ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on the impact and application mechanism of digitalization empowerment of Chinese enterprises along the "Belt and Road" to improve international performance

Xiaoli Zhang[1,*]

[1] Anhui Technical College of Mechanical and Electrical Engineering Anhui, 241000, Anhui, China

SUMMARY: *Under the background of market system differences, supply chain fluctuations and increasing complexity of cross-border collaboration along the "Belt and Road", this paper constructs a digital empowerment and internationalization performance analysis framework based on 1824 enterprise-year samples, and uses text mining, multi-source data fusion, machine learning and dual machine learning methods. To identify the impact and mechanism of digital empowerment on the internationalization performance of Chinese enterprises. The results showed that the digital empowerment index was significantly positively correlated with internationalization performance, and the correlation coefficient was 0.472. The digitization empowerment coefficient in the baseline regression is 0.286, which is still 0.219 after including enterprise characteristics and external environment variables, and the net effect of dual machine learning is 0.207. Heterogeneity analysis shows that the coefficient of high-tech enterprises group increases from 0.139 to 0.258, with an increase of 0.119. Research shows that digital empowerment can significantly improve the internationalization performance of enterprises by optimizing resource allocation, enhancing organizational collaboration and improving governance capabilities. This paper expands the research path of enterprise internationalization from the perspective of computer technology embedding, and also provides methodological support and practical reference for the high-quality operation of Chinese enterprises along the market.*

KEYWORDS: *Digital empowerment; International performance; Multi-source data fusion; Machine learning*

1 Introduction

In the context of the reconstruction of the global industrial chain, the expansion of the digital economy and the rising uncertainty of international economic and trade, the countries along the "Belt and Road" have become an important space for Chinese enterprises to carry out cross-border investment, trade cooperation and value chain embedding. However, there are system differences, market fluctuations, increasing difficulties in supply chain coordination and increasing risks of transnational operation. It also makes the internationalization development of enterprises face practical problems such as unstable performance, lagging response and insufficient resource allocation efficiency. Wu and Yin (2025) proposed that the "Belt and Road" is not only a regional cooperation initiative, but also affects the way and depth of Chinese enterprises' participation in the global value chain at the institutional level

*0125000366@ahcme.edu.cn
https://doi.org/10.65102/is2026296
Published: 30 April, 2026

[20]. Zhang et al. (2024) proposed that the productivity of OFDI under the "Belt and Road" scenario is significantly different, and the improvement of internationalization quality and performance of enterprises has obvious structural characteristics [3]. Wen et al. (2025) proposed that the Belt and Road Initiative has an important impact on the sustainable development performance of Chinese enterprises, indicating that the operation along the belt and Road has shifted from scale expansion to quality improvement [19]. This shows that the real demand of Chinese enterprises along the Belt and Road is no longer in "going global", but in how to achieve a higher level of international performance in a complex external environment.

Digital empowerment provides a new research perspective to solve this problem. Dong et al. (2024) proposed that digitalization can significantly affect the export performance of enterprises, but its effect is constrained by the common conditions of enterprises and external situations [1]. Fan et al. (2024) proposed that digital development has a significant role in promoting OFDI of Chinese listed enterprises [2]; Wu et al. (2024) proposed that cross-border e-commerce and trade digitalization can enhance enterprise export resilience [8]. It can be seen that digital empowerment not only helps to improve the efficiency of enterprise information acquisition, market identification and cross-border collaboration, but also may further improve the internationalization performance by optimizing resource allocation, enhancing organizational interaction and improving risk response ability. Therefore, the research on digital empowerment of Chinese enterprises along the "Belt and Road" has not only clear practical relevance, but also important theoretical value.

At the same time, it is necessary to integrate computer technology into the study of enterprise internationalization. Appiah et al. (2025) proposed that digitalization has been deeply embedded in the early internationalization process of enterprises, and relevant research needs to further strengthen the data-driven identification and technical explanation framework [13]. Cahen et al. (2025) proposed that global digital capabilities have become an important foundation for contemporary international enterprises to build competitive advantages [16]. This means that traditional research methods relying on static financial indicators and single questionnaire data are difficult to fully reveal the true mechanism of digital empowerment. It is urgent to use computer technologies such as text mining, multi-source data fusion, machine learning and intelligent analysis to improve the accuracy of enterprise digital level identification, performance measurement and mechanism testing. Based on this, this paper takes the Chinese-funded enterprises along the "Belt and Road" as the research object, analyzes the influence relationship, action path and application mechanism of digital empowerment on internationalization performance, and empathizes computer technology methods in the research to construct a framework of digital empowerment identification and internationalization performance evaluation. Compared with the existing research, the innovation of this paper is mainly reflected in focusing on the specific scenario of Chinese-funded enterprises along the "Belt and Road", strengthening the mechanism link between digital empowerment and international performance, and improving the technical content and practical value of the research from three levels of data identification, method integration and application interpretation.

2 Theoretical basis and influence mechanism analysis

Digital empowerment is not a simple superposition of information systems, platform tools or online business, but a systematic process in which enterprises rely on data resources, digital platforms, algorithmic models and network infrastructure to restructure and optimize research and development, procurement, production, logistics, marketing, investment and service links, so as to improve the efficiency of resource allocation, organizational collaboration capabilities and cross-border response speed. Correspondingly, internationalization performance should not only be understood as the growth of overseas sales or investment scale, but should be comprehensively reflected in the overall performance of enterprises in terms of export resilience, export stability, export sustainability, overseas investment efficiency, cross-border merger and acquisition effectiveness and international operation income. Wu and Xu (2025) proposed that digital innovation has become an important driving force for enterprise internationalization [12]; Yang et al. (2025) proposed that research on the internationalization of digital enterprises is shifting from traditional market entry analysis to research on digital capabilities, platform connectivity and network collaboration [14]. Li et al. (2025) proposed that digital platform-based enterprises show stronger characteristics of cross-border connection and resource integration in the process of internationalization [15]. Zhang et al. (2025) further pointed out that the collaboration between digital platform and internal digital capabilities has a significant supporting effect on the international expansion of enterprises [17]. It can be seen that the relationship between digital empowerment and internationalization performance is not isolated, but there is an internal coupling logic.

