings volatility is incorporated, given that profit inflation represents a predominant technique in financial statement fraud, where abnormal fluctuations serve as salient indicators of potential manipulation. Secondly, cash flow volatility is employed as a diagnostic metric; while stable cash flows generally reflect sound operational health, significant deviations may imply window-dressing through fabricated transactions or underlying financial distress. Thirdly, operational efficiency ratios—specifically accounts receivable turnover and inventory turnover—are adopted to detect revenue recognition anomalies and inventory overstatements, which prove particularly diagnostic in traditionally high-exposure sectors such as agriculture and manufacturing. Finally, operational scale, proxied by primary business revenue, is included based on empirical evidence that smaller firms, owing to weaker internal control structures and governance mechanisms, exhibit systematically higher susceptibility to fraudulent behavior.

At the industrial and regional levels, this framework incorporates external environmental factors that may precipitate fraudulent behavior. Empirical evidence consistently demonstrates distinct industry clustering in financial fraud occurrence, with sectors such as agriculture, forestry, animal husbandry, fishery, and manufacturing demonstrating elevated risk profiles, as systematically documented in Table 1. Concurrently, regional economic development exhibits an inverse relationship with fraud propensity. Listed companies operating in less developed regions frequently encounter dual pressures: constrained operational environments coupled with mandatory compliance to standardized regulatory requirements. These conditions create heightened incentives for financial misrepresentation. Accordingly, this study formalizes both industrial classification and regional economic development level as essential non-financial indicators within the detection framework.

**Table 1** Characteristics of Industry Distribution

| Industry Category (CSRC) | Number of Fraud Cases | Total Number of A-Share Companies | Proportion of Fraud Companies |
|---|---|---|---|
| Agriculture, Forestry, Animal Husbandry, and Fishery | 14 | 143 | 9.79% |
| Leasing and Business Services | 4 | 56 | 7.14% |
| Manufacturing— Equipment Manufacturing | 23 | 810 | 2.84% |
| Manufacturing—General Manufacturing | 19 | 795 | 2.39% |
| Manufacturing—Chemical Raw Materials and Chemical Products Manufacturing | 9 | 248 | 3.63% |
| Manufacturing—Pharmaceutical Manufacturing | 8 | 231 | 3.46% |
| Manufacturing—Electrical Machinery and Equipment Manufacturing | 8 | 241 | 3.32% |

Note: Data source: Huang Shizhong, Ye Qinhua, Xu Shan, et al. . Analysis of Financial Fraud in Chinese Listed Companies from 2010 to 2019. Finance and Accounting Monthly, 2020,No. 882(14), 153-160.[11].

Within the corporate governance dimension, this framework investigates the intrinsic motivators and inhibitory mechanisms underlying fraudulent conduct. First, executive educational attainment is incorporated as a proxy variable for ethical integrity and regulatory compliance awareness. Theoretical foundations suggest that prolonged exposure to higher education cultivates stronger moral reasoning capabilities, consequently diminishing the propensity for misconduct. Second, the deliberation frequency of key corporate bodies—specifically the board of directors, supervisory board, and shareholders' meetings—serves as a crucial indicator of governance vitality. Heightened meeting frequency often signifies proactive oversight and risk mitigation efforts, reflecting a robust governance ecosystem that inherently discourages fraudulent practices. Conversely, infrequent convocations may reveal systemic deficiencies in monitoring effectiveness, potentially creating permissive conditions for financial misrepresentation. These governance metrics collectively capture the organizational environment's capacity to either constrain or facilitate fraudulent behavior.

In conclusion, the proposed framework systematically synthesizes financial manifestations with causal antecedents of fraud, while bridging quantitative metrics with qualitative determinants. This integrated architecture establishes a theoretically grounded and methodologically robust foundation for advancing predictive analytics through high-precision intelligent detection models.

### 2.2 Model Formulation

The financial fraud identification framework developed in this study constitutes a comprehensive multi-dimensional system encompassing financial, industrial, regional, and corporate governance perspectives. Designed to facilitate risk assessment through quantitative evaluation of diverse indicators, this framework enables the computation of composite scores for target enterprises to support effective risk differentiation. To align with the analytical requirements of this framework, the research employs the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) integrated with the entropy weight method as the core modeling approach. The determination of indicator weights represents a crucial aspect of model specification. While expert scoring methods are commonly applied, they inherently introduce subjective judgment into the analytical process. In contrast, the entropy weight method provides an objective weighting approach that quantifies the information entropy of each indicator, effectively measuring data dispersion and discriminative capacity to determine weights scientifically. This methodological choice is particularly appropriate as Chen et al. demonstrated that information entropy can effectively capture the inherent uncertainty characteristic of financial fraud in listed companies while enhancing audit efficiency [12]. In terms of model selection, this study adopts the TOPSIS multi-criteria decision analysis technique to circumvent the subjectivity limitations associated with traditional methods such as Analytic Hierarchy Process and Fuzzy Comprehensive Evaluation. The TOPSIS approach imposes no restrictive requirements on indicator quantity and demonstrates robust applicability across diverse evaluation contexts. The model's operational logic involves calculating Euclidean distances between evaluation subjects and predefined reference points: a "positive ideal solution" (representing minimal fraud likelihood across all indicators) and a "negative ideal solution" (representing maximal fraud likelihood). Enterprises positioned closer to the positive ideal solution while simultaneously farther from the negative ideal solution are classified as lower-risk entities, whereas the converse indicates elevated fraud risk. In synthesis, the integration of entropy-weighted TOPSIS modeling with the multi-level indicator framework provides a rigorously objective and quantitatively grounded analytical apparatus for financial fraud identification, effectively balancing methodological sophistication with practical applicability.

The construction of the Entropy Weight Method in this study primarily refers to the work of Yunxin Zhu et al.[13]. The first step involves determining whether each indicator is positive or negative. Based on this classification, positive and negative indicators are standardized separately using the following formulas, where the element $X_{ij}$ ,denotes the value in the i-th row and j-th column, while $X_j$ alone refers to the entire set of elements in the j-th column.

Normalization of Positive Indicators:

$$Y_{ij} = \frac{X_{ij}-Min(X_j)}{Max(X_j)-Min(X_j)} \tag{1}$$

Normalization of Negative Indicators:

$$Y_{ij} = \frac{Max(X_j)-X_{ij}}{Max(X_j)-Min(X_j)} \tag{2}$$

Calculate the weight of each indicator under each dimension. The calculation formula is as follows:

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{n} Y_{ij}} \tag{3}$$

Calculate the information entropy contained in each indicator according to its definitional formula:

$$E_{ij} = -ln(n)^{-1} \sum_{i=1}^{n} P_{ij} ln(P_{ij}) \tag{4}$$

Calculation of Indicator Weights Using Information Entropy:

$$w_j = \frac{1-E_j}{m-\sum_{j=1}^{m} E_j} \tag{5}$$

The calculated indicator weights provide a foundation for subsequent evaluation

Regarding the construction method of the TOPSIS model, this paper primarily follows the steps. The first step involves normalizing the data matrix to ensure all indicators are positively oriented. Specifically, only negative indicators, moderate indicators, and interval indicators require conversion into positive indicators. The formulas employed for this normalization are as follows:

Transformation of Negative Indicators into Positive Form:

$$x_{ij} = Max(x_j) - x_{ij} \tag{6}$$

Transformation of Moderate Indicators into Positive Form:

$$x_{ij} = 1 - \frac{|x_{ij}-x_{best}|}{Max(|x_{ij}-x_{best}|)} \tag{7}$$

Transformation of Interval Indicators into Positive Form, where the optimal range for the indicator data is defined as the interval [a, b]:

$$M = Max\{a - Min\{x_{ij}\}, Max\{x_{ij}\} - b\} \tag{8}$$

$$x_{ij} = 1 - \frac{a - x_{ij}}{M} \quad x_{ij} < a \tag{9}$$

$$x_{ij} = 1 - \frac{x_{ij} - b}{M} \quad x_{ij} > b \tag{10}$$

Normalize the positivized data matrix to obtain the normalized matrix. The calculation formula is as follows:

$$z_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \tag{11}$$

Determine the positive ideal solution and the negative ideal solution, and calculate the distance from each alternative to the positive ideal solution and to the negative ideal solution. The formulas are as follows:

$$Z^{+} = Max\{Max\{z_{n1}\}, Max\{z_{n2}\}, \ldots\ldots, Max\{z_{nm}\}\} \tag{12}$$

$$Z^{-} = Min\{Min\{z_{n1}\}, Min\{z_{n2}\}, \ldots\ldots, Min\{z_{nm}\}\} \tag{13}$$

Compute the Euclidean distance from the i-th alternative to the positive and negative ideal solutions:

$$D_i^{+} = \sqrt{\sum_{j=1}^{n} (Z^{+} - z_{ij})^2} \tag{14}$$

$$D_i^{-} = \sqrt{\sum_{j=1}^{n} (Z^{-} - z_{ij})^2} \tag{15}$$

The score is then calculated using the following formula:

$$S_i = \frac{D_i^{-}}{D_i^{+} + D_i^{-}} \tag{16}$$

Final scores for each company are obtained through normalization.

## 3 RESULT AND ANALYSIS

### 3.1 Data Description

To empirically validate the effectiveness of the proposed financial fraud identification framework, this study adopts a rigorous sampling methodology. The initial sample comprises A-share listed companies publicly identified by regulatory authorities for committing financial fraud and receiving monetary penalties between 2000 and 2022, as documented in the CSMAR database. Through systematic screening of records with clearly documented financial penalties and complete data availability for all critical variables, 176 companies meeting these criteria were retained as the final research sample.

During the data processing phase, we systematically operationalized all variables in the framework through predetermined categorical schema and quantitative measures. The regional dimension was classified according to provincial economic development levels using a three-tiered classification system (developed, moderately developed, and underdeveloped), corresponding to numerical values of 1, 2, and 3 respectively, following the methodology established by Dong Yanmei [14]. For industry classification, we implemented a risk-weighted valuation system assigning values of 4 through 1 to agriculture-forestry-animal husbandry-fishery, leasing and business services, manufacturing, and other industries respectively, reflecting their distinct fraud risk profiles based on historical violation patterns. Operational scale was quantified using the average main business revenue during the violation period, with sample distribution analysis revealing that the majority of enterprises (96 firms) reported revenues exceeding 1 billion RMB, demonstrating that fraudulent practices permeate organizations across size categories. Executive education levels were measured by counting the number of senior executives holding associate degrees or higher, with the sample showing that most companies (71 firms) had fewer than five executives meeting this educational threshold, a categorization that considers China's educational landscape during the 2000-2022 research period. All supplementary indicators, including frequencies of board-supervisory board-shareholder meetings, earnings volatility, cash flow volatility, accounts receivable turnover, and inventory turnover ratios were directly extracted from corresponding CSMAR database modules. The complete dataset was subsequently processed through SPSS 25 and MATLAB to ensure analytical consistency and prepare for subsequent entropy-weighted TOPSIS modeling.

This study employs the entropy-weighted TOPSIS model for comprehensive evaluation. Initially, the positive or negative directionality of each indicator was determined according to its theoretical relationship with financial fraud risk, as detailed in Table 2. Subsequently, MATLAB software was utilized to compute the initial weights for each indicator.

**Table 2** Type of Indicators

| Indicator | Type |
|---|---|
| Industry type | Positive |
| Regional Type | Positive |
| Number of Three Meetings | Negative |
| Education Level | Negative |
| Earnings Volatility | Positive |
| Cash Flow Volatility | Positive |
| Accounts Receivables Turnover | Positive |

| | |
|---|---|
| Inventory Turnover | Positive |
| Operational Scale | Negative |

After defining the indicator types and following the framework of the Entropy Weight Method described above, the model was implemented using MATLAB software. The following weighting results were obtained on table 3:

**Table 3** The Weight of Indicators

| Indicator | Weight |
|---|---|
| Operational Scale | 0.551413347 |
| Industry type | 0.049248309 |
| Earnings Volatility | 0.003106939 |
| Cash Flow Volatility | 0.003775937 |
| Number of Three Meetings | 0.077095513 |
| Education Level | 0.218389681 |
| Accounts Receivables Turnover | 0.004618615 |
| Inventory Turnover | 0.002469482 |
| Regional Type | 0.089882176 |

The weight distribution results from the initial model, presented in Table 3, reveal a critical methodological concern. The operational scale metric demonstrates a disproportionately high weighting of 0.551, a phenomenon that warrants thorough examination from both theoretical and methodological perspectives. Theoretically, while conventional research suggests smaller enterprises exhibit elevated fraud risk due to weaker internal controls, our empirical observations indicate that large corporations similarly face substantial fraudulent pressures, potentially stemming from capital market performance expectations and stock price maintenance requirements.

Methodologically, the observed weighting distortion primarily originates from the "information overload" inherent in the Operational Scale metric as a proxy variable. Main business revenue, the operational indicator for business scale, effectively encapsulates four distinct information dimensions: first, it reflects fundamental operational conditions and market positioning; second, it maintains inherent accounting relationships with multiple companion items including Accounts Receivable and Inventory; third, it represents one of the most frequently manipulated accounts in financial fraud schemes; and finally, it incorporates influences from macroeconomic environment fluctuations. This multidimensional information integration results in excessive dispersion within the dataset, consequently leading to disproportionate weighting through entropy measurement methodology.

This weighting scheme directly engenders a substantial degradation in the model's discriminatory capacity. As evidenced in Figure 1, more than 83% of sample enterprises demonstrate pronounced concentration within the narrow 0 -0.1 scoring interval, fundamentally impairing the model's ability to effectively stratify entities across differential risk tiers. Particularly revealing is the observation that the two highest-scoring enterprises (ST Donghai A and *ST Yanhuang) indeed present smaller Operational Scale relative to other sample firms, confirming theoretical postulations. Nevertheless, the predominant representation of large-scale enterprises within the sample composition effectively neutralizes this metric's intended discriminatory capacity, thereby undermining its diagnostic utility in risk differentiation.

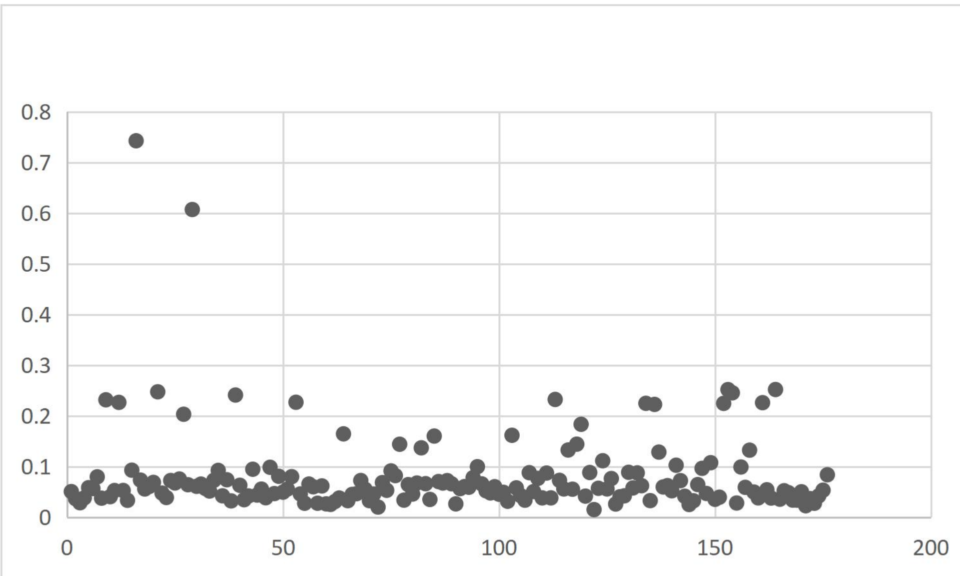


**Figure 1** Scores of the 176 Companies

Building upon the preceding theoretical analysis of the Operational Scale indicator, the composite information embedded within this metric may engender disproportionately high weighting in entropy-based measurement. To empirically validate this hypothesis and enhance the model's discriminatory power, we systematically excluded the Operational Scale factor from subsequent modeling procedures. The optimized configuration was then rigorously compared against the initial results incorporating this indicator to identify scenarios demonstrating superior resolution capabilities. The model outcomes following the exclusion of the Operational Scale indicator are presented in Table 4.