From the perspective of mechanism, the key reason why digital empowerment can improve international performance is that it changes the way enterprises handle cross-border complex information and organize international business activities. Hongliang et al. (2024) proposed that digital transformation can enhance the export resilience of enterprises [4]; Zhang et al. (2024) proposed that digital transformation can help improve the stability of enterprises' export [5]; Xu et al. (2025) proposed that digital transformation can prolong the duration of enterprises' export and further improve their financial performance [11]. Zhu (2025) pointed out that digital transformation has a significant role in promoting OFDI of Chinese enterprises [9]. Xuan and Yan (2025) proposed that there is a significant correlation between digital transformation, internationalization and cross-border M&A performance [10]. Together, these studies show that digital empowerment is not only reflected in the expansion of overseas business scale, but more importantly, by improving the efficiency of information identification, reducing transaction costs, enhancing cross-regional coordination and optimizing the decision-making process, promoting the internationalization of enterprises from extensive expansion to quality improvement.

The realization of this process is inseparable from the support of computer technology. Text mining can identify the characteristics of digital transformation from corporate annual reports, announcements, social responsibility reports and overseas business disclosures. Multi-source data fusion can integrate export, investment, M&A, finance, platform transactions and regional institutional environment data. Machine learning can help to identify the marginal impact, heterogeneous performance and nonlinear characteristics of digital empowerment on internationalization performance. So that the research breaks through the limitations of traditional single financial indicators and questionnaire data. Khurram et al. (2024) proposed that there is a linkage relationship between digital transformation and overseas investment performance and ESG performance [6]. Guo et al. (2024) proposed that digital transformation can improve export performance through ESG responsibility [7]. Capelleras et al. (2025) proposed that there may be a curve relationship between digitalization and export propensity, which is affected by institutional environmental factors [18]. This shows that computer technology is not only a means to realize digital empowerment, but also an important tool to identify its action path and boundary conditions.

On this basis, this paper argues that digital empowerment affects internationalization performance mainly through three intermediary paths: resource allocation efficiency, organizational collaboration capability and responsibility governance level. Firstly, digital empowerment can improve the perception ability of enterprises to overseas market information, customer needs and supply chain status, so as to optimize the allocation of investment, export and M&A resources. Second, digital platforms and intelligent systems can enhance information connectivity and process collaboration between headquarters and overseas branches, suppliers and customers, and improve international business efficiency. Third, digital transformation may also enhance legitimacy and recognition in host markets by improving ESG governance and digital innovation capabilities. At the same time, the performance effect of digital empowerment will also be adjusted by the difference of institutional environment, the level of digital infrastructure in the host country and the original internationalization ability of the enterprise. That is to say, the higher the degree of digitalization does not necessarily bring the same range of performance improvement, and the strength of its effect also depends on the institutional and technical conditions of the enterprise. Based on the above analysis, this paper proposes the following research hypothesis: digital empowerment has a significant positive impact on the internationalization performance of Chinese enterprises along the "Belt and Road", and this impact will be transmitted through mechanisms such as resource allocation, organizational collaboration and ESG responsibility, and will be jointly regulated by the application level of computer technology, the difference of institutional environment and digital infrastructure conditions. The comparison between related studies and this study is shown in Table 1.

Table 1: Comparison between some of the references and this study

Reference	Focus	Method	Implications
Hongliang et al. (2024) [4]	Digital transformation and export resilience	Empirical analysis based on Chinese enterprise samples	Indicates that digitalization helps enhance firms' recovery capability in international operations when facing external shocks
Zhang et al. (2024) [5]	Digital transformation and export stability	Panel data analysis of listed companies	Reveals that digitalization promotes the continuity and stability of firms' international operations
Khurram et al. (2024) [6]	Digital transformation, outward investment performance, and ESG	Association analysis of investment performance	Suggests that ESG may serve as an important mediating variable through which digitalization affects internationalization performance
Guo et al. (2024) [7]	Digital transformation, ESG responsibility, and export performance	Empirical econometric analysis	Reinforces the explanation that digitalization improves internationalization performance through responsibility governance
Xuan and Yan (2025) [10]	Digital transformation, degree of internationalization, and cross-border M&A performance	Empirical study in the context of cross-border M&A	Extends the impact of digitalization to more complex international market entry modes
Yang et al. (2025) [14]	Research progress on the internationalization of digital firms	Systematic literature review	Provides theoretical support for digital capability, platform logic, and network collaboration
Zhang et al. (2025) [17]	Digital platforms, internal digitalization, and SME internationalization	Mechanism analysis and theoretical synthesis	Indicates that the synergy between platformization and internal digitalization is an important driver of internationalization
Capelleras et al. (2025) [18]	Nonlinear relationship between digitalization and export propensity	Cross-country comparative study	Shows that the institutional environment moderates the strength of digitalization's impact on internationalization performance
This Study	Digital empowerment and internationalization performance of Chinese enterprises along the Belt and Road	Combination of text mining, multi-source data fusion, machine learning, and mechanism testing	Embeds computer technology into internationalization performance research and comprehensively identifies direct effects, mediating pathways, and modera

3 Research design and calculation method construction

3.1 Sample source, data collection and preprocessing

This paper takes the Chinese listed enterprises with export, foreign direct investment, cross-border merger and acquisition or overseas subsidiary layout along the "Belt and Road" as the initial sample. During the study period, the analysis data set is constructed according to "enterprise identification - cross-source capture - entity alignment - text calculation - feature fusion - cleaning and standardization", and the overall process is shown in Figure 1. Firstly, the sample identification crawled the enterprise annual report, ESG report, company announcement, overseas investment announcement, merger and acquisition disclosure, financial statements and regional operation data with the listed company code as the main key. Then, according to the name of the host country, the location of the project, the source of overseas income and the registration place of the subsidiary, whether the enterprise belongs to the operation sample along the "Belt and Road" is determined. The business identification function along the line is defined as follows:

$$B_{it} = \begin{cases} 1, & \text{x} \\ 0, & \text{Others} \end{cases} \tag{1}$$

x represents that the knot enterprise i has export, investment, merger and acquisition or overseas institution records in the countries along the Belt and Road in year t. Only firm-year observations with $B_{it}=1$ are retained for subsequent processing.

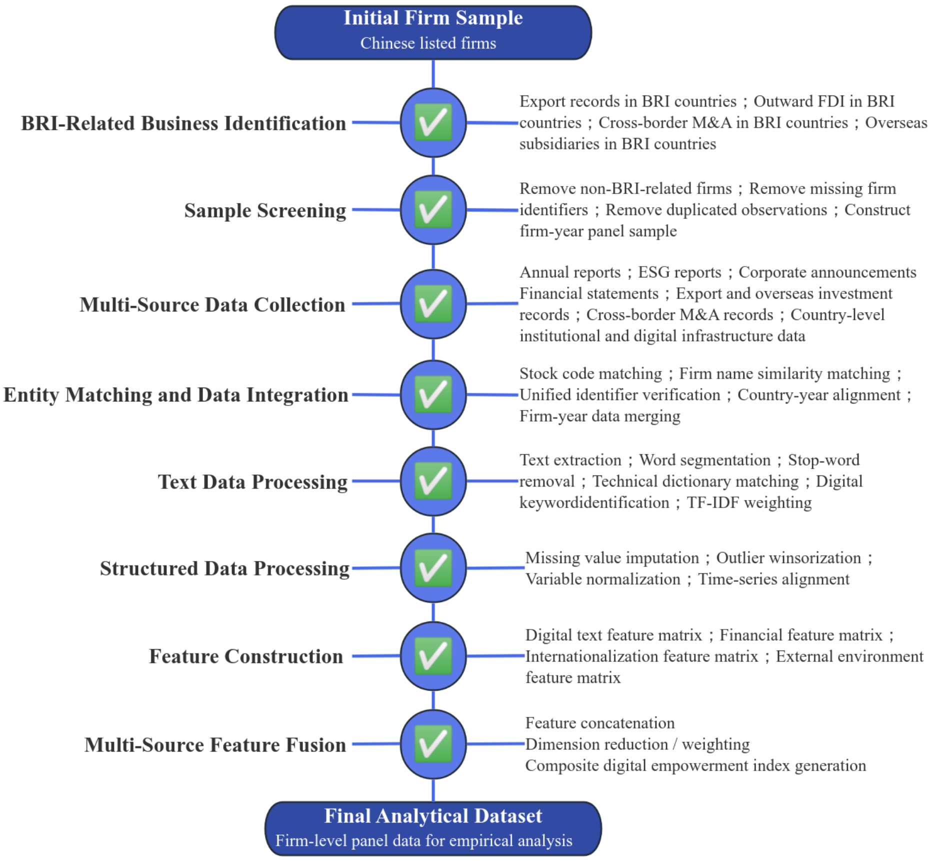

Figure 1: Flowchart of sample construction, data scraping, and preprocessing

After the cross-source crawling is completed, it is necessary to solve the problems of inconsistent enterprise names, inconsistent expression of overseas projects, and confusion of entities with the same name. In this paper, a three-level alignment strategy of "exact matching of securities code + similarity matching of enterprise name + unified social credit code verification" is adopted. For any two records a and b, construct an entity matching score:

$$S_{ab} = \omega_1 \cdot sim_{name}(a,b) + \omega_2 \cdot sim_{code}(a,b) + \omega_3 \cdot sim_{id}(a,b) \tag{2}$$

Here, $\omega1+\omega2+\omega3=1$, sim_{name} is the enterprise name edit distance similarity, sim_{code} is the stock code consistency index, and sim_{id} is the uniform social credit code consistency index. When $S_{ab} \geq \tau$, it is identified as the same subject; otherwise, it is entered into the manual review library. This step can reduce duplicate records and mismatch errors caused by cross-platform grasping. The sample sources, variable types and processing methods are shown in Table 2.

Table 2: Sample sources, variable types and preprocessing methods

Data Module	Main Fields	Data Form	Technical Processing	Output Result
Basic Firm Information	Firm name, firm code, industry, place of registration	Structured	Primary key matching, entity disambiguation	Master firm index table
Financial and Operational Data	Revenue, net profit, R&D, proportion of overseas revenue	Structured	Missing value imputation, winsorization, normalization	Financial feature matrix
International Business Data	Export records, OFDI, M&A amount, overseas subsidiaries	Semi-structured	Time alignment, country code mapping	Internationalization performance feature matrix
Annual Report and ESG Texts	Digitalization-related expressions, platform construction, algorithm application, data governance	Unstructured	Word segmentation, denoising, TF-IDF, vectorization	Digital text feature matrix
External Environment Data	Host-country institutional environment, digital infrastructure	Structured	Country-year matching	Moderating variable matrix