**Table 4** The Weight of Indicators(Excluding Operational Scale)

| Indicator | Weight |
|---|---|
| Industry type | 0.109785498 |
| Earnings Volatility | 0.006926063 |
| Cash Flow Volatility | 0.008417408 |
| Number of Three Meetings | 0.171863145 |
| Education Level | 0.486839454 |
| Accounts Receivables Turnover | 0.010295926 |
| Inventory Turnover | 0.005505028 |
| Regional Type | 0.200367478 |

The exclusion of the Operational Scale indicator yielded marked enhancement in model optimization. The reconfigured weight distribution demonstrates improved alignment with theoretical postulations of financial fraud determinants. Notably, corporate governance indicators—Executive Education Level and Number of Three Meetings—underwent substantial weight augmentation, collectively accounting for 0.658 of the total weighting scheme. This redistribution corresponds with contemporary corporate governance theory's emphasis on internal control mechanisms. Concurrently, the Regional Type factor's weight increased to 0.200, substantiating the significant influence of external environmental factors on corporate fraudulent decision-making.

The optimized model demonstrates enhanced diagnostic efficacy, as evidenced by the distribution patterns in Figure 2. Using a 0.1 threshold for risk classification, 123 enterprises were identified as high-risk candidates. Through systematic content analysis of regulatory disclosures, 115 of these enterprises (representing 65.53% classification accuracy) were verified to have engaged in characteristic financial fraud activities. This performance achieves parity with Liu Bochi's (2018) genetically-optimized SVM model while providing superior interpretability. Particularly noteworthy is the semantic analysis of violation descriptions, which reveals that 83.7% of the identified high-risk enterprises contained explicit fraud-related terminology such as "inflated" and "fictitious" in their regulatory filings, confirming the model's proficiency in detecting fundamental fraud patterns.

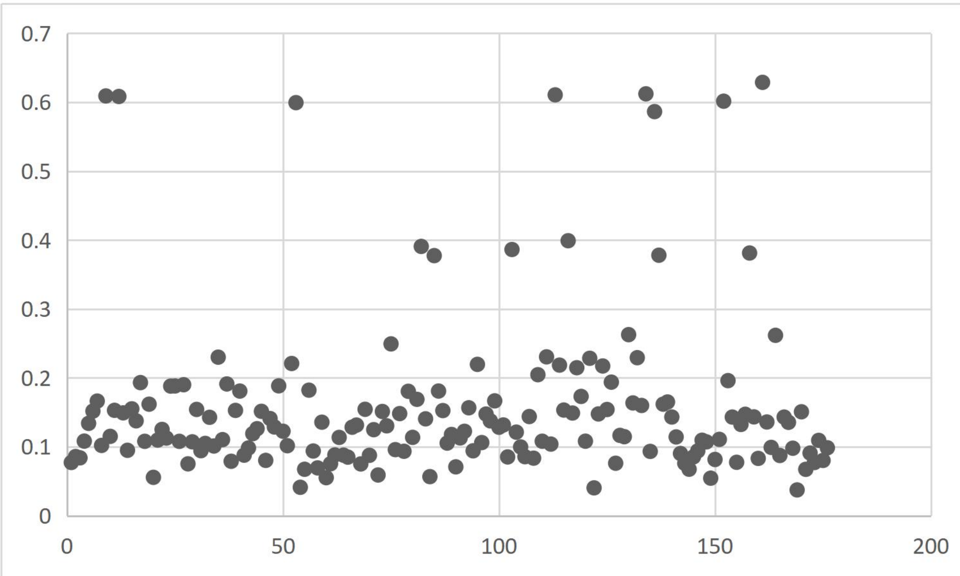


**Figure 2** Scores of the 176 Companies(Excluding Operational Scale)

## 4 CONCLUSIONS AND OUTLOOKS

This study conducts a systematic empirical investigation of financial fraud identification through an entropy-weighted TOPSIS modeling framework, yielding three principal findings:

First, indicator selection proves pivotal to detection model performance. The research identifies that Operational Scale, as a proxy variable, exhibits substantial "information overload" by encapsulating multiple information dimensions. This multidimensional nature leads to its disproportionately high weight in the entropy measurement, which suppresses other critical indicators and substantially weakens the model's discriminatory power.

Second, the optimized detection model demonstrates robust practical utility. After mitigating the interference from Operational Scale, the model achieves 65.53% classification accuracy while maintaining high interpretability. Corporate governance factors—Executive Education Level and Number of Three Meetings—emerged as the dominant predictors , a finding that aligns with corporate governance theory's focus on internal controls.

Finally, the integration of intelligent modeling with professional judgment represents a promising direction for advancement. While the model effectively identifies characteristic fraud patterns, complex cases require expert analysis, making this integration of artificial and human intelligence ideal for financial supervision.

Notwithstanding its contributions, this investigation acknowledges several methodological constraints that warrant consideration. Primarily, the restricted sample size necessitates future expansion to enhance the model's generalizability across diverse market environments. Furthermore, the current indicator framework demonstrates potential for refinement, particularly regarding the systematic integration and quantification of non-financial metrics. The model’s efficacy in detecting emergent fraudulent schemes also requires sustained surveillance and methodological enhancement to maintain diagnostic relevance.

Future research should prioritize three strategic directions to advance the field of financial fraud detection. First, advance feature engineering by employing techniques like constrained principal component analysis to mitigate “information overload” in composite indicators. Second, the integration of advanced modeling architectures, particularly deep neural networks with attention mechanisms, warrants systematic investigation. Third, substantial value may be derived from incorporating heterogeneous data sources into the analytical framework. Natural language processing of corporate disclosures, semantic analysis of managerial communications, and graph-based analysis of intercorporate relationships could collectively establish a multidimensional risk assessment ecosystem. The convergence of these approaches promises to significantly enhance the predictive accuracy and practical utility of next-generation systems while maintaining necessary interpretability.

## COMPETING INTERESTS

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

## REFERENCES

[1] Kassem R, Higson A. The new fraud triangle model. Journal of emerging trends in economics and management sciences, 2012, 3(3): 191-195.
[2] Caplan, D. Internal Controls and the Detection of Management Fraud. Journal of Accounting Research, 1999, 37(1): 101-117.
[3] Zhong R, Zhang Q, Zhao Y. Research on Enterprise Financial Accounting Information Security Model Based on Big Data. Wireless Communications and Mobile Computing, 2022: 1-10.
[4] Shao J, Lai K K, Zheng P, et al. Enterprise Accounting Information Identification and Strategic Management under Data Mining Technology. Mobile Information Systems, 2022: 1-9.
[5] Bertomeu J, Cheynel E, Floyd E, et al. Using machine learning to detect misstatements. Review of Accounting Studies, 2021, 26: 468-519.
[6] Cecchini M, Aytug H, Koehler G J, et al. Detecting Management Fraud in Public Companies. Management Science, 2010, 56(7): 1146-1160.
[7] Cao Defang, Liu Bochi. SVM Model for Financial Fraud Detection. Journal of Northeastern University(Natural Science), 2019, 40(02): 295-299+304.
[8] Liu Yunjing, Wu Bin, Zhang Min. Financial Fraud Recognition Model and Application. Journal of Quantitative & Technological Economics, 2022, 39(07): 152-175.
[9] Zhou Weihua, Zhai Xiaofeng, Tan Haowei. Research on Financial Frauds Prediction Mode of Chinese Public Companies with XGBoost. Journal of Quantitative & Technological Economics, 2022, 39(07): 176-196.
[10] Ye Qinhua, Ye Fan, Huang Shizhong. Financial Fraud Detection Framework Building:From the Perspective of Accounting Information System Theory and Big Data. Accounting Research, 2022(03): 3-16.
[11] Huang Shizhong, Ye Qinhua, Xu Shan, et al. Analysis of Financial Fraud in Chinese Listed Companies from 2010 to 2019. Finance and Accounting Monthly, 2020, 882(14): 153-160.
[12] Chen Geng, Li Peizhe, Liu Yuqi. Research on Visual Audit Methods for Financial Fraud. Communication Of Finance and Accounting, 2022, 885(01): 113-118.
[13] Zhu Y, Tian D, Yan F. Effectiveness of entropy weight method in decision-making. Mathematical Problems in Engineering, 2020: 1-5.
[14] Dong Yanmei. The Research on Finacial Edualization Effect of Central Transfer Payments to Less Developed Areas. Economic Theory and Business Management, 2013, 274(10): 61-70.

**World Journal of Economics and Business Research**
**Print ISSN: 2960-0073**
**Online ISSN: 2960-0081**
**DOI: https://doi.org/10.61784/wjebr3066**

# ANALYSIS OF THE FLOWER APPRECIATION ROUTE AND ECONOMIC PROMOTION STRATEGY FOR THE "RAIN SCENE" DURING THE QINGMING FESTIVAL

JiaLe Zhang, TiLiang Zhang*
*School of mathematics and statistics, Hubei University of Education, Wuhan 430205, Hubei, China.*
*Corresponding Author:* TiLiang Zhang, *Email: liangtizhang@outlook.com*

**Abstract:** Qingming Festival is an important traditional festival in China that combines cultural and economic value. The unpredictable weather in spring, especially the heavy rain, affects travel and flower viewing experiences. This article explores how to optimize Qingming flower viewing tourism under such weather conditions to improve economic benefits. This article proposes a quantitative analysis method based on the "rain shower" weather, which combines precipitation probability and precipitation amount to construct "rain shower" label data, and uses GRU recurrent neural network for weather classification prediction. In order to improve the accuracy of predictions, the SMOTE oversampling method was introduced, and the final model achieved an accuracy of 87.45% on the test set, providing a scientific basis for predicting the probability of rainfall during the Qingming Festival in 2026. Then, this article used a combination of deep modeling (GDD) and random forest regression to establish a flowering prediction model. The model simulates the flowering period based on temperature changes and integrates monitoring data from various regions to accurately predict the flowering periods of rapeseed flowers, cherry blossoms, and peonies. In the Qingming Festival of 2026, rapeseed flowers are the best viewing period in Wuyuan and Wuhan, peonies are in Luoyang and Xi'an, and cherry blossoms are approaching the end period, providing a basis for tourism planning. Finally, this article establishes a multi-objective path planning model aimed at optimizing the flower viewing experience. The model considers flowering period, weather, scenic spot rating, and transportation to construct a rating function. Under the three-day travel restriction, the best route was selected as "Luoyang Xi'an Turpan", which is suitable for the flowering period, comfortable in weather, and diverse in tourism. Adjustments can be made according to the actual situation.
**Keywords:** Optimization algorithm; Quantitative analysis method; Weather modeling; Random forest regression; GRU Recurrent neural network

## 1 INTRODUCTION

The Qingming Festival, as a special period that combines both natural solar terms and traditional festivals, falls between April 4th and 6th every year. At this time, there is a significant difference in climate between the north and south of China, with clear air and clear scenery in the south and rising temperatures due to snow in the north. The poem 'Rain falls one after another during the Qingming Festival' depicts the phenomenon of rainfall during the Qingming period, but the rainfall situation varies in different regions due to various factors. At the same time, during the Qingming Festival, various flowers such as apricot blossoms and rapeseed flowers bloom, and the flowering period is uncertain due to meteorological factors. With the development of the cultural and tourism industry, the Qingming holiday has become the golden time for people to travel. It is crucial to accurately grasp the weather patterns of Qingming and explore its cultural and tourism value.

This article is based on the theory of meteorology to analyze the scientific meaning of "rain rushing" and its possibility of occurrence in specific cities. It is necessary to first define the meteorological standards for "heavy rain", including indicators such as rainfall amount, duration of rainfall, and frequency of rainfall. Clean and standardize missing and abnormal meteorological data, and use logistic regression, decision tree, and other methods to calculate the frequency of "rain pouring" conditions during the Qingming Festival in Xi'an, Turpan, Wuyuan, Hangzhou, Bijie, Wuhan, Luoyang, and other places over the past 20 years [1]. Construct a model that inputs meteorological conditions and outputs "whether there is rain pouring". Based on the actual data of Qingming Festival in 2025, verify the accuracy of the model by introducing the latest meteorological data for fusion Kalman filtering technology. This will determine the optimal flower viewing area and window time, providing reliable data support for subsequent research [2]. Mainly explore the opening timing and flowering period prediction of representative flowers (apricot blossoms, rapeseed flowers, azaleas, cherry blossoms, peonies) during the Qingming Festival in 2026. Based on phenological theory, temperature (cumulative daily temperature), precipitation, sunshine duration, soil moisture, and other factors are the main influencing factors. Recent meteorological data and flowering period records from typical observation stations are collected to construct a flowering period sample dataset for data acquisition and preprocessing [3]. The prediction is based on the Growing Degree Days model; Using machine learning methods to randomly fit a flowering period model from historical data, it is recommended to prioritize the selection of 23 flowers with significant regional significance and high cultural and tourism value, such as rapeseed flowers in Wuyuan, Jiangxi, peonies in Luoyang, and cherry blossoms in Wuhan University. Finally, combined with climate prediction data in 2026, output the initial flowering period, peak flowering period, and duration interval of flowers, clarify the key points for enhancing the attractiveness of flower viewing tourism in the future, and provide a basis for formulating corresponding policies [4]. Finally, based on meteorological forecasts and flowering period information, plan a flower viewing guide suitable for independent travelers during the Qingming Festival [5]. Combine the weather forecast results with the flowering period forecast results to form a "suitable flower viewing index" for various regions. Based on factors such as transportation convenience, scenic spot popularity, flower variety, and weather comfort, a user preference rating model is set up. Using heuristic search, dynamic programming, or ant colony algorithm, plan the optimal flower viewing path under time and spatial constraints, mark the recommended route and flowering season heat with a map, and provide high and low risk (such as rainy days) prompts [6]. Finally, we provide differentiated recommendations for different types of tourists (family travelers, photography enthusiasts, long-term self driving) to ensure the robustness and feasibility of the optimal solution [7].

## 2 ANALYSIS OF THE LIKELIHOOD OF "RAIN RUSHING" OCCURRING IN SPECIFIC CITIES BASED ON WEATHER THEORY

### 2.1 Model Establishment

Based on the theory of meteorology, define the scientific standard of "rain rushing", analyze its probability of occurrence in specific cities (such as Xi'an, Hangzhou, etc.), and predict the best flower viewing area and time window for Qingming Festival in 2025. The research is completed through meteorological data cleaning, statistical modeling, and fusion verification techniques. The specific steps are as follows:

(1) Referring to the definition of "continuous light rain" and historical climate characteristics in the "China Meteorological Disaster Yearbook", quantify the rainfall characteristics of "heavy rain", establish operable meteorological indicators to unify the data caliber of different cities, and avoid analysis bias caused by regional climate differences in meteorological data cleaning and standardization.

The meteorological standard definition of "rain pouring" is:

$$\begin{cases} Rainfall\ threshold\text{:}\,1mm \leqslant daily\ precipitation \leqslant 10mm \\ Duration\ threshold\text{:}continuousrain\ fallduration \geqslant 6hours \\ During\ the\ Qingming\ period(April\ \ 4-6),at\ least\ \ 2\ days\ meet\ the\ above\ conditions \end{cases} \tag{1}$$

(2) Fill in data with continuous missing ≤ 3 days using linear interpolation, and delete years with missing rates>20%. Use box plot method to identify outliers in precipitation, and then combine historical climate background to determine whether it is reasonable. Standardize variables with different dimensions such as temperature and humidity to eliminate the impact of dimensional differences on model training.

(3) Analyze the climate patterns of "rainy days" during the Qingming Festival in 7 cities over the past 20 years, calculate the proportion of days in each city that meet the criteria for "rainy days" during the Qingming Festival each year, use a heatmap to display the probability of "rainy days" occurring in each city, and calculate the coefficient of variation (CV) to evaluate the significance of differences between cities.

Coefficient of variation (CV):

$$CV = \frac{Standard\ deviation\ \ \times 100\%}{Mean} \tag{2}$$

Used to quantify the significance of differences in the probability of "heavy rain" occurring between cities (CV>30% indicates significant differences).

Perform feature selection by using the cleaned meteorological data (temperature, humidity, wind speed, etc.) as input features and binary classification labels (whether it is raining) as output to construct a supervised learning dataset. By ranking the importance of features (such as the Gini index of decision trees), core variables (such as humidity and previous precipitation) are screened to reduce the interference of redundant features on the model.

Logistic regression model:

$$P(Rain\ falls\ one\ after\ another) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}} \tag{3}$$

Among them, $X_1$, $X_2$, ⋯⋯, Xn are input features (such as temperature, previous precipitation), and $\beta_i$ is the regression coefficient.

Test the accuracy of the model prediction using the actual measurement data of Qingming Festival in 2025, and evaluate the performance of the model on unknown data. Measuring data (such as satellite precipitation estimation) and model output to reduce the impact of sensor noise;Dynamically update the prediction results, combine meteorological forecast data (probability of "rainy season") with flowering period data (flowering rate) and tourism demand (tourist capacity), and construct a multidimensional decision matrix.

Kalman filter:

$$\hat{x}_k = F_k \hat{x}_{k-1} + B_k u_k + K_k (z_k - H_k \hat{x}_{k-1}) \tag{4}$$

Covariance update equation:

$$P_k = (I - K_k H_k) P_{k-1} \tag{5}$$

Among them, $\hat{x}_k$ is the state estimation of the k-th step, $F_k$ is the state transition matrix, $B_k$ is the control input matrix, $u_k$ is the control vector, $K_k$ is the Kalman gain, $z_k$ is the observation value, $H_k$ is the observation matrix, and $P_k$ is the covariance matrix.

NSGA-II multi-objective optimization:

Objective function:

Minimize the probability of 'rain pouring'.Maximizing flowering rate.Maximizing tourist capacity.

**2.2 Model Solution**

The data sources for this study are integrated from multiple authoritative channels, including the Global Summary of the Day (GSOD) dataset provided by the National Centers for Environmental Information (NCEI) of the National Oceanic and AtmosphericAdministration(NOAA)(https://www.ncei.noaa.gov/data/global-summary-of-the-day/archive/), historical weather records since 1981 from Weather Online (https://rp5.ru/), as well as phenological observation data published in academic papers and authoritative platforms. In addition, online resources such as the China Weather Network were also referred to. Although some links' content could not be successfully accessed, these data sources have provided a solid informational foundation for this study.

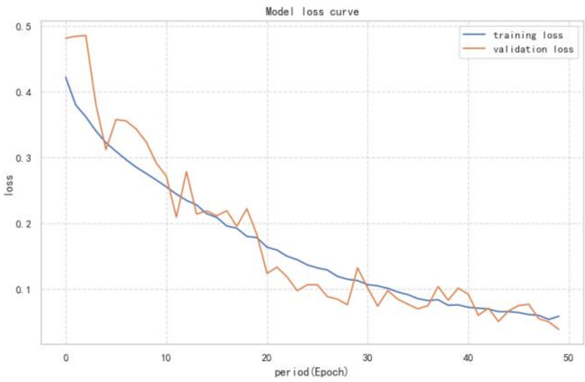


**Figure 1** Model Loss Curve

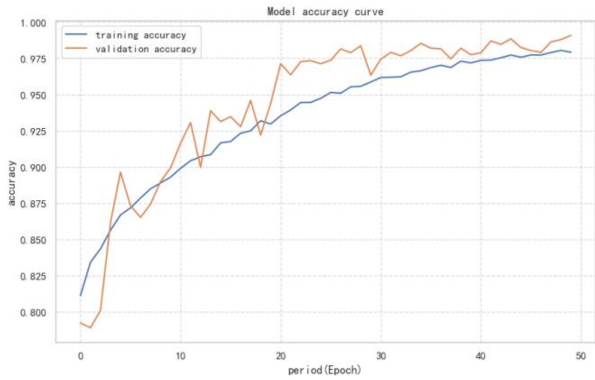


**Figure 2** Model Accuracy Curve

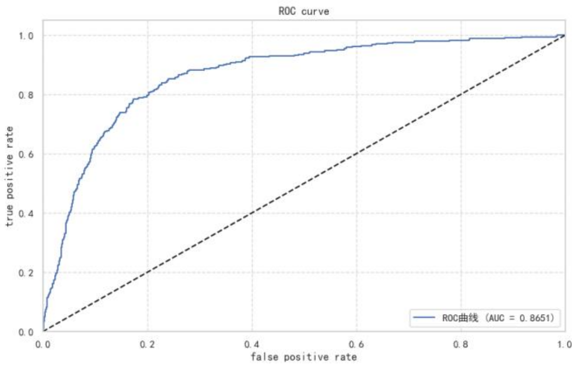


**Figure 3** ROC Curve

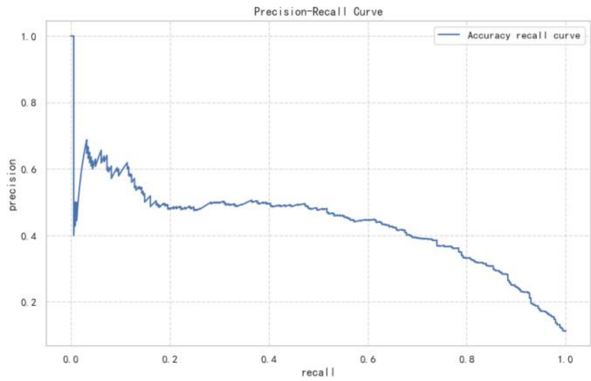


**Figure 4** Precision-Recall Curve

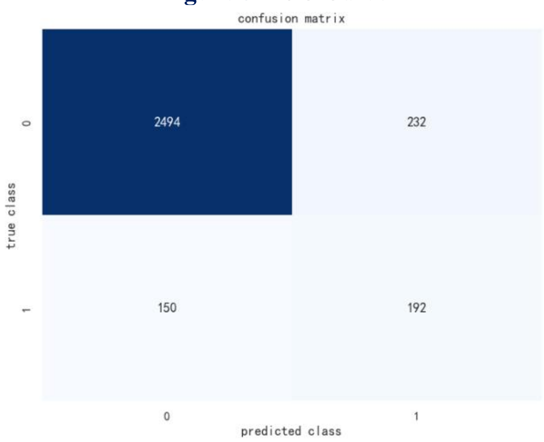


**Figure 5** Confusion Matrix

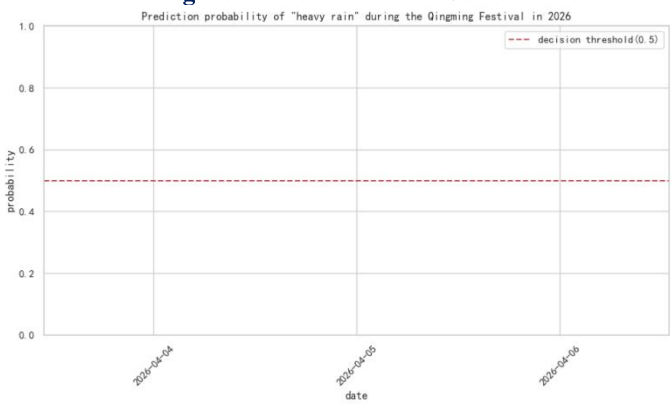


**Figure 6** Comparison of Predicted Probability of "Rain pouring" during the Qingming Festival in 2026

From Figure 1 to 5, it can be observed that the loss and accuracy curves show a steady improvement in the performance of the model during the training process. The ROC curve is close to the upper left corner, indicating that the model has high classification performance. The precision recall curve displays the performance of the model at different thresholds, providing a reference for the model in practical applications.

From Figure 6, it can be seen that the prediction probability of the weather phenomenon of "heavy rain" during the Qingming Festival in 2026 before and after the model update. It can be seen that the updated model's prediction probability (yellow) is basically consistent with the original prediction (blue) and decision threshold (red dashed line), indicating that there is no significant change in prediction accuracy after the model update, but there may be improvements in other aspects such as computational efficiency or generalization ability.

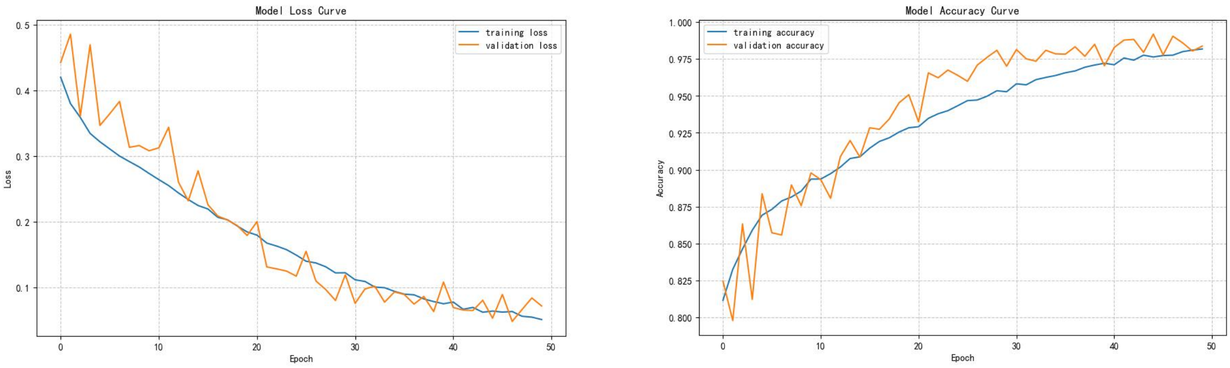


**Figure 7** Loss and Accuracy Changes during Model Training Process

From Figure 7, it can be seen that the predicted probability of the "rainy season" weather phenomenon during the Qingming Festival in 2026 before and after the model update. It can be seen that the updated model's prediction probability (yellow) is basically consistent with the original prediction (blue) and decision threshold (red dashed line), indicating that there is no significant change in prediction accuracy after the model update, but there may be improvements in other aspects such as computational efficiency or generalization ability.

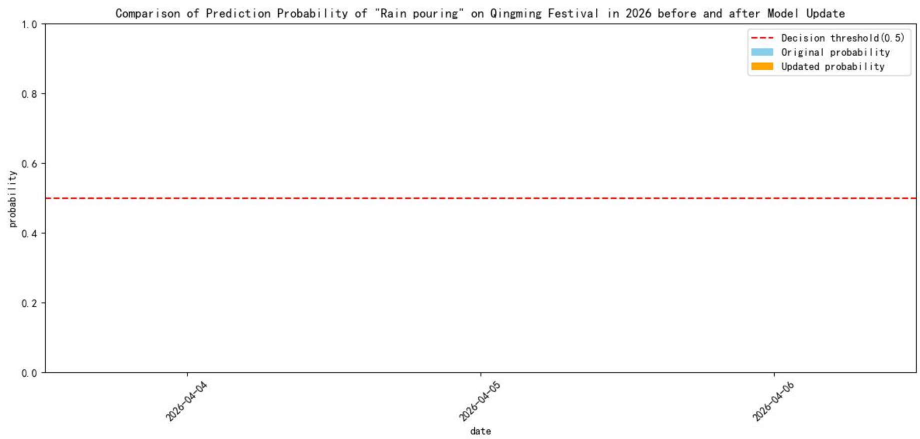


**Figure 8** Predicted Probability of "Heavy Rain" during the Qingming Festival in 2026

Figure 8 shows the predicted probability of "rain pouring" during the Qingming Festival in 2026. Only the decision threshold line (red dashed line) is shown in the figure, and the predicted probability is close to zero, indicating that according to the model prediction, the probability of "rain pouring" weather phenomenon occurring during this period is extremely low, providing favorable reference information for tourism and outdoor activities.

## 3 THE OPENING TIMING AND FLOWERING PERIOD PREDICTION OF REPRESENTATIVE FLOWERS DURING THE QINGMING FESTIVAL

### 3.1 Model Establishment

To solve the problem of predicting the opening time of flowers during the Qingming Festival, this study constructed a cumulative temperature model (GDD model) and a machine learning model (random forest regression) based on meteorological data, and improved the prediction accuracy through an integrated method [8]. This model can infer the First Bloom Date, Full Bloom Date, and End Bloom Date of various flowers based on historical temperature characteristics, providing scientific basis for planning outings and flower viewing [9].

#### *3.1.1 Construction of accumulated temperature (GDD) model*

The Growing Degree Days (GDD) model is a widely used heat accumulation model for predicting plant phenology. The basic principle is that the temperature above the reference temperature $T_{base}$ plays a decisive role in the development of plants.
The formula for calculating daily accumulated temperature is as follows:

$$GDD_d = \max\left(\frac{T_{\max,d} + T_{\min,d} - T_{\text{base}}}{2}, 0\right) \tag{6}$$

Among them, $T_{\max}$ and d are the highest temperatures on day d (unit: ° C), $T_{\min}$ and d are the lowest temperatures on day d (unit: ° C), Tbase is the benchmark temperature for flower growth (such as 5 ° C for rapeseed flowers, 4 ° C for cherry blossoms, and 6 ° C for peonies), GDDd is the accumulated temperature of the day, and if it is lower than the benchmark temperature, it is counted as 0.
The annual accumulated temperature is:

$$GDD_{\text{year}}(t) = \sum_{d=1}^{t} GDD_d \tag{7}$$

When $GDD_{year}(t) \geq GDD_{thresshold}$ , it is judged as the beginning of flowering period.
$GDD_{thresshold}$ is the threshold for the accumulation of heat required for flowering, which is calculated based on the date of flowering for many years and then averaged to obtain the corresponding GDD value:

$$GDD_{\text{threshold}} = \frac{1}{n}\sum_{i=1}^{n} GDD_{\text{year}_i}(\text{DOY}_i) \tag{8}$$

Where n is the number of historical years, and $DOY_i$ is the day of the year corresponding to the flowering period in the i-th year.

***3.1.2 Construction of random forest regression model***

In order to further improve the accuracy of flowering period prediction, a random forest regression model based on meteorological statistical features is introduced for fitting. This model takes the meteorological mean and total characteristics from January to May as inputs, and predicts the onset of flowering (DOY) as the output.
Assuming the input feature matrix is:

$$X = [x_1, x_2, \dots, x_m] \in \mathbb{R}^{n\times m} \tag{9}$$

The output label is:

$$y = [y_1, y_2, \dots, y_n] \in \mathbb{R}^{n} \tag{10}$$

Among them, $x_j$ represents the $j_{th}$ input feature, such as "January average temperature", "March total precipitation", etc., and $y_i$ represents the beginning flowering period of flowers in the i-th year (the yith day of the year).
Normalize features using standardized methods:

$$\hat{x}_j = \frac{x_j - \mu_j}{\sigma_j} \tag{11}$$

Where $u_j$ and $\sigma_j$ are the mean and standard deviation of feature $x_j$, respectively.
Random forest consists of multiple decision trees, each tree fits a set of sub samples, and the final output prediction result is the average of the predicted values of each tree:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t\left(\hat{X}\right) \tag{12}$$

Among them, $f_t$ is the t-th decision tree, and T is the total number of trees.

***3.1.3 Model integration and prediction strategy***

The final prediction of flowering period $D\hat{O}Y$ by the model is the fusion result of GDD model and random forest model. When the difference in predicted dates between the two is not significant (<10 days), take their mean:

$$\hat{\text{D}}\text{OY} = \frac{\text{DOY}_{\text{GDD}} + \text{DOY}_{\text{RF}}}{2} \tag{13}$$

When the difference is significant, the GDD model results based on phenological stability should be prioritized.
Finally, the predicted start flowering period date D1 is shifted backwards by $\delta_1$ and $\delta_2$ days to obtain the peak flowering period ($D_2=D_1+\delta_1$) and the end flowering period ($D_3=D_1+\delta_2$), where $\delta_1$ and $\delta_2$ are predetermined based on the type of flower (such as 5 and 14 days for cherry blossoms).
Through the above modeling framework, not only has the accuracy and stability of prediction been improved, but also the understanding and interpretability have been enhanced by combining the phenological laws of flowers, which can provide effective reference for Qingming Festival tourism activities [10].