In the text data processing, this paper extracts the corpus from the chapters of annual reports such as "management discussion and analysis", "business review", "technological innovation", "overseas operation" and "digital construction". After unified transcoding to UTF-8, the noise is removed, and the table fragments, meaningless symbols, stop words, low-frequency words and repeated sentences are eliminated. Then, the enterprise digital dictionary and the "Belt and Road" international business dictionary are combined for word segmentation and word mapping. In order to improve the accuracy of technical feature recognition, the digital dictionary not only includes general words such as "digital transformation", "platform construction" and "intelligent manufacturing", but also includes computer technology words such as "cloud computing", "big data", "artificial intelligence", "algorithm optimization", "ERP", "CRM", "cross-border e-commerce system" and "supply chain collaboration platform". For the digitized text intensity of enterprise iii in year t, TF-IDF weighting is used to calculate:

$$TFIDF_{k,it} = tf_{k,it} \cdot \log\left(\frac{N}{df_k + 1}\right) \tag{3}$$

$$DTI_{it} = \frac{\sum_{k=1}^{m} \alpha_k \cdot TFIDF_{k,it}}{L_{it}} \tag{4}$$

where $tf_{k,it}$ is the term frequency of term k in enterprise i and year t, df_k is the number of documents containing the term, N is the total number of documents, αk is the term weight, Lit is the total number of terms in the document. This index is used to characterize the investment in digital technology, the data governance capability, and the degree of intelligent system embedding. The process of text processing and feature generation is shown in Figure 2.

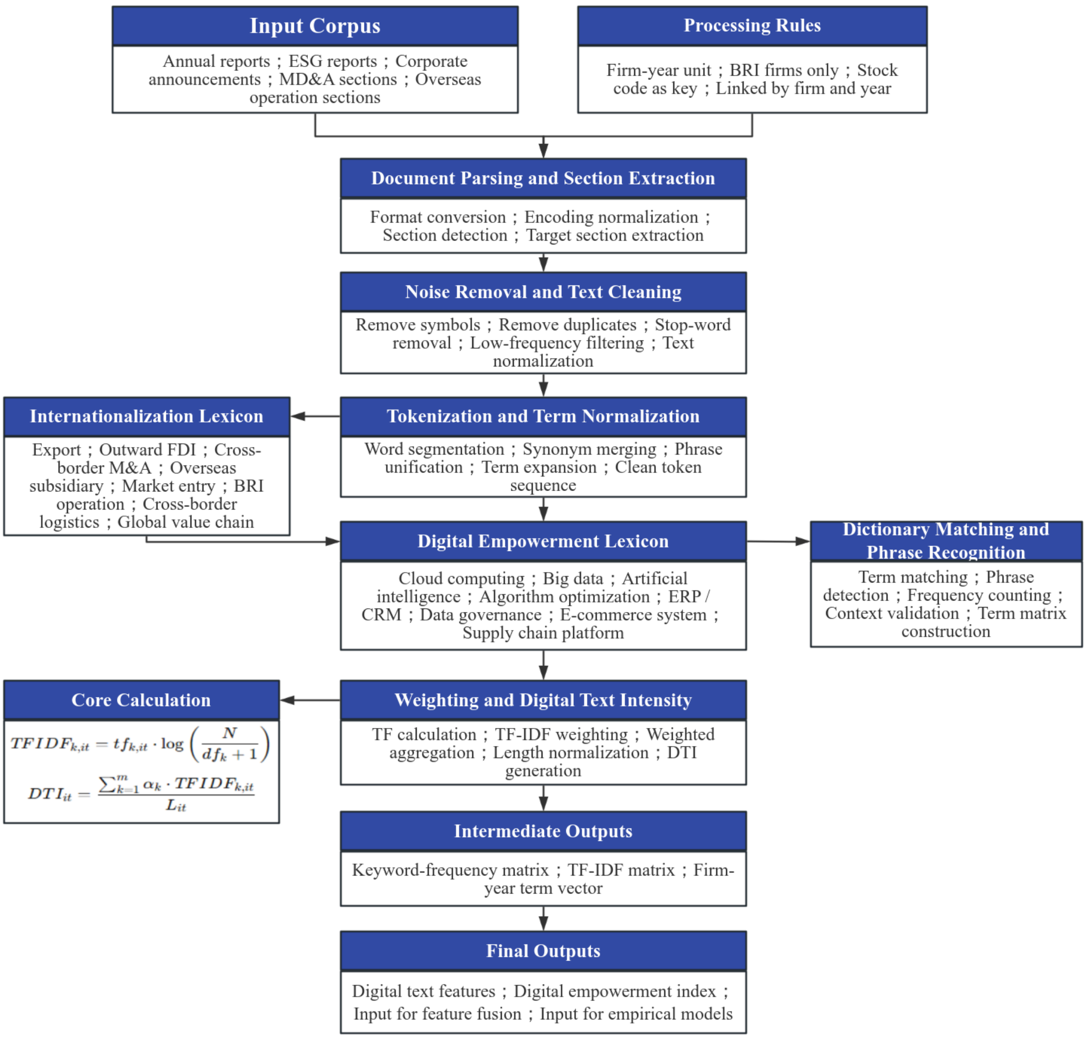

Figure 2: Illustration of text cleaning, term identification, and digital feature generation

Structured data preprocessing mainly includes missing correction, outlier handling, dimension unification and time alignment. The combination strategy of "k-nearest neighbor imputation + industry-annual mean backfilling" was used for missing values of continuous variables. For missing values for xit, we compute the between-sample distance:

$$d(i,j) = \sqrt{\sum_{r=1}^{p} (x_{ir} - x_{jr})^2} \tag{5}$$

Then the nearest K neighbors are selected for weighted interpolation:

$$\hat{x}_{it} = \frac{\sum_{j \in N_K(i)} \frac{1}{d(i,j) + \varepsilon} x_{jt}}{\sum_{j \in N_K(i)} \frac{1}{d(i,j) + \varepsilon}} \tag{6}$$

Here, ε is the smoothing term that prevents the denominator from being zero. Outliers are treated with bilateral 1% quantile tail reduction, and the original variable is denoted as x. The tail reduction result is as follows:

$$x^w = \min\{ \max(x, P_1), P_{99}\} \tag{7}$$

Among them, P_1 and $P_9$9 are the 1% and 99% quantiles, respectively. To eliminate dimensional differences, all continuous variables were further standardized by Z-score:

$$z_{it} = \frac{x_{it} - \mu_x}{\sigma_x} \tag{8}$$

Among them, the variables related to internationalization performance, digital characteristics and control variables are unified into the framework of the enterprise-year panel, and the external institutional environment and digital infrastructure indicators are matched according to the country-year dimension.