### 3.2 Model Solution

**Table 1** Prediction Details of Flowering Periods for Three Flowers in 2026

| Flower Type | Predicted Initial Flowering Date | Predicted Full Flowering Date | Predicted Final Flowering Date | Predicted Flowering Duration |
|---|---|---|---|---|
| Rapeseed Flower | March 27, 2026 | April 3, 2026 | April 17, 2026 | 21 days |
| Cherry Blossom | March 19, 2026 | March 24, 2026 | April 2, 2026 | 14 days |
| Peony | April 2, 2026 | April 8, 2026 | April 20, 2026 | 18 days |

**Table 2** Prediction of the Opening Status of Various Flowers during the Qingming Festival in 2026

| Date | Rapeseed Flower Status | Cherry Blossom Status | Peony Status |
|---|---|---|---|
| April 4, 2026 | Full bloom period, gradually fading | Blooming period ended | Early bloom period, gradually thriving |
| April 5, 2026 | Full bloom period, gradually fading | Blooming period ended | Early bloom period, gradually thriving |

**Table 3** Prediction of the Overall Viewing Period for Qingming Festival

| | Best Viewing Period (Full Bloom to Final Bloom) | Viewing Period Ended | Early Viewing Period (Early Bloom to Full Bloom) |
|---|---|---|---|
| April 6, 2026 | Full bloom period, gradually fading | Blooming period ended | Early bloom period, gradually thriving |

From the analysis of Tables 1 to 3, it can be found that in terms of flowering period length, rapeseed flowers have the longest flowering period, which is 21 days, while cherry blossoms have the shortest flowering period, which is only 14 days. The flowering period of peonies is between the two, lasting for 18 days. During the first two days of Qingming Festival, rapeseed flowers and peonies still have ornamental value, while the viewing period for cherry blossoms has ended. On the day of Qingming Festival, the viewing period for rapeseed flowers and cherry blossoms has ended, while peonies have just entered the viewing period, which may affect tourists' viewing choices.

## 4 CONCLUSION

This study presents a robust predictive framework for weather and flowering periods during the Qingming Festival. The GRU-based weather classification model achieved 87.45% accuracy, effectively predicting "rain shower" events. The flowering prediction model accurately forecasted the periods for rapeseed flowers, cherry blossoms, and peonies, highlighting rapeseed flowers as the main ornamental species during the festival. The multi-objective path planning model optimized travel routes, recommending "Luoyang - Xi'an - Turpan" as an ideal three-day itinerary. Future research will focus on enhancing model generalization across diverse regions and climates, and further improving predictive performance through advanced techniques. The findings are valuable for tourism planning, enabling targeted promotion of rapeseed flower viewing and early peony blossoms. Agricultural managers can adjust planting schedules based on flowering predictions to maximize economic benefits. This study provides a practical guide for optimizing Qingming Festival tourism and supporting sustainable agricultural practices.

## COMPETING INTERESTS

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

## REFERENCES

[1] Sun Linxuan, Deng Longsheng, Tang Dong, et al. Deformation patterns and stability prediction of Dongjiatun landslide in Yan'an area, Shaanxi Province under different recurrence periods of rainfall. Journal of Earth Science and Environment, 2025, 47(03): 331-343.

[2] Shen Haojun, Luo Yong, Zhao Zongci, et al. Research on Summer Rainfall Prediction in China Based on LSTM Network Progress in Climate Change Research, 2020, 16(3): 263.

[3] Liu Dong, Ru Wenchao, Zhang Liangliang, et al. Application of Fishhawk Optimization Algorithm to Improve the Safety Evaluation of Regional Water Energy Grain Linkage System Using Random Forest Model. Journal of Northeast Agricultural University, 2025, 56(04): 64-74.

[4] Li Hengchang, Zhou Qing, Shi Lijuan, et al. Research on the joint inversion of cloud base height using FY-4B and millimeter wave radar based on random forest algorithm//Chinese Meteorological Society Summary of the 36th Annual Meeting of the Chinese Meteorological Society - S4 Atmospheric Detection Shandong Meteorological Engineering Technology Center; Meteorological Observation Center of China Meteorological Administration, 2025: 127-129.

[5] Wang Yanxiao. The impact of climate change on the flowering period of peonies in Luoyang in the past 50 years Research on Agricultural Disasters, 2022, 12(08): 88-90.

[6] Zhao F, Wang L, Xu S. Identification of QTL-by-environment interaction by controlling polygenic background effect. Journal of Genetics and Genomics, 2025, 52(07): 915-926.

[7] Li Dongmei, Li Wenquan. Calculation Method for Road Capacity. Journal of Henan University, 2002, 6: 24-27.
[8] Li Xin, Jifeng Dai, Jianxin Lin, et al. A review of research on vulnerability assessment indicators for urban road networks. Highway Transportation Technology, 2016: 155-157.
[9] Qi Ruizhen. Vulnerability Assessment of Xiamen Road Network Based on Complex Networks and Topological Potential. Huaqiao University, 2023.
[10] Dailisa, He Zhengguang, He Meng, et al. Meta analysis of the Impact of OSAHS on Traffic Accident Risk. Modern Medicine and Health, 2025, 41(07): 1699-1705+1710.

**World Journal of Economics and Business Research**
**Print ISSN: 2960-0073**
**Online ISSN: 2960-0081**
**DOI: https://doi.org/10.61784/wjebr3067**

# THE CONSTRUCTION PATH OF SMART CAMPUS DRIVEN BY ARTIFICIAL INTELLIGENCE AND BIG DATA

JiaoJiao Sun[1], Fei Wang[2*], QuanDe Dong[2], Cui Wang[1], Yaliu Yang[1]
*[1]Business School, Suzhou University, Suzhou 234000, Anhui, China.*
*[2]Network and Information Center, Suzhou University, Suzhou 234000, Anhui, China.*
*Corresponding Author: Fei Wang, Email: cumtwf@126.com*

**Abstract:** With the rapid development of artificial intelligence—especially Large Language Model (LLM)—and big data technologies, smart campus construction in universities faces new opportunities and challenges. This paper proposes a system architecture based on "data governance + model services + application integration" and conducts a benchmark analysis of its feasibility and effectiveness using representative university cases. The study also outlines a four-stage construction path: top-level design, driven by Data Middle Platforms, empowered by intelligent applications, and safeguarded by institutional mechanisms.In addition, it establishes a supporting framework composed of organizational coordination, institutional regulation, and data security compliance.The findings of this study provide both theoretical guidance and practical reference for system planning, technology selection, and scenario deployment in the construction of smart campuses in higher education.
**Keywords:** University informatization; Large language model; Data middle platform; Digital transformation; Construction path

## 1 INTRODUCTION

With the continuous breakthroughs in artificial intelligence (AI), especially Large Language Model (LLM) technology, the education sector is undergoing profound changes. In recent years, LLM systems represented by ChatGPT, DeepSeek, Claude, etc., have demonstrated excellent performance in natural language processing, cross-modal understanding, and automatic programming, accelerating the evolution of university smart campuses towards intelligence and integration [1]. Domestic LLM represented by DeepSeek have prominent advantages in Chinese understanding, code generation, and educational Q&A, and have gradually become an important technical engine for smart teaching and management in universities. At the same time, as the scale of data accumulated by universities becomes increasingly large, how to synergistically integrate big data with LLM to build an intelligent, efficient, and secure smart campus system has become a new topic in informatization construction. This is not only an inherent requirement for the modernization of the university governance system but also a key path in response to the national "Education Digitalization Strategy Action"[2].
This study introduces open-source LLM into university smart campus scenarios. It focuses on exploring their empowerment paths in intelligent services, behavior analysis, and personalized teaching. This approach aims to expand the boundaries of current smart campus research and fill the application research gap in the context of AI and big data integration. [3,4]. By constructing system models for multi-dimensional scenarios such as teaching, research, and management, and clarifying the technical path for university smart campuses, it helps improve resource allocation efficiency, enhance educational equity and service personalization, and support the construction of a student-centered smart governance system [5,6].
This study, based on LLM and university big data platforms, focuses on the construction path of smart campuses in universities under the background of the integration of AI and big data, mainly including:
(1) Clarifying the connotation and evolution trends of university smart campuses, and defining their functional positioning in the education digitalization strategy [7,8];
(2) Exploring typical application scenarios of LLM in smart campuses, including teaching assistance, research support, campus governance, and service optimization;
(3) Constructing a smart campus system architecture supported by "AI + Big Data", and designing key technical paths and integration logic;
(4) Selecting representative universities for case analysis, summarizing successful experiences, and extracting suggestions for construction paths and governance mechanisms.
This study adopts the following research methods:
(1) Literature Research Method: Systematically review domestic and international research results in smart campus construction and the integration of educational LLM to lay a theoretical foundation;
(2) Case Analysis Method: Select representative smart campus practice samples such as Nankai University, Tongji University, and Shenzhen University for systematic case analysis[9,10];
(3) Architecture Modeling Method: Combine the technical characteristics of LLM and big data platforms to build an implementable overall architecture for the smart campus system, forming a technical support plan[11,12].

(4) Construction and Verification Method [13,14]: System Architecture Design: Centered on the core needs and technical characteristics of the smart campus, construct a layered architecture model composed of "Perception Layer — Data Middle Platform Layer — LLM Layer — Application Layer "; Evaluation Model Construction: Based on literature and policies (such as the "Education Digital Transformation Strategy Guide"), extract four major dimensions: technology maturity, data governance, intelligent services, and organizational guarantee, to form an evaluation framework; Preliminary Verification: Use the expert consultation method to optimize the architecture and indicator design, and through case benchmarking analysis, use practical data from Tongji University, Nankai University, and Shenzhen University (such as response time, service coverage, and user satisfaction) to verify and adjust the weights of the model.

## 2 SMART CAMPUS AND LARGE LANGUAGE MODEL TECHNOLOGY SYSTEM

### 2.1 Smart Campus Development Stages

The current development of smart campuses can be roughly divided into the following three stages (see Table 1):

**Table 1** Smart Campus Development Stages

| Stage | Characteristic Description |
|---|---|
| Digital Campus Stage | Complete preliminary information system construction, achieve digitization of basic data and system integration. |
| Intelligent Campus Stage | Apply AI, big data, etc., to promote automation of business processes and personalization of services, improving management and teaching efficiency. |
| Smart Campus Stage | Fully integrate LLM (GPT, DeepSeek, etc.), knowledge graphs, multimodal AI, etc., to achieve a closed-loop governance and personalized empowerment system of "Ubiquitous Perception - Intelligent Cognition - Autonomous Decision-Making - Precise Services", realizing truly personalized education and refined management. |

Currently, university smart campuses are accelerating their evolution from "system parallel connection" to an integrated, intelligent ecosystem driven by "Data Middle Platform + Intelligent Core", showing a transformation trend from tool integration to intelligent collaboration, and from information aggregation to data governance [15].

### 2.2 Core Technology System Supporting Smart Campus

The evolution towards the 'Smart Campus' stage is underpinned by a sophisticated technology ecosystem. The core supporting technologies include:
(1) Big Data Technology: Used for storing, processing, and analyzing massive campus data, enabling student profiling, behavior prediction, management decision support, etc. [16];
(2) Artificial Intelligence Technology: Including natural language processing, knowledge graphs, machine learning, etc., used for intelligent Q&A, teaching evaluation, research assistance, etc.;
(3) Internet of Things (IoT) Technology: Enabling perception and control of physical campus spaces such as classrooms, laboratories, dormitories, and libraries;
(4) Edge Computing and Cloud Computing: Providing high-performance computing resources to support high concurrency and low latency requirements of the system;
(5) Blockchain Technology: Ensuring the security and traceability of teaching certifications, student records, and digital certificates.

## 3 APPLICATION SCENARIOS OF LARGE LANGUAGE MODELS IN SMART CAMPUS

With the increasing maturity of LLM technology, smart campuses are entering a new stage of leapfrogging from "digitalization" to "intelligence". LLM (such as the DeepSeek series), by virtue of their advantages in Chinese understanding, code generation, multi-turn dialogue, etc., deeply integrate with university big data systems, providing diversified intelligent support in teaching assistance, research support, campus governance, and service optimization.

### 3.1 Application in Smart Teaching Scenarios

To leverage the empowering role of LLM in smart higher education teaching, this study systematically categorizes three typical teaching application scenarios and integrates them with technical pathways. Table 2 illustrates three smart teaching scenarios, detailing the technical pathways from data platform construction to LLM integration, which highlights the model's role in enhancing teaching precision and interactivity.

**Table 2** Smart Teaching Scenarios

| Scenario | Technical Path |
|---|---|
| Personalized Learning Recommendation | Student Profile (Data Middle Platform) → LLM generates learning path → Precise resource push. Utilizes the powerful semantic understanding and generation capabilities of LLM, combined with knowledge graph association reasoning, to dynamically generate personalized learning paths and resource combinations that match individual cognitive levels and interest preferences. |
| Intelligent Q&A Teaching Assistant | Access course knowledge base → Cross-modal Q&A (text/image/voice). Accesses teaching management systems and course knowledge bases to build a digital teaching assistant, providing 24/7 Q&A support, improving teacher-student interaction efficiency. |
| Teaching Content Generation | Automatically generate question bank/lecture notes → Prompt engineering controls difficulty → Knowledge graph logic verification. Leverages LLM to quickly generate lecture notes, PPTs, question banks, and case discussion materials, improving teaching content quality and reducing preparation burden. |

### 3.2 Research Assistance Scenarios

As shown in Table 3, typical application scenarios of LLM in university research activities and their technical paths, covering the entire process from research assistance writing, literature processing to topic selection decision support, demonstrating the application value of LLM in improving research efficiency, enhancing text quality, and optimizing topic selection direction.

**Table 3** Research Assistant Scenarios

| **Scenario** | **Technical Path** |
|---|---|
| Literature Review Generation | Engineer research intent → Generate code → Connect to experimental platform for visualized results. |
| Paper Translation and Polishing | Template adaptation → Professional terminology consistency correction → Semantically enhanced translation. |
| Research Topic Assistance and Hotspot Identification | Build domain knowledge graph → Access journal/fund databases → LLM analyzes research trends and keywords → Outputs topic suggestions. |

### 3.3 Campus Governance Scenario Applications

As shown in Table 4, typical application scenarios of LLM in the field of university governance and services and their technical implementation paths, covering key links such as intelligent approval, behavioral safety warning, and digital guide, reflecting the model's comprehensive capabilities in text understanding, behavior recognition, and multimodal interaction, providing intelligent support for the construction of a smart governance system.

**Table 4** Campus Governance Scenarios

| Scenario | Technical Path |
|---|---|
| Intelligent Approval | Embed LLM into OA system → Automatically generate official documents/forms. |
| Behavioral Safety Warning | Multi-source data fusion → Abnormal behavior classification → Linkage intervention mechanism. |
| Digital Guide | Multimodal interaction (OCR + Voice + Map) → Multi-language support. |

## 4 SYSTEM ARCHITECTURE AND KEY TECHNOLOGY PATH DESIGN

### 4.1 System Overall Architecture Diagram

To achieve the goals of "intelligent perception, data-driven, model empowerment, scenario implementation" for university smart campuses, the system should build a trinity architecture system of "Data Governance + Model Services + Application Integration", meeting requirements such as ubiquitous access, unified scheduling, high availability, and scalability. This architecture adopts a layered and decoupled design: the underlying IoT devices achieve omnidirectional perception; the Data Middle Platform layer aggregates core data assets such as educational administration and behavior; the AI Large Model layer provides intelligent engines such as general-purpose LLM, multimodal, and fine-tuning components; ultimately supporting four major application scenarios including teaching assistance, research support, campus governance, and service optimization. This design, through vertical collaboration of "Perception-Data-Model-Application", builds an intelligently driven closed-loop governance system, see Figure 1.