In order to avoid measurement bias caused by single source features, this paper conducts multi-source fusion on text feature matrix Tit, financial operation matrix Fit and international business matrix I_{it}, and constructs enterprise comprehensive feature vector:

$$Z_{it} = [\lambda_1 T_{it}, \lambda_2 F_{it}, \lambda_3 I_{it}] \tag{9}$$

Here, $\lambda_1+\lambda_2+\lambda_3=1$. Then, principal component extraction or entropy value weighting are used to generate the digital empowerment representation values and basic explanatory variables required for subsequent models. Using the principal components method, the QTH principal component score would be:

$$PC_{q,it} = a_{q1} z_{1,it} + a_{q2} z_{2,it} + \cdots + a_{qn} z_{n,it} \tag{10}$$

The overall score is written as follows:

$$DE_{it} = \sum_{q=1}^{Q} \rho_q \, PC_{q,it} \tag{11}$$

Here, ρ_q is the variance contribution rate. Through this process, text, financial and international business information can be uniformly mapped into the same computational space, which provides highly consistent input data for subsequent digital empowerment identification and internationalization performance modeling.

3.2 Identification of Enterprise Digital empowerment Level based on text Mining

After sample screening, text crawling and preprocessing, this paper further identifies the digital empowerment level of enterprises based on text mining methods. The process does not stop at simple keyword counting, but according to the order of "domain dictionary construction - semantic expansion - sentence level recognition - chapter weighting - enterprise annual aggregation", the annual report, ESG report, announcement and overseas business text are transformed into computable digital enabling indicators. The overall recognition process is shown in Figure 3. In order to ensure that the recognition object is consistent with the research topic, this paper only keeps the text fragments related to enterprise operation, technology deployment, platform construction, data governance, intelligent decision-making and cross-border business digitization, and uniformly maps them to the enterprise-year level for feature calculation.

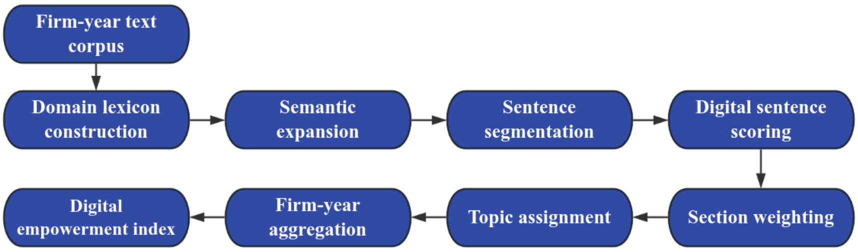

Figure 3: Flowchart of identifying the level of digital empowerment of enterprises based on text mining

Let the set of texts of firm i in year t be denoted $D_{it}=\{s_{1,it},s_{2,it},...,s_{n,it}\}$, where sj,it denote the JTH sentence or semantic unit. The first step in digitally enabling recognition is the construction of a domain dictionary. Based on the topics of enterprise digital transformation, platform system, data governance, algorithm application and cross-border digital business, this paper manually organizes the core seed word set K(0)={k1,k2,... km}, and then the pre-trained word vector model is used to semantically expand the highly relevant terms. For candidate word w and seed word k, the semantic similarity is defined as follows:

$$sim(w,k) = \frac{\mathbf{v}_w \cdot \mathbf{v}_k}{\|\mathbf{v}_w\|\|\mathbf{v}_k\|} \tag{12}$$

Here, v_w and v_k are the word vector representations of candidate and seed words, respectively. When $\max_{k\in K^{(0)}} sim\,(w,k) \geq \delta$, this candidate is incorporated into the extended dictionary K(1). Finally, the digitization empowering dictionary K is obtained, which includes six terms: infrastructure digitization, business process digitization, data governance, intelligent analysis, platform collaboration and cross-border digital operation. The purpose of this step is to reduce the identification bias caused by differences in enterprise expressions and avoid the omission problem caused by only relying on fixed keywords.

In the sentence-level recognition stage, this paper calculates both the word frequency weight and the semantic relevance probability for each sentence sj,it. Firstly, TF-IDF is used to measure the importance of digital terms in sentences. If the term frequency of term k in enterprise i, year t and sentence j is tfk,jit, and its inverse document frequency is $idf_k = \log\frac{N}{df_k+1}$, then the weighted value of the term is:

$$TFIDF_{k,jit} = tf_{k,jit} \cdot \log\left(\frac{N}{df_k + 1}\right) \tag{13}$$

where N is the total number of enterprise-year documents, and df_k is the number of documents in which term k occurs. Furthermore, the lexicon strength score of the sentence is obtained by summing the TF-IDF values of all digitized terms in the sentence according to the importance coefficient ωk:

$$L_{j,it} = \sum_{k\in K} \omega_k \cdot TFIDF_{k,jit} \tag{14}$$

It is still difficult to distinguish "digital actual deployment" from "general strategic expression" by only relying on dictionary matching, so this paper further introduces sentence vector semantic recognition. Suppose the sentences sj,it are encoded by the pre-trained language model to obtain vector representations hj,it, and the prototype vector of digitization positive class is c+, then the sentence-level digitization semantic probability is defined as follows:

$$P_{j,it} = \sigma\big(\mathbf{W}\mathbf{h}_{j,it} + b\big) \tag{15}$$

where σ(·) is the Sigmoid function, W and b are the classification layer parameters. In order to balance the interpretability and semantic recognition ability of rules, this paper adopts the hybrid scoring method of "dictionary strength + semantic probability" to comprehensively judge the degree of digitization empowerment of sentences:

$$Score_{j,it} = \lambda \cdot \frac{L_{j,it} - \min(L)}{\max(L) - \min(L)} + (1-\lambda) \cdot P_{j,it} \tag{16}$$

Here, λ∈[0,1] represents the fusion weight of the lexicon score and the semantic probability. If $Score_{j,it} \geq \tau s$, the sentence is identified as a digitally empowered valid sentence. Otherwise, it is regarded as low relevance or noise sentence. In this way, highly relevant expressions such as "deploy ERP system", "build cross-border supply chain collaboration platform", "carry out algorithm-based customer profiling" can be distinguished from low-information sentences that are only promotional in nature.

Due to the different information density in different chapters of an enterprise, this paper performs chapter weighting on sentence scores in the document aggregation stage. The enterprise text is divided into five chapters of "technology innovation", "operation and management", "overseas business", "strategic planning" and "social responsibility". Suppose that the chapter belonging to the sentence sj and it is c(j), and the corresponding chapter weight is ηc(j). Then the text digitization intensity index of enterprise i in year t is:

$$DTI_{it} = \frac{\sum_{j=1}^{n} \eta_{c(j)} \cdot Score_{j,it}}{\sum_{j=1}^{n} \eta_{c(j)}} \tag{17}$$

Among them, the chapters of "technological innovation", "overseas business" and "operation and management" are given higher weights to enhance the sensitivity of the indicators to real technology deployment and international business digitalization scenarios. In order to further identify the internal structure of digital empowerment, this paper adopts the topic mapping method and divides the effective sentences into six dimensions: digital infrastructure, platform collaboration, data governance, intelligent analysis, supply chain digitalization and cross-border business digitalization. Let the topic score of the r-th dimension in enterprise i and year t be Tr,it, then it can be written as follows:

$$T_{r,it} = \frac{\sum_{j=1}^{n} \mathbb{I}\,(z_{j,it} = r) \cdot Score_{j,it}}{\sum_{j=1}^{n} \mathbb{I}\,(z_{j,it} = r)} \tag{18}$$

Here, $z_{j,it}$ represent the topic category to which the sentence belongs, and I(·) is an indicator function. If a soft assignment based on topic probability is used, then the indicator function is replaced by the topic probability $\pi_{r,jit}$ is sufficient.

After obtaining the overall intensity index DTIit and the scores of each dimension Tr,it, this paper further constructs the comprehensive index of enterprise digital empowerment. First, standardize each index:

$$Z_{q,it} = \frac{X_{q,it} - \mu_q}{\sigma_q} \tag{19}$$

where $X_{q,it}$ represent the overall intensity or the score of a certain dimension, μq and σq are the mean and standard deviation of this index, respectively. Then, the entropy method is used to determine the weight of each dimension. Let the proportion of the QTH index in the sample after standardization be:

$$p_{q,it} = \frac{Z_{q,it}}{\sum_{i=1}^{N} \sum_{t=1}^{T} Z_{q,it}} \tag{20}$$

Then its information entropy is as follows:

$$e_q = -\frac{1}{\ln(NT)} \sum_{i=1}^{N} \sum_{t=1}^{T} p_{q,it} \ln(p_{q,it}) \tag{21}$$

The weights are defined as follows:

$$w_q = \frac{1 - e_q}{\sum_{q=1}^{Q} (1 - e_q)} \tag{22}$$

Accordingly, the comprehensive index of enterprise digital empowerment is obtained as follows:

$$DE_{it} = \sum_{q=1}^{Q} w_q \, Z_{q,it} \tag{23}$$

where Q contains the overall text intensity and each topic dimension indicator. The final DE_{it} is the measurement result of digitization empowerment level at the enterprise-year level, and serves as the core explanatory variable in the subsequent internationalization performance model.

In order to check the stability of the text recognition results, this paper adds two checks in the training and application stages. Firstly, a certain proportion of sentence samples from high-frequency enterprises were randomly selected for manual review, and the consistency rate between manual annotation and model judgment was calculated. Secondly, the smoothness test is carried out on the digital enabling index of the same enterprise in adjacent years. If there is an abnormal jump, the original text is backloaded to check whether there is a format capture error or a concentrated accumulation of non-operational expressions. After the above processing, the identification results of enterprise digital empowerment level can simultaneously reflect the explicit expression of text, semantic potential features and chapter structure differences, and be consistent with the enterprise-year multi-source database formed in Section 3.1.

3.3 Construction of internationalization performance evaluation index based on multi-source data fusion

In this paper, internationalization performance is defined as the comprehensive operation results of enterprises in the process of export, investment, merger and acquisition and overseas operation along the "Belt and Road". It not only pays attention to the internationalization scale, but also pays attention to the stability, sustainability, integration efficiency, growth quality and risk-adjusted return of internationalization. Based on this, this paper integrates the annual report division information, export business records, foreign direct investment data, cross-border merger and acquisition disclosure, overseas subsidiary distribution and host country environmental data at the enterprise-year level, and constructs a multi-source internationalization performance evaluation index system. The relevant indicators, data sources and measurement calibres are shown in Table 3.