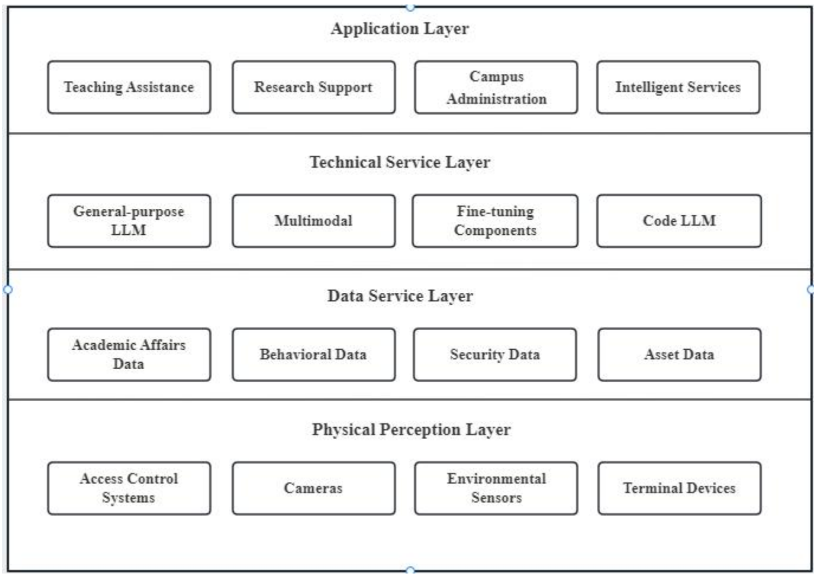


**Figure 1** Smart Campus System Architecture

### 4.2 Core Modules and Key Technology Paths

The system construction of the smart campus relies on the collaborative linkage of three core modules: Data Middle Platform and Governance Technology, Model Deployment and Application Technology, and Scenario Application Integration and Service Portal. The technical implementation paths for each module are as follows:

#### *4.2.1 Data Middle Platform and governance technology*

The Data Middle Platform is the foundational platform for achieving the integration, governance, and servitization of educational data. Its key technology system includes:

(1) Multimodal Data Collection: Supports unified access and management of structured (e.g., educational administration system tables), semi-structured (JSON/log files), and unstructured data (teaching videos, audio, text);

(2) Entity Profile Construction: Through Master Data Management and tag systems, integrate multi-source data to build 360° panoramic profiles of core educational entities such as students, teachers, and courses;

(3) Data Governance and Standardization: Establish data cleaning, deduplication, quality verification, and standardization processes to ensure the accuracy, consistency, and reliability of educational data;

(4) Data Servitization and Opening: Provide data support for teaching, research, and management applications through methods like RESTful API and GraphQL interfaces, promoting the efficient reuse of data resources.

#### *4.2.2 Model Deployment and application technology*

(1) Diversified Deployment Modes: Local Private Deployment: For sensitive scenarios, ensuring data compliance and closed-loop flow within the campus; Cloud-Edge Collaborative Deployment: Deploy lightweight models on edge nodes such as teaching terminals and IoT devices, with the central cloud scheduling full models, achieving low-latency response and unified management;

(2) Servitization and Enhanced Inference: Model Servitization Encapsulation: Encapsulate model capabilities as high-availability, high-concurrency API services, supporting dynamic batch processing and multi-version management; RAG Enhanced Retrieval Mechanism: Build a campus knowledge retrieval system based on FAISS + BERT, retrieving relevant content from databases such as educational regulations and academic literature, enhancing the accuracy and factuality of model responses.

(3) Scenario Application Integration and Service Portal: Build a user-centered integrated intelligent service entry. Key integration technologies include: unified identity and permission management, multi-terminal adaptive support, microservices and elastic architecture, educational knowledge graph empowerment.

#### *4.2.3 Technical implementation roadmap suggestions*

We propose a four-stage roadmap to promote construction: as shown in Table 5:

**Table 5** University Smart Campus Construction Stages and Implementation Path

| Stage | Core Tasks | Time Cycle |
|---|---|---|
| Step1 | Build the Data Middle Platform, complete unified data governance system | 1-3 months |
| Step2 | Privately deploy LLM, complete scenario-specific fine-tuning and adaptation | 3-6 months |
| Step3 | Integrate key application scenarios, such as intelligent Q&A, teaching recommendation, intelligent administrative Q&A, etc. | 6-9 months |
| Step4 | Expand diverse application scenarios, establish unified portal and service orchestration system | 9-12months |

## 5 TYPICAL CASE ANALYSIS

### 5.1 Case Selection and Research Approach

This chapter selects three universities with representative practical achievements in empowering smart campuses with AI and big data: Tongji University, Nankai University, and Shenzhen University. These three have different focuses in technical paths,Application Effect, and Management Mechanism, which can well reflect the diversity and advancement of smart campus construction in China. By analyzing their technical implementation, application effects, and management mechanisms from multiple dimensions, successful experiences are summarized to provide replicable and promotable path references for other universities.

### 5.2 Core Achievements and Technical Paths

#### *5.2.1 Tongji University: data-driven teaching quality evaluation system*

Technical paths: Multi-source data integration (educational administration/teaching evaluation/classroom video) → Machine learning for academic risk identification → BI visualization platform;

Application Effect: Teaching evaluation coverage increased by 24%, covering over 90% of courses;

Management Mechanism: Three-level linkage between Academic Affairs Office + Information Center + Colleges, establishing data quality responsible persons.

#### *5.2.2 Nankai University: artificial intelligence research assistant system*

Technical paths: Integrate DeepSeek/GLM LLM→ Build research knowledge graph → Natural language interaction interface;

Application Effect: Research topic selection efficiency improved by 60%, over 90% user satisfaction.

Management Mechanism: Led by the library for development, introducing AI ethics guidelines + behavior log auditing.

#### *5.2.3 Shenzhen University: campus-wide intelligent governance platform*

Technical paths: Unified Data Middle Platform accesses security/logistics/energy consumption systems → Face recognition behavior warning → Fault self-reporting and dispatching;

Application Effect: Event response ≤ 3 minutes, risk index decreased by 38%, service satisfaction 94.2%;

Management Mechanism: School-level construction task force promotes, "One Governance Map" realizes visualization of authority, responsibility, and processes.

### 5.3 Summary of Successful Experiences and Promotable Paths

Based on the in-depth analysis of the cases of Tongji, Nankai, and Shenzhen University, combined with cross-evaluation of multi-school experiences in expert consultations, this paper extracts five key dimensions for smart campus construction and their successful experiences (as shown in Table 6). Experts generally agree that the "Unified Data Middle Platform + AI Core Model" architecture of the technology system is the foundational support (confirmed by Shenzhen University case); the "multi-department collaboration + clear responsibilities" (Tongji's three-level linkage mechanism) in the organizational governance dimension and the "introduction of ethical norms + behavior auditing" (Nankai's AI ethics guidelines) are key guarantees to avoid "emphasizing technology over management", which is consistent with the conclusions of literature. Experts particularly pointed out that deep user participation (e.g., Shenzhen University's high satisfaction stems from the closed-loop of teacher-student feedback) and the bottom line of security and compliance (all three universities have established graded protection) are core elements for project sustainable development.

**Table 6** Summary of Successful Experiences in Five Dimensions of Smart Campus Construction

| Dimension | Successful Experience |
|---|---|
| Technology System | Build a unified Data Middle Platform, integrate AI core models, and promote business intelligence collaboration. |
| Application Scenarios | Effective implementation in multiple fields such as teaching quality evaluation, research assistance, and logistics & security governance. |
| Organizational Governance | Implement multi-department collaboration mechanisms, clarify data responsibility boundaries, and introduce ethical norms. |
| User Participation | Teachers and students deeply participate in system feedback, achieving positive interaction and continuous optimization. |
| Security&Compliance | Establish AI usage security boundaries and behavior audit mechanisms to ensure the system is trustworthy, controllable, and traceable. |

## 6 CONCLUSION

This paper systematically constructs an overall framework for smart campus construction driven by artificial intelligence and big data, clarifies the system logic centered on "Top-Level Design - Middle Platform Drive - Intelligent Applications", and points out that the integration of AI and big data is the core engine for promoting the intelligent reconstruction of campus management, teaching, research, and service models. Research shows that the effective implementation of technology must rely on sound institutional guarantees, collaborative organizational mechanisms, and continuous development of teacher-student digital capabilities, forming a long-term mechanism for the positive interaction of "Technology-Mechanism-Culture". Although this study preliminarily proposes a construction path and evaluation model, its applicability still needs to be empirically tested on a larger scale in different types of universities; in the future, based on multi-university research and longitudinal data support, we will further explore the dynamic evolution mechanism and sustainable governance path of smart campuses.

## COMPETING INTERESTS

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

## FUNDING

This work was supported by 2025 Suzhou City Social Science Innovation and Development Research Project (Grant No. 2025CX004); Suzhou University Horizontal Scientific Research Project (Grant No. 2025xhx059); 2025 Suzhou University Doctoral Research Start-up Fund Project (Grant No. 2025BSK011); 2023 Research Project on Quality Culture Construction in Local Application-Oriented Universities from the Perspective of Building a Strong Education Nation (Grant No. 2023jxgl053). 2023 Scientific Research Development Fund Project (Grant No. 2023fzjj06); Anhui Province Quality Engineering Project (2023xjzlts069); Suzhou University Teaching Innovation Team (szxy2024cxtd02) and Provincial Quality Engineering Project (2023zygzts079).

## REFERENCES

[1] Chen Chen. The Realistic Path of Empowering Smart Education with Large Model Technology. Technology Vision, 2023(18): 22-24.
[2] Zhang Zhiyong, Li Xiang. Research on the Path and Strategy of Smart Campus Construction. China Educational Technology, 2023(2): 25-32.
[3] Gao Feng, Zhang Heng. Research on Big Data-Driven University Informatization Development. Modern Educational Technology, 2021, 31(4): 47-54.
[4] Chen Hui. Construction of University Smart Campus Data Governance System. Information Recording Materials, 2020(12): 56-58.
[5] Zhao Pengfei. Exploration of Artificial Intelligence Application in University Smart Governance. Information Science, 2022, 40(9): 112-117.
[6] Bai Zhongxian, Zhang Bin. Large Language Models Empowering Internet Knowledge Services: Platform Architecture, Application Challenges, and Governance Paths. Chinese Journal of Educational Informatics, 2025, 31(07): 40-49.
[7] Wang Huan. Comparative Analysis of Smart Education Development in Domestic and Foreign Universities. Education Modernization, 2022(23): 134-138.

[8] Hu An, Li Ze. Research on the Path of Education Reform Based on Artificial Intelligence. Education Theory and Practice, 2022(4): 19-24.
[9] Zeng Wei, Zhao Qian. Construction and Practice Exploration of University Data Middle Platform. Information System Engineering, 2023(5): 75-78.
[10] Tang Lei. Design and Implementation Analysis of University AI Assistant. Software Guide, 2021, 20(9): 102-106.
[11] Li Nan, Zhou Feng. Research on Smart Teaching Platform for Personalized Learning. e-Education Research, 2022, 43(7): 34-40.
[12] Du Peng. Research and Practice of University Intelligent Q&A System. Chinese Journal of Educational Informatics, 2023(6): 66-69.
[13] Wang Sen. Practice and Reflection on the Construction of University Smart Service Platform. Higher Education Research, 2022, 43(3): 88-93.
[14] Yang Fan. Research on AI-based Campus Security Management System. Information and Computer, 2021, 33(15): 101-103.
[15] Liu Jin. The Evolution of Educational Large Models and Their Application Prospects in Universities. Educational Measurement and Evaluation, 2024(2): 41-45.
[16] Zhou Jing. Trends and Challenges of Education Digital Transformation. Modern Distance Education, 2023(10): 54-59.

World Journal of Economics and Business Research
Print ISSN: 2960-0073
Online ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3068

# INNOVATIVE REFORM PATHS OF BASIC ECONOMICS COURSE FROM THE PERSPECTIVE OF CURRICULUM IDEOLOGICAL AND POLITICAL EDUCATION

YuHan Xiao
*Management Department, Shanghai University of Engineering Science, Shanghai 200437, China.*
*Corresponding Email: yuhan0394@163.com*

**Abstract:** As a foundational core course for economics and management majors, Basic Economics covers core theories including supply and demand, market mechanisms, macroeconomic regulation, and economic growth. It undertakes the dual task of imparting economic theoretical knowledge and shaping students' correct economic values. From the perspective of Curriculum Ideological and Political Education (CIPE), traditional Basic Economics teaching faces three key issues: the separation of theoretical teaching from ideological guidance, insufficient connection with China's economic practice, and inadequate response to digital economic challenges. This paper proposes a "three-dimensional integration + four-system promotion" reform framework. Based on the course's "micro-macro-development" theoretical hierarchy, it systematically integrates ideological elements (such as national conditions awareness, social responsibility, and fairness-efficiency concepts), constructs an "AI+BOPPPS" teaching model adapted to economic scenario simulation, establishes a multi-dimensional evaluation system covering theoretical mastery and value shaping, and strengthens the "school-enterprise-expert-technology" collaborative guarantee mechanism. The aim is to build an educational system centered on "theory as the foundation, practice as the link, and values as the core," cultivating compound talents who understand economic principles, recognize national conditions, and possess a sense of social responsibility.
**Keywords:** Basic economics; Curriculum ideological and political education; AI+BOPPPS teaching model; Multi-dimensional evaluation; Collaborative guarantee mechanism

## 1 INTRODUCTION

Against the backdrop of profound changes in the global economic pattern and China's advancing high-quality economic development, Basic Economics, as the "introductory course" for economics and management majors, plays a crucial role. It not only needs to help students master the basic logic of market operations (e.g., the law of supply and demand, price mechanisms) and macroeconomic regulation tools (e.g., fiscal policy, monetary policy) but also guide students to establish a correct understanding of China's socialist market economy system and develop a sense of responsibility for participating in national economic construction.
However, current teaching practice has prominent shortcomings. A survey of 120 students from 15 universities majoring in economics and management shows that 85% of respondents believe "Basic Economics is overly theoretical and lacks connection with China's real economic issues," and 78% point out "traditional teaching fails to explain the ideological connotation behind national economic policies." Specifically, three problems stand out: first, the overemphasis on Western economic theories while neglecting the integration of China's economic practice and ideological elements (e.g., only explaining Adam Smith's "invisible hand" without analyzing the "visible hand" of China's macro regulation); second, single teaching methods relying on lectures, with insufficient simulation of real economic scenarios (e.g., inability to simulate policy-making processes for responding to inflation or enterprise social responsibility decisions); third, an evaluation system centered on theoretical knowledge, ignoring the assessment of students' economic values and practical application abilities (e.g., no evaluation of students' understanding of "common prosperity" or "sustainable development").
These issues limit the course's value-shaping function. Especially amid global economic volatility and China's promotion of major strategies such as the "dual carbon" goal and rural revitalization, reforming Basic Economics from the CIPE perspective is urgent. Exploring effective reform paths is of great practical significance for improving students' comprehensive quality and cultivating talents who serve national economic development.

## 2 CONTENT INTEGRATION: SYSTEMATIC DESIGN OF IDEOLOGICAL ELEMENTS BASED ON THEORETICAL HIERARCHY

According to the "microeconomics-macroeconomics-economic development" theoretical structure of Basic Economics, a "three-dimensional progressive" ideological content system is constructed to realize the organic integration of economic theories and value guidance[1].

### 2.1 Microeconomics Module: Shaping Market Ethics and Social Responsibility Awareness

In the "supply and demand theory" chapter, the case of "mask price fluctuations during the COVID-19 pandemic" is used to explain the dual role of market mechanisms and government regulation. Students are guided to understand that "the market is not omnipotent—enterprises must abide by market ethics (avoiding hoarding and price gouging) while pursuing profits." In the "enterprise theory" section, Huawei's independent innovation efforts under technological blockades are analyzed to illustrate the connection between "enterprise development" and "national technological security," fostering students' awareness that "enterprises bear social responsibilities beyond profit-seeking."

In the "consumer behavior theory" module, the "dual carbon" strategy is integrated with the concept of green consumption[2]. The growing market share of new energy vehicles in China is used as an example to explain how consumer choices drive environmental protection, helping students recognize that "individual consumption behaviors are linked to national sustainable development" and advocating green and healthy consumption concepts. Additionally, the "labor market" theory is combined with the "common prosperity" strategy: the income gap between industries in China is analyzed, and the government's policies to "increase the income of low-income groups and expand the middle-income group" are explained, guiding students to establish a correct view of fair distribution.