Table 3: Internationalization performance evaluation index system

Dimension	Indicator Name	Symbol	Measurement Definition	Data Source	Direction
International Operating Intensity	Overseas Revenue Ratio	ORRit	Ratio of overseas operating revenue to total operating revenue	Annual reports, segment reports	Positive
International Market Coverage	Belt and Road Market Coverage	BCOVit	Standardized value of the number of Belt and Road countries in which the firm operates	Overseas subsidiaries, export and investment records	Positive
Export Continuity Capability	Export Duration	EDUit	Number of years with continuous export activities	Export records	Positive
Export Stability Capability	Export Stability Index	ESTit	Derived from the inverse transformation of export volatility	Export records	Positive
Outward Investment Expansion	OFDI Intensity	OFDIit	Standardized value of the number or amount of OFDI projects in Belt and Road countries in the current year	Outward investment records	Positive
M&A Integration Performance	Cross-border M&A Performance	MAPit	Degree of operational improvement or success rate after merger and acquisition completion	M&A announcements, financial data	Positive
Overseas Operational Efficiency	Overseas Return on Assets	OROAit	Ratio of overseas business profit to overseas assets	Segment reports, annual reports	Positive
International Growth Quality	Overseas Revenue Growth Rate	OGRit	Growth of current overseas revenue relative to the previous period	Annual reports, segment reports	Positive
International Business Resilience	Overseas Business Recovery Index	RESit	Degree of recovery in overseas revenue after external shocks	Annual reports, segment reports	Positive
Risk-adjusted Performance	Overseas Return-to-Risk Ratio	RAPit	Ratio of overseas returns to the volatility of overseas business operations	Financial data, overseas operation data	Positive

Since the information related to internationalization performance is distributed in different sources, and the same index often has problems such as caliber difference, time misalignment and repeated disclosure, this paper does not directly use the original value of a single database, but performs weighted fusion of similar variables based on source reliability, disclosure integrity and time consistency. Let the multi-source observations of index q in enterprise i and year t be xq,it(s), then the fusion result is as follows:

$$\tilde{x}_{q,it} = \sum_{s=1}^{S_q} \phi_s \, x_{q,it}^{(s)}, \quad \sum_{s=1}^{S_q} \phi_s = 1 \tag{24}$$

Here, ϕs denotes the weight of the s-th source. Audit annual reports, official divisional statements and official investment records are assigned high weights, while announcement summaries and secondary collation data are assigned relatively low weights. Through this processing, the comparability and stability of the firm-year indicators can be improved while preserving multi-source information.

In the process of index generation, financial indicators such as the proportion of overseas revenue, the return on overseas assets and the growth rate of overseas revenue are directly calculated based on the integrated enterprise financial and division information. The business coverage along the Belt and Road is identified according to the number of countries where the enterprise has exported, invested, acquired or set up overseas institutions along the Belt and Road, and uniformly converted into a comparable proportion index; The export duration is generated based on the continuous export records of the enterprise, reflecting its ability to maintain the international market. The export stability index is inversely constructed based on the export volatility under the rolling window. The smaller the volatility is, the higher the stability is. The priority of M&A integration performance is calculated based on the operational improvement and transaction completion status after the completion of M&A; The overseas business recovery index reflects the ability of enterprises to resume overseas operations after external shocks. After the above processing, data from different sources, different structures and different scales are uniformly transformed into internationalization performance indicators under the same evaluation framework.

Considering the significant differences in dimension, distribution and dispersion of each index, missing repair, outlier processing and standardization conversion should be completed before comprehensive evaluation. Missing values are first imputed according to the data of the same enterprise in adjacent years; If consecutive years were missing, the industry-year median was used to fill in. Outliers are processed by quantile tail reduction to weaken the interference of extreme samples on the results. After the direction is unified, all indicators enter the objective weighting stage. In order to avoid the deviation caused by subjective weight setting, this paper uses the entropy method to determine the weight of each index. Let the standardized index value be zq,it, then its weight is:

$$w_q = \frac{1 - e_q}{\sum_{q=1}^{Q}(1 - e_q)} \tag{25}$$

Here, e_q is the information entropy of the QTH index. This method can make the indexes with higher dispersion degree and stronger information distinguishing ability obtain higher weights, so as to enhance the ability of the evaluation results to identify enterprise differences.

After obtaining the index weights, this paper uses the TOPSIS method to calculate the comprehensive index of enterprise internationalization performance. By comparing the relative distance between the enterprise and the optimal state and the weakest state, the multi-dimensional index is ranked comprehensively. Let the comprehensive index of internationalization performance of enterprise i in year t be IP_{it}, then its expression is as follows:

$$IP_{it} = \frac{D_{it}^{-}}{D_{it}^{+} + D_{it}^{-}} \tag{26}$$

Here, D_{it}^{+} represents the distance from the firm to the positive ideal solution, and D_{it}^{-} represents the distance from the firm to the negative ideal solution. The larger the value of IP_{it}, the closer the enterprise is to the international operation state of high coverage, high continuity, high stability, high efficiency and high resilience.

The resulting comprehensive index of internationalization performance absorbs multi-dimensional information such as export, investment, merger and acquisition and overseas operation at the same time, which can reflect the real level of internationalization operation of Chinese enterprises along the "Belt and Road".

3.4 Design of influence Effect identification Model based on Machine Learning

After completing the construction of enterprise digital empowerment index DEit and internationalization performance composite index IPit, this paper further uses machine learning methods to identify the impact effect of digital empowerment on internationalization performance. Since the internationalization performance of enterprises is often affected by multiple factors such as scale, capital structure, R&D intensity, industry attributes, overseas layout foundation, institutional environment and digital infrastructure at the same time, and there may be nonlinear relationships, interactions and sample heterogeneity between digital empowerment and internationalization performance, this paper does not use a single linear model for direct estimation. Instead, a two-stage identification framework of "nonlinear fitting - effect purification - result interpretation" is constructed.