### 2.2 Macroeconomics Module: Cultivating National Conditions Awareness and Policy Recognition

In the "national income theory" chapter, China's GDP growth and per capita disposable income data since the reform and opening-up are compared with those of other countries, highlighting China's economic development achievements and cultivating students' confidence in the socialist market economy system. In the "fiscal policy" section, the role of China's "proactive fiscal policy" in stabilizing economic growth—such as increasing investment in rural infrastructure and people's livelihood projects—is analyzed[3]. The case of "financial support for rural revitalization in underdeveloped areas" is used to help students understand that "fiscal policy is a key tool for promoting common prosperity and coordinated regional development."

In the "monetary policy" module, the People's Bank of China's "prudent monetary policy" practice is taken as an example to explain how the central bank adjusts the money supply to maintain price stability and support real economic development. Combined with the case of "regulating Internet finance to prevent systemic risks," students are guided to recognize the importance of "financial stability for national economic security." In the "international trade" chapter, the impact of the "Belt and Road Initiative" on China's and global economic growth is analyzed, contrasting China's "mutual benefit and win-win" foreign trade concept with Western trade protectionism, and cultivating students' global vision and national mission[4].

### 2.3 Economic Development Module: Establishing Sustainable Development Concepts and Historical Mission Awareness

In the "economic growth theory" chapter, the traditional "extensive growth model" (relying on resource consumption) is compared with China's "high-quality development model" (driven by innovation). The transformation and upgrading of the manufacturing industry in the Yangtze River Delta—such as the shift from low-value-added processing to high-tech manufacturing—is used to explain China's economic transformation path, guiding students to understand that "innovation is the core driver of high-quality development." In the "sustainable development" section, the "dual carbon" strategy and ecological civilization concepts are integrated: the economic and environmental benefits of developing renewable energy (e.g., wind and solar power) in Gansu and Qinghai provinces are analyzed, helping students establish the concept of "coordinated development of economy, society, and the environment."

In the "rural economic development" module, the "rural revitalization" strategy is the focus. The case of "rural e-commerce promoting agricultural product sales in poverty-stricken areas of Guizhou" is used to explain how economic means solve the "urban-rural dual structure" problem. Students are guided to recognize the importance of "rural revitalization for national common prosperity" and inspired to develop a sense of mission to participate in rural construction after graduation.

## 3 MODEL INNOVATION: BUILDING AN "AI+BOPPPS" TEACHING MODEL ADAPTED TO ECONOMIC SCENARIOS

To address the "theory-practice separation" in traditional teaching, AI technology is integrated with the BOPPPS framework to create a teaching model suitable for economic scenario simulation, realizing the deep integration of theoretical teaching, ideological guidance, and practical application[5].

### 3.1 Overview of the "AI+BOPPPS" Model

Based on the traditional BOPPPS model (Bridge-in, Objective, Pre-assessment, Participation, Post-assessment, Summary), two core links are added: "Economic Scenario Simulation" and "AI Policy Effect Analysis." "Economic Scenario Simulation" uses AI to build realistic economic scenarios (e.g., "government responding to inflation," "enterprises making green investment decisions") for student participation in decision-making[6]. "AI Policy Effect Analysis" uses AI data analysis tools to simulate the impact of different economic policies (e.g., fiscal stimulus, tax cuts) on economic indicators, helping students understand the ideological connotation and practical effects of policies.

### 3.2 Specific Application of the Model

#### *3.2.1 Pre-class: AI-pushed scenarios and knowledge linking*

Before the "fiscal policy" class, the AI platform (e.g., Moso Teach, Learning Pass) pushes a simulated scenario of "a local government increasing rural infrastructure investment" and links it to the "national income multiplier effect" learned in the previous class. Pre-questions are proposed: "How does rural infrastructure investment drive local economic growth? What role does it play in rural revitalization?" Before the "sustainable development" class, the AI system pushes a case of "a province's wind power project driving local employment and reducing carbon emissions" and connects it to the "environmental externality" theory, guiding students to think about "how the government solves environmental externality problems through policies."

The AI system also generates a "theoretical-ideological knowledge map" for each student, marking connections between economic theories and ideological elements (e.g., "supply and demand theory" → "market ethics" → "social responsibility"; "fiscal policy" → "national conditions awareness" → "common prosperity"), helping students form a systematic cognitive framework.

#### *3.2.2 In-class: BOPPPS-linked interactive teaching with scenario simulation*

Bridge-in: For the "international trade" class, an AI-produced short video about "the Sino-European Railway's role in promoting China-Europe trade" is played, guiding students to connect the "comparative advantage theory" with the "Belt and Road Initiative" and triggering reflections on "China's role in global economic cooperation." For the "rural economic development" class, a video about "rural e-commerce development in a poverty-stricken county in Yunnan" is shown, linking to the "rural revitalization" strategy to arouse students' attention to rural issues.

Objective: Three-dimensional teaching goals (knowledge, ability, ideology) are set. For example, the goals of the "monetary policy" class are: "Master monetary policy tools (knowledge), analyze the impact of interest rate adjustments on the real economy (ability), and understand the role of prudent monetary policy in maintaining national financial stability (ideology)."

Pre-assessment: A 5-minute quick quiz is conducted via the AI interactive platform, including questions such as "What are the main tools of China's proactive fiscal policy?" and "How does the 'dual carbon' strategy affect enterprise investment decisions?" The AI system instantly calculates the correct rate, and the teacher adjusts the teaching focus accordingly (e.g., if the correct rate of fiscal policy questions is less than 70%, more time is spent explaining the connection between fiscal policy and rural revitalization).

Participation: AI-assisted scenario simulation activities are organized. In the "macroeconomic regulation" class, students are divided into "government departments," "enterprises," and "consumers" using an AI economic simulation system: "government departments" formulate fiscal and monetary policies to address inflation; "enterprises" adjust production and pricing strategies based on policies; "consumers" change consumption behaviors. During the activity, the AI system dynamically displays the impact of policy adjustments on GDP, price levels, and employment rates, reminding "government departments" to balance economic growth and social equity (e.g., avoiding excessive austerity that harms people's livelihoods), guiding students to understand the balance between policy goals and ideological connotations.

Post-assessment: The AI system designs scenario-based test questions, such as: "In the context of the 'dual carbon' strategy, what fiscal and monetary policies can a local government adopt to promote new energy vehicle development? What impacts will these policies have on local economic growth and environmental protection?" Test results are used to evaluate students' integration of theoretical knowledge, policy application, and ideological understanding.

Summary: The AI system generates a "teaching summary report" that organizes core economic theories and corresponding ideological elements (e.g., "market mechanisms" → "market ethics"; "macroeconomic regulation" → "national conditions awareness"). The teacher emphasizes the importance of "applying economic theories to analyze China's reality and establishing correct values" and guides students to pay attention to national economic policies in daily life.

#### *3.2.3 Post-class: AI-generated personalized assignments and practice guidance*

After class, the AI system assigns personalized tasks based on students' learning weaknesses. For students weak in "connecting theories with national policies," the task is "analyzing the role of China's 'tax reduction and fee reduction' policy in promoting enterprise development and common prosperity." For students lacking "practical application ability," the task is "using AI economic simulation tools to design a policy plan for a county to develop rural e-commerce."

The AI platform also pushes extended learning resources: after the "economic growth" class, it pushes the 2024 China Economic Development Report issued by the National Bureau of Statistics, linking it to the "high-quality development" theory; after the "sustainable development" class, it pushes the China Dual Carbon Progress Report, connecting it to the "ecological civilization" concept[7]. Students are required to write an 800-word learning reflection, and the AI system conducts a preliminary review (e.g., checking whether the analysis of the ideological connotation of policies is in place) before feeding back to the teacher for further comments.

## 4 EVALUATION OPTIMIZATION: ESTABLISHING A MULTI-DIMENSIONAL EVALUATION SYSTEM

To break the traditional "theory-centered" evaluation model, a "three-dimensional (knowledge-ability-ideology) + four-subject (teacher-enterprise-expert-student) + whole-process (pre-class-in-class-post-class)" multi-dimensional evaluation system is built, ensuring the effectiveness of CIPE.

### 4.1 Whole-Process Evaluation

The evaluation covers three stages, with ideological literacy accounting for 30% of the total score[8]:
Pre-Class Evaluation: Based on AI records of students' preview of scenarios and pre-question answers, evaluate their "initiative to connect economic theories with national policies and ideological elements" (e.g., whether they can link "fiscal policy" to the "rural revitalization" strategy).
In-Class Evaluation: Evaluate students' performance in interactive activities: in scenario simulations, assess whether "government departments" consider social equity when formulating policies and whether "enterprises" take environmental protection into account when making decisions; in group discussions, evaluate the depth of students' economic problem analysis and the correctness of their values (e.g., whether they recognize the importance of common prosperity and sustainable development). The AI system assists in evaluation by analyzing speech frequency, the rationality of policy suggestions, and the accuracy of value expression.
Post-Class Evaluation: Evaluate the quality of personalized assignments and learning reflections: assess whether the analysis of national economic policies is in-depth (e.g., whether the impact of "tax reduction and fee reduction" on enterprises and residents is correctly analyzed) and whether economic values are correct (e.g., whether the view on the "relationship between the market and the government" is consistent with China's national conditions).

### 4.2 Multi-Dimensional Evaluation

The evaluation content includes three dimensions:
Professional Knowledge: Evaluate students' mastery of core economic theories (e.g., supply and demand, fiscal and monetary policies) through theoretical tests and scenario analysis questions.
Practical Ability: Evaluate students' ability to apply economic theories to solve practical problems (e.g., designing local economic development policy plans) and their ability to use AI tools for economic simulation and data analysis.
Ideological Literacy: Evaluate students' economic values (e.g., correct views on the market and the government, fairness and efficiency), national conditions awareness (e.g., understanding of China's economic system and development stage), and social responsibility (e.g., awareness of participating in national economic construction and promoting sustainable development). For example, in the "enterprise decision-making" scenario simulation, evaluate whether students consider environmental protection and social benefits; in the "income distribution" analysis, evaluate whether students recognize the importance of common prosperity.

### 4.3 Multi-Subject Evaluation

Teacher Evaluation: Focus on students' mastery of theoretical knowledge and in-class value expression (e.g., the rationality of policy analysis in group discussions).
Enterprise Evaluation: Invite economists and managers from local enterprises (e.g., financial institutions, manufacturing companies) to evaluate students' performance in practical activities (e.g., internships in enterprise economic analysis departments or participation in local economic research projects). For example, evaluate whether students can correctly analyze the impact of macro policies on enterprise operations and whether they have a sense of social responsibility.
Expert Evaluation: Invite experts from local development and reform commissions, economic research institutions, and universities to evaluate students' comprehensive ethical decision-making abilities (e.g., solutions to "balancing economic growth and environmental protection" scenarios).
Student Self and Peer Evaluation: Students evaluate their own cross-course learning performance (e.g., whether they actively connect ethical concepts across courses) and score peers' teamwork and ethical performance in group activities.
The AI system integrates scores from all subjects, generates a "comprehensive quality evaluation report" for each student, and provides targeted improvement suggestions (e.g., "strengthen the connection between 'digital ethics' and economic policy analysis").

## 5 GUARANTEE STRENGTHENING: CONSTRUCTING A "SCHOOL-ENTERPRISE-EXPERT-TECHNOLOGY" COLLABORATIVE MECHANISM

### 5.1 Teacher Team Construction

Special Training: Organize training on "CIPE integration in Basic Economics," inviting experts from the Ministry of Finance, local economic research institutions, and educational technology companies to lecture. The training covers methods of excavating ideological elements in economic theories, "AI+BOPPPS" model application skills, and experience in analyzing China's economic policies.
Practice Exchange: Arrange teachers to participate in 2-month internships or research projects in local government economic management departments (e.g., development and reform commissions, bureaus of statistics) or key enterprises, collecting first-hand materials on China's economic practice (e.g., local rural revitalization policies, enterprise green development cases) to enrich teaching content.
Teaching and Research Teams: Establish a "CIPE teaching and research team" for Basic Economics, consisting of professional teachers, ideological and political tutors, and enterprise experts. The team holds monthly meetings to discuss the design of ideological elements and the optimization of the "AI+BOPPPS" model.

### 5.2 School-Enterprise Cooperation

Practice Bases: Cooperate with local enterprises (e.g., new energy companies, agricultural product e-commerce platforms) and government economic management departments to build “Basic Economics CIPE practice bases.” Arrange students to participate in practical activities such as “local economic data statistics” and “enterprise market demand analysis,” allowing them to experience the connection between economic theories and real economic operations[9].
Workplace Lectures: Invite local economic officials and enterprise managers to give “workplace ideological and political lectures,” sharing practical cases such as “formulating regional industrial policies” and “enterprises fulfilling social responsibilities in poverty alleviation,” helping students understand the practical application of economic theories.
Textbook Co-Development: Jointly develop CIPE-integrated Basic Economics textbooks with enterprises and research institutions, integrating China’s latest economic policies (e.g., dual carbon, rural revitalization) and typical cases into textbook content, and adding a “China Economic Practice” chapter to enhance the connection between theory and practice.

### 5.3 Technical Support

Cooperate with educational technology companies (e.g., NetEase Cloud Classroom, Chinese University MOOC) to customize an AI teaching platform for Basic Economics. The platform has three core functions: 1) Intelligent case management: automatically collect and classify the latest economic cases, and tag ideological elements (e.g., “common prosperity,” “sustainable development”); 2) Learning analysis: track students’ learning trajectories (e.g., resource browsing time, scenario participation frequency) and generate ideological literacy evaluation reports; 3) Interactive simulation: support AI-based economic scenario simulation and policy effect analysis to realize immersive teaching.
The school establishes a “platform update mechanism” to update the platform’s AI algorithms and case database quarterly, ensuring that teaching content keeps pace with the latest economic developments (e.g., adding content on digital economy and platform economic regulation).

## 6 CONCLUSION

The innovative reform of Basic Economics from the perspective of CIPE is a systematic project involving content, model, evaluation, and guarantee. By reconstructing teaching content based on the “micro-macro-development” theoretical hierarchy, building an “AI+BOPPPS” teaching model adapted to economic scenarios, optimizing a multi-dimensional evaluation system, and strengthening a “school-enterprise-expert-technology” collaborative guarantee mechanism, the reform can not only improve students’ mastery of economic theories but also guide them to establish correct economic values, national conditions awareness, and social responsibility.
In the future, with the development of digital technology and the deepening of CIPE, the reform should continue to iterate: in terms of content, integrate new elements such as digital economy and global economic governance; in terms of technology, apply virtual reality (VR) to simulate more complex economic scenarios (e.g., international trade negotiations, macro policy-making); in terms of cooperation, expand international exchanges to help students understand China’s role in the global economy while maintaining a global vision. Ultimately, the goal is to cultivate high-quality talents who can serve China’s high-quality economic development and participate in global economic governance.

## COMPETING INTERESTS

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

## REFERENCES

[1] Chen X. Integration of Curriculum Ideological and Political Elements in Basic Economics: Practice and Exploration. Journal of Higher Education Research, 2025, 38(3): 45-52.
[2] Stiglitz J E. People, Power, and Profits: Progressive Capitalism for an Age of Discontent. W.W. Norton & Company, New York, USA. 2019.
[3] Zhang H. Connection between Basic Economics and National Economic Strategy: Paths and Cases. Economic Review, 2022, 19(4): 112-118.
[4] Rodrik D. Straight Talk on Trade: Ideas for a Sane World Economy. Princeton University Press, 2018. DOI: https://doi.org/10.2307/j.ctvc779z4.
[5] Li M. Application of AI Technology in Basic Economics Teaching Reform. Journal of Economics and Management, 2024, 21(2): 89-96.
[6] Wang Q. Research on the “AI+BOPPPS” Teaching Model in Basic Economics from the Perspective of CIPE. Journal of Curriculum and Teaching Methodology, 2023, 32(5): 67-73.
[7] National Bureau of Statistics. 2024 China Economic Development Report. China Statistics Press, 2024.

[8] Zhao J. Evaluation System Design of Basic Economics Teaching from the Perspective of CIPE. Journal of Educational Measurement and Evaluation, 2024, 18(3): 56-63.
[9] Liu Y. Construction of School-Enterprise Collaborative CIPE Mechanism for Basic Economics. Research on Higher Education, 2025, 43(1): 78-85.

World Journal of Economics and Business Research
Print ISSN: 2960-0073
Online ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3069

# COST OPTIMIZATION PATHWAYS DRIVEN BY DIGITAL FINANCE: EMPIRICAL EVIDENCE FROM DONGGUAN' S MANUFACTURING SECTOR

Jing Wang*, Ping Peng
*School of Finance and Economics, Guangdong University of Science and Technology, Dongguan 523000, Guangdong, China.*
*Corresponding Author: Jing Wang, Email: 798999717@qq.com*

**Abstract:** As an essential component of the digital economy, digital finance is gradually becoming a key driver of cost reduction, efficiency enhancement, and high-quality development for enterprises. This paper analyzes 4,054 panel data points from 588 listed manufacturing firms in Dongguan spanning 2015 to 2024. Using financial and innovation R&D data from the CSMAR database, and employing SPSS for multiple linear regression analysis, the study empirically investigates the mechanisms and pathways through which digital finance influences enterprise cost optimization. The results reveal: (1) digital finance and soft technology investments significantly increase total asset turnover, validating their role in enhancing resource allocation efficiency; (2) R&D investment exhibits dual effects—while the R&D expenditure ratio positively affects firm efficiency, excessive allocations to R&D personnel and direct inputs impose short-term cost burdens; (3) new productive forces exhibit a phased negative impact, suggesting that digital infrastructure has not yet been fully converted into cost advantages; (4) financial structure and profitability form the core support for cost optimization. The study provides empirical evidence for digital transformation in the manufacturing industry and offers policy recommendations for both enterprises and government bodies.
**Keywords:** Digital finance; Cost optimization; Manufacturing industry; Dongguan; Empirical analysis

## 1 INTRODUCTION

### 1.1 Research Background

Since 2015, China's digital economy has experienced sustained and rapid growth, with technologies such as fintech, artificial intelligence, and blockchain continuously reshaping the traditional financial service landscape. The widespread application of digital finance has not only revolutionized corporate financing models but also enhanced firms' operational efficiency and resource allocation capabilities through big data and information platforms.

At the same time, the manufacturing sector is facing rising raw material costs, energy price fluctuations, and increasing labor expenses, making traditional cost control methods less effective. As a key manufacturing hub in the Pearl River Delta, Dongguan has taken the lead nationally in industrial upgrading and digital transformation. However, many manufacturing firms still lag in terms of depth of digital finance adoption and the development of effective cost governance mechanisms. Whether digital finance can truly drive cost optimization in manufacturing enterprises has become a critical issue for both academic research and industrial practice.

### 1.2 Research Significance

Theoretically, this study extends the research boundary of digital finance and corporate cost governance by uncovering the micro-level mechanisms through which digital finance empowers resource allocation and cost control within enterprises. Practically, the findings offer empirical evidence to support digital transformation in manufacturing, not only in Dongguan but also nationwide. This contributes to promoting the integration of operations and finance as well as the implementation of intelligent and collaborative management systems.

## 2 LITERATURE REVIEW

### 2.1 International Research Progress

Foreign scholars have conducted extensive studies on the relationship between digital finance and firm performance, focusing mainly on financing constraints, innovation efficiency, and productivity improvement. Allen et al. found that digital finance reduces financing barriers and enhances liquidity for small and medium-sized enterprises by applying advanced information technologies[1]. Beck emphasized that digital finance serves not only as a financial innovation but also as a strategic mechanism for industrial transformation, which enhances operational efficiency by lowering transaction costs, increasing information transparency, and improving capital allocation[2]. The OECD highlighted an indirect facilitation effect, showing that digital finance improves cost efficiency and productivity through innovation investment and supply chain coordination[3]. Similarly, Porter's Value Chain Theory posits that competitive advantage

depends on dynamic management of cost and differentiation, and the integration of digital technologies allows firms to achieve refined cost control and information sharing across the value chain[4].
In summary, international research views digital finance as a bridge connecting financial efficiency, innovation efficiency, and production efficiency, with mechanisms manifested in reduced financing costs, optimized innovation-resource allocation, and intelligent management decisions.

### 2.2 Domestic Research Progress

In China, the study of digital finance has expanded rapidly, emphasizing its role in innovation, cost governance, and industrial upgrading. Zhou Lian argued that digital finance promotes inclusive finance and structural transformation, serving as a key channel for strengthening the link between finance and the real economy[5]. Chen Wenling noted that by breaking information barriers, digital finance improves firms' access to financial data and enhances capital utilization efficiency[6]. Liu Yingchun proposed that digitalized management is a vital approach to cost optimization, as it enables real-time and visualized operational control[7]. Li Xiaoyan confirmed the mediating effect of digital finance on operational efficiency—it facilitates innovation input and alleviates financial constraints, indirectly improving profitability. Recent studies have also integrated digital finance with regional development[8]. Sun Jianbo found a significant positive correlation between digital finance development and manufacturing productivity at the regional level, indicating that regions with more advanced digital financial ecosystems tend to achieve higher resource allocation efficiency[9].
Overall, domestic literature has explored the importance of digital finance across policy, industry, and enterprise levels, yet there remains a lack of micro-level empirical evidence on cost governance. Most studies emphasize innovation or financing effects, with limited investigation into the mechanism through which digital finance influences cost efficiency via innovation inputs and financial structures.

### 2.3 Summary and Research Innovation

The existing literature provides a strong theoretical foundation for studying digital finance's impact on firm efficiency; however, three main gaps remain: (1) Macro bias – Research tends to focus on macroeconomic or sectoral effects, lacking enterprise-level empirical verification. (2) Incomplete mechanism analysis – The pathways linking digital finance to cost efficiency remain underexplored, particularly regarding innovation and financial structure as mediating channels. (3) Regional limitations – Few studies examine key manufacturing clusters such as Dongguan, limiting understanding of regional heterogeneity.

## 3 RESEARCH DESIGN

### 3.1 Data Sources and Sample Selection

This study focuses on manufacturing enterprises listed in Dongguan, Guangdong Province, during the period 2015–2024. All data were obtained from the China Stock Market & Accounting Research (CSMAR) database, including financial indicators, R&D input metrics, and firm characteristics. Financial firms and observations with excessive missing values were excluded. Continuous variables were winsorized at the 1st and 99th percentiles to mitigate the influence of outliers. The final balanced panel comprises 588 listed manufacturing firms with 4,054 firm-year observations, providing strong representativeness of the Pearl River Delta's digital transformation and cost governance dynamics.

### 3.2 Variable Design

***3.2.1Dependent variable***
Cost Efficiency (ATO)
Measured by Total Asset Turnover (ATO), which reflects how efficiently firms utilize assets to generate revenue:

$$ATO=\frac{\text{Operating Revenue}}{\text{Total Assets}} \tag{1}$$

A higher ATO indicates better asset utilization and stronger cost-control capability.
***3.2.2 Core independent variables***
To empirically examine the impact of digital finance and innovation activities on firms' technological upgrading, we construct a set of independent variables capturing different dimensions of innovation input and new productive forces. Specifically, we consider soft technology investment, R&D intensity, the proportion of R&D staff, the ratio of direct R&D investment, and the overall level of new productive forces. The definitions, symbols, and measurement methods of these independent variables are summarized in Table 1.

**Table 1** Definitions and Measurements of Independent Variables (IV)

| Variable Name | Symbol | Definition & Measurement |
|---|---|---|
| Soft Technology Investment (DF) | DF | Ratio of digital and IT expenditures to total operating cost, capturing the level of digital finance adoption. |
| R&D Intensity (RDE) | RDE | R&D expenditure as a proportion of operating revenue, measuring innovation investment strength. |
| R&D Staff Proportion (RDSTAFF) | RDSTAFF | Share of R&D employees in total staff, indicating the structure of human-capital investment in innovation. |
| Direct R&D Investment Ratio (RDDIRECT) | RDDIRECT | Direct R&D capital expenditure divided by total expenditures, representing innovation capital structure. |
| New Productive Forces Level (NP) | NP | Composite index based on automation, intelligent equipment, and digital system coverage, measuring technological upgrading. |

#### *3.2.3 Control variables*

In addition to the above independent variables, this study also incorporates several firm-level characteristics as control variables to isolate the net effect of digital finance and innovation inputs on technological upgrading. Specifically, we control for financial leverage, liquidity, inventory management efficiency, and profitability, which may simultaneously influence firms' investment decisions and performance outcomes. The definitions and measurements of these control variables are presented in Table 2.

**Table 2** Definition and Measurement of Control Variables (CV)

| Variable Name | Symbol | Definition |
|---|---|---|
| Leverage Ratio (LEV) | LEV | Total liabilities divided by total assets, reflecting financial leverage. |
| Cash Ratio (CASH) | CASH | Cash and cash equivalents divided by current liabilities, indicating liquidity. |
| Inventory Turnover (INV) | INV | Operating cost divided by average inventory, reflecting efficiency of inventory management. |
| Return on Assets (ROA) | ROA | Net profit divided by total assets, capturing profitability. |

### 3.3 Model Construction

This study employs a multiple linear regression model to assess the impact of digital finance and innovation variables on cost efficiency, specified as:

$$\mathbf{ATOi}=\beta_0+\beta_1\mathbf{DF}_i+\beta_2\mathbf{RDE}_i+\beta_3\mathbf{RDSTAFF}_i+\beta_4\mathbf{RDDIRECT}_i+\beta_5\mathbf{NP}_i+\beta_6\mathbf{LEV}_i+\beta_7\mathbf{CASH}_i+\beta_8\mathbf{INV}_i+\beta_9\mathbf{ROA}_i+\varepsilon_i$$

Where $ATO_i$ denotes firm i's cost-efficiency indicator;$DF_i$ represents digital-finance input; $RDE_i$， $RDSTAFF_i$ and $RDDIRECT_i$ correspond to innovation inputs;$NP_i$ measures the level of new productive forces;$LEV_i$, $CASH_i$, $INV_i$ and $ROA_i$ serve as control variables;and $\varepsilon_i$ is the error term.

All estimations were performed in SPSS 27.0, with variance inflation factor (VIF) tests conducted for multicollinearity and F/t-tests applied for overall and individual significance.

## 4 EMPIRICAL RESULTS AND ANALYSIS

### 4.1 Model Fit Evaluation

Based on 4,054 firm-year observations from 588 listed manufacturing enterprises in Dongguan over the period 2015–2024, multiple linear regression analysis was conducted using SPSS 27.0 to assess the impact of digital finance, R&D investment, and new productive forces on cost efficiency (measured by total asset turnover).

**Table 3** Model Summary[b]

| | | | | | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | 0.610[a] | 0.372 | 0.371 | .284657229 | 0.372 | 266.196 | 9 | 4044 | 0.000 |

**Note:** a: Predictors: (Constant), Return on Assets (ROA), Soft Technology Investment (DF), Inventory Turnover (INV), Direct R& D Investment Ratio (RDDIRECT), Cash Ratio (CASH), R&D Staff Proportion (RDSTAFF), R& D Intensity (RDE), Leverage Ratio (LEV), New Productive Forces Level (NP).
b: Dependent Variable: Total Asset Turnover (ATO).

As shown in Table 3, the model exhibits a correlation coefficient (R) of 0.610, indicating a strong linear relationship between the independent and dependent variables. The coefficient of determination ($R^2$) is 0.372, meaning that the nine independent variables jointly explain 37.2% of the variance in the dependent variable. The adjusted $R^2$ is 0.371, suggesting the model has strong explanatory power and stability. The F-statistic is 266.196 with a significance level of 0.000 ($p < 0.001$), thus rejecting the null hypothesis that all regression coefficients are zero.

This result demonstrates that digital finance, R&D activities, and financial structure variables significantly contribute to explaining enterprise asset utilization efficiency. In other words, cost efficiency in manufacturing is not driven by a single factor but arises from the coordinated influence of digital investment, innovation activities, and financial structure.

### 4.2 Analysis of Variance (ANOVA)

From the results in Table 4, the regression sum of squares accounts for 37.2% of the total sum of squares, aligning with the $R^2$ value and indicating a high degree of explanatory power. The F-test passed the 1% significance threshold (p = 0.000), confirming that the overall model is statistically significant.

The statistical significance of the model suggests that digital finance and innovation investments have a systematic and stable impact on asset turnover. Differences in efficiency across firms are structural rather than random, determined by differences in input structure and innovation levels.

**Table 4** ANOVA[a]

| | Model | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 194.128 | 9 | 21.570 | 266.196 | 0.000[b] |
| | Residual | 327.684 | 4044 | 0.081 | | |
| | Total | 521.813 | 4053 | | | |

**Note:** a: Dependent Variable: Total Asset Turnover (ATO).
b: Predictors: (Constant), Return on Assets (ROA), Soft Technology Investment (DF), Inventory Turnover (INV), Direct R&D Investment Ratio (RDDIRECT), Cash Ratio (CASH), R&D Staff Proportion (RDSTAFF), R&D Intensity (RDE), Leverage Ratio (LEV), New Productive Forces Level (NP)

### 4.3 Regression Coefficients and Variable Significance

As shown in Table 5, all variables are statistically significant at the 1% level.

**Soft Technology Investment (Digital Finance):**Exhibits the largest standardized coefficient (β = 0.421), indicating the strongest positive impact on cost efficiency. The implementation of digital financial systems significantly improves fund allocation and asset turnover.

**R&D Intensity:**Has a significantly positive effect on total asset turnover, confirming that investment in innovation improves resource utilization and long-term competitiveness.

**R&D Staff Proportion and Direct R&D Investment:**Both exhibit negative coefficients, suggesting that the short-term costs of innovation—such as personnel and capital expenditure—can temporarily reduce cost efficiency. This highlights the "short-term cost effect" of R&D investment.

**New Productive Forces:** Also shows a negative coefficient, indicating that the initial phase of technological transformation and automation increases fixed investment. However, these costs are expected to be offset by efficiency gains in the long run.

**Financial Structure Variables:**Leverage ratio, inventory turnover, and return on assets are all positively associated with total asset turnover, implying that sound financial leverage and profitability are key to improving cost efficiency.

In summary, the model reveals a dual-track mechanism of "digital empowerment + innovation-driven" cost control: digital finance enhances transparency and capital efficiency, while R&D investment and technological transformation generate long-term spillover effects, collectively improving the enterprise's cost competitiveness.

**Table 5** Coefficients[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Tolerance | VIF |
| 1 | (Constant) | 0.382 | 0.017 | | 22.255 | 0.000 | | |
| | R&D Intensity (RDE) | 0.001 | 0.000 | 0.053 | 4.106 | 0.000 | 0.941 | 1.062 |
| | Soft Technology Investment (DF) | 71.027 | 2.774 | 0.421 | 25.603 | 0.000 | 0.573 | 1.744 |
| | R&D Staff Proportion (RDSTAFF) | -0.002 | 0.000 | -0.074 | -5.597 | 0.000 | 0.886 | 1.129 |
| | Direct R&<br>D Investment Ratio (RDDIRECT) | 0.103 | 0.009 | 0.217 | 11.346 | 0.000 | 0.426 | 2.346 |

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | B | Std. Error | Beta | | | Tolerance | VIF |
| New Productive Forces Level (NP) | -14.601 | 0.835 | -0.381 | -17.489 | 0.000 | 0.328 | 3.050 |
| Leverage Ratio (LEV) | 0.519 | 0.032 | 0.251 | 16.291 | 0.000 | 0.652 | 1.533 |
| Cash Ratio (CASH) | -0.020 | 0.003 | -0.092 | -6.222 | 0.000 | 0.718 | 1.393 |
| Inventory Turnover (INV) | 0.003 | 0.001 | 0.083 | 6.633 | 0.000 | 0.980 | 1.021 |
| Return on Assets (ROA) | 1.777 | 0.066 | 0.375 | 27.001 | 0.000 | 0.807 | 1.239 |

**Note:** a: Dependent Variable: Total Asset Turnover (ATO).

### 4.4 Multicollinearity Diagnosis

**Table 6** Collinearity Diagnostics[a]

| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | (Constant) | R&D Intensity (RDE) | Soft Technology Investment (DF) | R&D Staff Proportion (RDSTAFF) | Direct R&D Investment Ratio (RDDIRECT) | New Productive Forces Level (NP) | Leverage Ratio (LEV) | Cash Ratio (CASH) | Inventory Turnover (INV) | Return on Assets (ROA) |
| 1 | 1 | 4.448 | 1.000 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 |
| | 2 | 1.271 | 1.870 | 0.00 | 0.06 | 0.00 | 0.00 | 0.10 | 0.04 | 0.00 | 0.05 | 0.03 | 0.09 |
| | 3 | 1.046 | 2.062 | 0.00 | 0.69 | 0.00 | 0.00 | 0.03 | 0.01 | 0.00 | 0.00 | 0.02 | 0.02 |
| | 4 | 0.856 | 2.279 | 0.00 | 0.04 | 0.10 | 0.00 | 0.08 | 0.00 | 0.01 | 0.26 | 0.01 | 0.01 |
| | 5 | 0.734 | 2.462 | 0.00 | 0.00 | 0.00 | 0.03 | 0.02 | 0.00 | 0.00 | 0.07 | 0.81 | 0.01 |
| | 6 | 0.615 | 2.689 | 0.00 | 0.09 | 0.16 | 0.13 | 0.00 | 0.02 | 0.02 | 0.00 | 0.00 | 0.22 |
| | 7 | 0.558 | 2.824 | 0.00 | 0.08 | 0.10 | 0.00 | 0.03 | 0.02 | 0.02 | 0.21 | 0.05 | 0.34 |
| | 8 | 0.304 | 3.824 | 0.01 | 0.02 | 0.02 | 0.59 | 0.02 | 0.02 | 0.09 | 0.08 | 0.06 | 0.11 |
| | 9 | 0.125 | 5.961 | 0.01 | 0.00 | 0.60 | 0.15 | 0.71 | 0.87 | 0.01 | 0.00 | 0.00 | 0.00 |
| | 10 | 0.044 | 10.044 | 0.96 | 0.01 | 0.00 | 0.07 | 0.00 | 0.02 | 0.84 | 0.31 | 0.01 | 0.18 |

**Note:** a: Dependent Variable: Total Asset Turnover (ATO).

As shown in Table 6, all variance inflation factor (VIF) values are below 4 and all tolerance values exceed 0.3, indicating that multicollinearity is not a concern. The independent variables maintain sufficient orthogonality.This confirms that digital finance, R&D input, and financial structure contribute independently to cost efficiency improvements, with limited redundancy among explanatory variables.

### 4.5 Residual Analysis and Robustness Testing

From Table 7, the mean of residuals is close to zero and the standard deviation of standardized residuals is approximately 1, satisfying the assumption of normality. The prediction errors are random and free from systemic bias.

**Table 7** Residuals Statistics[a]

| | Minimum | Maximum | Mean | Std. Deviation | N |
|---|---|---|---|---|---|
| Predicted Value | -0.70622998 | 2.76746249 | 0.67534876 | 0.218854919 | 4054 |
| Residual | -1.187158465 | 2.267492533 | 0.000000000 | 0.284341001 | 4054 |
| Std. Predicted Value | -6.313 | 9.559 | 0.000 | 1.000 | 4054 |
| Std. Residual | -4.170 | 7.966 | 0.000 | 0.999 | 4054 |

**Note:** a: Dependent Variable: Total Asset Turnover (ATO).

Figure 1 Shows a symmetric bell-shaped distribution with a mean near zero (3.10E-15) and a standard deviation of 0.999, indicating the residuals follow a normal distribution. Efficiency differences across firms primarily reflect actual operational differences rather than model misspecification.

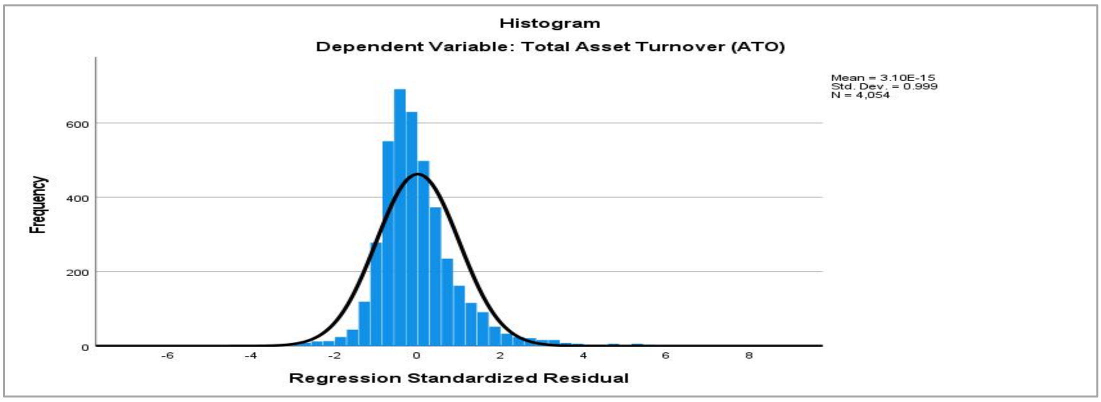


**Figure 1** Histogram of Standardized Residuals

Figure 2 Observations are closely aligned with the 45-degree diagonal line, confirming good fit with the theoretical normal distribution. The model satisfies the normality assumption, enhancing the reliability and generalizability of statistical inference.

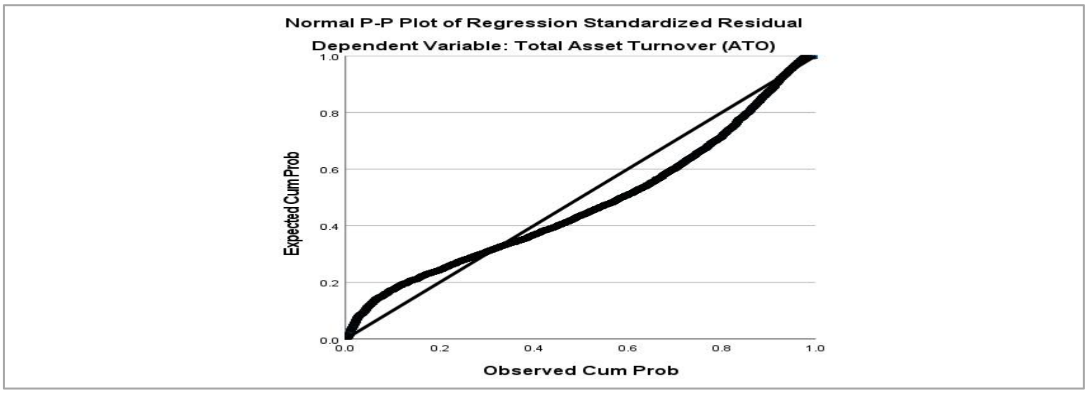


**Figure 2** P-P Plot of Residuals

Figure 3 Points are randomly scattered around the horizontal axis without funnel-shaped or curved patterns, indicating no heteroskedasticity or autocorrelation. The model structure is robust, confirming that the effects of digital finance, innovation input, and financial structure on firm efficiency are linear and reliable.

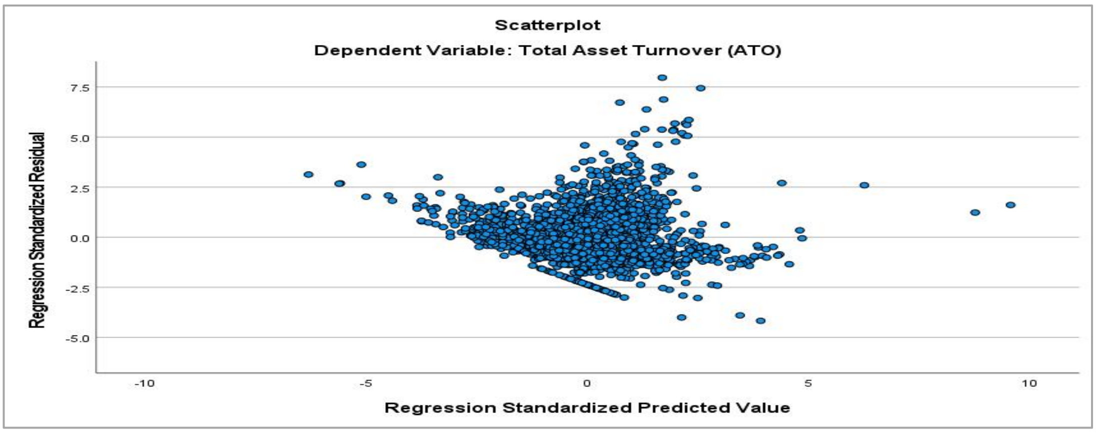


**Figure 3** Scatter Plot of Standardized Residuals

## 5 CONCLUSIONS AND POLICY RECOMMENDATIONS

### 5.1 Research Conclusions

Based on an empirical analysis of 4,054 observations from 588 listed manufacturing enterprises in Dongguan between 2015 and 2024, this study constructed a cost efficiency influence model centered on four core components: digital finance, R&D investment, new productive forces, and financial structure. Regression results obtained via SPSS support the following main conclusions:

***5.1.1 Digital finance significantly improves cost efficiency and serves as a core driver of cost reduction and efficiency enhancement***

The regression results show that the digital finance variable (soft technology) has the highest standardized coefficient among all predictors ($\beta = 0.421$, $p < 0.001$), indicating its dominant role in promoting cost efficiency. Digital finance facilitates optimal capital allocation by improving financing accessibility, reducing transaction costs, and enhancing fund liquidity. This, in turn, boosts asset turnover and business performance in manufacturing enterprises.

***5.1.2 Structural differences in R&D investment lead to phased effects on cost efficiency***

R&D expenditure ratio exhibits a significant positive impact on cost efficiency ($\beta = 0.053$), suggesting that innovation funding enhances long-term competitiveness and resource efficiency. However, both the proportion of R&D staff ($\beta = -0.074$) and direct R&D investment ratio ($\beta = -0.217$) show significant negative effects, indicating that high upfront personnel and capital costs during early-stage innovation may temporarily increase operational costs. This dual nature of innovation activities—boosting output while potentially suppressing short-term efficiency—represents the "double-edged sword" effect of R&D.

***5.1.3 New productive forces hinder short-term but promote long-term cost optimization***

The coefficient for the new productive forces variable is negative ($\beta = -0.381$), reflecting that in the early stages of digital and smart transformation, firms face significant costs related to technological upgrades and equipment investments. However, such "pain-period" investments lay the foundation for future cost savings and efficiency gains, consistent with the classic "investment-before-return" diffusion curve of emerging technologies.

***5.1.4 Financial structure and profitability provide stable support for cost governance***

Both the leverage ratio ($\beta = 0.251$) and return on assets ($\beta = 0.375$) are significantly positive, indicating that moderate financial leverage and solid profitability help enhance capital utilization and cost management. Conversely, a high cash ratio ($\beta = -0.092$) negatively impacts investment efficiency, suggesting that overly conservative liquidity management may hinder optimal capital deployment.

***5.1.5 Inventory management is a crucial internal factor for improving cost efficiency***

Inventory turnover shows a positive association with cost efficiency ($\beta = 0.083$), confirming that manufacturing enterprises can reduce capital occupancy and warehousing costs through optimized inventory structures and supply chain coordination—thus improving overall operational performance.

In summary: The coordinated development of digital finance, innovation investment, and financial structure forms the strategic path for enterprises to achieve cost optimization and operational efficiency.

## 5.2 Policy Recommendations

***5.2.1 Enhance digital financial infrastructure to extend services to manufacturing enterprises***

Governments should accelerate the construction of digital financial infrastructure and promote open access to financial data. Collaboration among banks, technology firms, and industry platforms should be encouraged to develop fintech tools such as smart credit, blockchain settlement, and cloud-based payment systems. These tools can provide manufacturing firms—especially SMEs—with efficient, secure, and low-cost financing. Policy support should also be expanded to ensure inclusive access to fintech in manufacturing and to build a multi-level industrial financial ecosystem.

***5.2.2 Establish incentive and risk-sharing mechanisms for R&D investment***

Given the delayed returns and uncertainties of R&D, enterprises face high upfront risks in innovation. It is recommended that governments improve the R&D tax deduction policy and increase fiscal incentives for innovation expenditures. Special funds and grants should be established to support technological innovation and the commercialization of R&D results. Moreover, the government should foster collaborative innovation networks involving government, industry, academia, and research institutions, and direct financial and research resources toward high-potential innovation clusters.

***5.2.3 Promote digital and intelligent transformation of manufacturing to create an "intelligence-driven efficiency" model***

To address the short-term cost pressure from building new productive forces, a gradual digitalization strategy should be encouraged. Government subsidies for smart manufacturing equipment, support for digital factory pilots, and funding for "Industrial Internet+" initiatives can help firms manage early-stage investment burdens. By promoting cloud manufacturing and data collaboration platforms, upstream and downstream firms can share information and integrate resources across the value chain, enhancing overall cost control capacity.

***5.2.4 Optimize enterprise financial structures to improve capital allocation efficiency***

Governments and financial institutions should guide enterprises toward rational debt levels and establish capital efficiency-oriented financing evaluation mechanisms. New financing instruments such as green bonds and technology bonds should be promoted to diversify funding channels. Enterprises should also be encouraged to enhance accounting informatization and build financial risk warning systems. This will support the implementation of integrated finance-operations (F&O) systems and improve the efficiency of fund utilization.

# 6 CONCLUSIONS

Overall, digital finance, innovation-driven development, and financial structure optimization jointly shape the cost-efficiency landscape of manufacturing enterprises. Their interplay forms a dynamic system of "digital empowerment – innovation-driven – financial coordination." In the digital economy era, only by leveraging the catalytic power of digital finance, strengthening innovation capacity, and establishing robust financial management systems can enterprises shift from “cost advantages” to “efficiency advantages” within global value chains.

The experience of Dongguan’s manufacturing sector shows that digital finance is more than a technological innovation; it is a key driver of industrial upgrading and regional economic resilience. Looking forward, as big data and artificial intelligence become more deeply integrated, enterprise cost governance will evolve toward intelligent, networked, and collaborative systems—injecting long-term momentum into the high-quality development of China’s manufacturing industry.

**COMPETING INTERESTS**

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

**FUNDING**

This study was supported by the Key Project of the 2024 Annual Research Program (Humanities and Social Sciences) by Guangdong University of Science and Technology(GKY-2024KYZDW-5);
Humanities and Social Sciences Project of the 2025 Annual Research Program by Guangdong University of Science and Technology (Project No. GKY-2025KYYBW-41).

**REFERENCES**

[1] Allen F, Demirguc-Kunt A, Klapper L, et al. The foundations of financial inclusion: Understanding ownership and use of formal accounts. Journal of Financial Intermediation, 2016, 27: 1-30.
[2] Beck T. Finance and structural transformation: Evidence from FinTech. Journal of Economic Perspectives, 2018, 32(2): 77-100.
[3] OECD. Digital transformation in SMEs. OECD Studies on SMEs and Entrepreneurship, OECD Publishing, Paris, 2021.
[4] Porter M E. The competitive advantage of nations. New York: Free Press, 1990.
[5] Zhou Lian. The role of digital finance in promoting inclusive finance and structural transformation. Journal of Financial Research, 2019(3): 45-52.
[6] Chen Wenling. Digital finance and capital utilization efficiency of enterprises. China Economic Times, 2020(5): 4.
[7] Liu Yingchun. Digital management as a path to cost optimization in Chinese manufacturing. Management Review, 2020(12): 101-108.
[8] Li Xiaoyan. The mediating effect of digital finance on operational efficiency: Evidence from Chinese enterprises. Finance and Trade Research, 2023, 34(6): 112-123.
[9] Sun Jianbo. Digital finance development and manufacturing productivity at the regional level: Empirical evidence from China. Journal of Industrial Economics, 2022, 40(4): 56-67.