The model input consists of three parts: the first is the core explanatory variable, which is the enterprise digitization empowerment index DEit obtained in Section 3.2. The second is the explained variable, which is the internationalization performance index IPit constructed in Section 3.3. The third is the control variable matrix Xit, which includes firm size, asset-liability ratio, R&D investment intensity, capital intensity, firm age, industry category, regional digital infrastructure, host country institutional environment and market openness. In order to prevent information leakage, this paper takes enterprise-year as the basic sample unit, adopts the hierarchical division method of "training set-validation set-test set", and sets intra-group constraints on the samples of the same enterprise in consecutive years to avoid the same enterprise entering the training set and test set at the same time in adjacent years. Categorical variables were coded one-hot, continuous variables were standardized as described above, and variables with high missing rates were not directly entered into the model, but were represented by missing indicator variables and imputed values.

In the first stage, the gradient boosting tree model is used to learn the nonlinear mapping relationship between digitization empowerment and internationalization performance. Considering that there are both high-dimensional structured features and complex interaction terms in the samples, XGBoost is preferred as the main learner in this paper. Let the KTH regression tree be fk(·), then the predicted performance value of firm i in year t is as follows:

$$\widehat{\mathrm{IP}}_{\mathrm{it}} = \sum_{k=1}^{K} f_k\,(\mathrm{DE}_{\mathrm{it}}, X_{\mathrm{it}}) \tag{27}$$

where K is the number of trees. The optimization objective of XGBoost consists of two parts: empirical loss and model complexity, and its objective function is as follows:

$$\mathcal{L} = \sum_{i,t} l\left(\mathrm{IP}_{\mathrm{it}}, \widehat{\mathrm{IP}}_{\mathrm{it}}\right) + \sum_{k=1}^{K} \Omega\,(f_k) \tag{28}$$

Here, $l(\cdot)$ is the squared loss function and Ω(fk) is the tree complexity penalty term, which is used to constrain the tree depth, the number of leaf nodes and leaf weights to avoid the model overfitting in the panel samples. The core hyperparameters include max_depth, eta, min_child_weightmin, subsample, and colsample_bytree. In this paper, Bayesian optimization or grid search is used to optimize the hyperparameters on the training set, and the mean square error, mean absolute error and R2 are used as the evaluation criteria on the validation set. Finally, a set of parameters with the minimum generalization error is selected to enter the testing phase.

Since the gradient boosting tree can automatically capture the threshold effects and high-order interactions between variables, this stage is mainly used to identify the overall predictive contribution and nonlinear effect characteristics of digitization empowerment on internationalization performance. In order to improve the interpretability of the results, SHAP method is further introduced to decompose the variable contributions. For any enterprise-year sample, the predicted value of internationalization performance can be decomposed into the sum of the benchmark forecast and the marginal contribution of each variable, where the SHAP value of the digitally enabled variable reflects its local influence direction and influence strength on the current sample. By sorting and clustering the SHAP values of the full sample, it is possible to identify whether digital empowerment shows significant heterogeneity under conditions of different sizes of enterprises, different industries, and different market layouts along the routes.

The second stage employs a dual machine learning approach to estimate the net effect of digitization empowerment on internationalization performance. Under the premise of controlling the influence of high-dimensional covariates, the machine learning model is used to fit the predicted part of the explained variable and the core explanatory variable respectively, and then the residual term is used to perform orthogonal regression, so as to reduce the omitted variable bias and model setting bias. In this paper, the following partial linear structure is set:

$$\mathrm{IP}_{\mathrm{it}} = \theta \mathrm{DE}_{\mathrm{it}} + g(X_{\mathrm{it}}) + \varepsilon_{\mathrm{it}}, \quad \mathrm{DE}_{\mathrm{it}} = m(X_{\mathrm{it}}) + \nu_{\mathrm{it}} \tag{29}$$

Here, $g(X_{it})$ and $m(X_{it})$ are unknown nonlinear functions learned by machine learning methods, and θ is the net influence coefficient of digitization empowerment that we focus on in this paper. The residuals are obtained by cross-fitting:

$$\widetilde{\mathrm{IP}}_{\mathrm{it}} = \mathrm{IP}_{\mathrm{it}} - \hat{g}(X_{\mathrm{it}}), \quad \widetilde{\mathrm{DE}}_{\mathrm{it}} = \mathrm{DE}_{\mathrm{it}} - \hat{m}(X_{\mathrm{it}}) \tag{30}$$

Then the residual error is used for quadratic regression, and the following is obtained:

$$\hat{\theta} = \left(\sum_{i,t} \widetilde{DE}_{it}^{2}\right)^{-1} \sum_{i,t} \widetilde{DE}_{it} \widetilde{IP}_{it} \tag{31}$$

The advantage of this estimator is that even if $g(\cdot)$ and $m(\cdot)$ are fitted by complex nonparametric learners, $\hat{\theta}$ can still maintain good robustness as long as the prediction error satisfies certain convergence conditions. In this paper, random forest, LightGBM and XGBoost are used as auxiliary learners at this stage to estimate $\hat{g}(X_{it})$ and $\hat{m}(X_{it})$ respectively, and the sensitivity of sample division to the results is reduced by five-fold cross-fitting.

In order to identify the heterogeneity of the impact effect of digital empowerment, this paper further groups the samples by firm size, technology intensity, market coverage breadth along the routes and the institutional environment of the host country, repeats the above dual machine learning estimation process respectively, and compares the group differences of θ^. At the same time, the interaction feature set (DEit×Hit) is added to the main model, where Hit represents the heterogeneous grouping variable or environmental variable, to test the marginal effect changes of digital empowerment in different contexts. If some interaction features show high contributions in both SHAP ranking and DML residual regression, it can be determined as an important source of adjustment affecting the effect.

In order to ensure the stability of the model output, this paper implements three types of control synchronously in the process of model training and effect recognition. Firstly, a rolling validation based on year is used to test the generalization ability of the model to samples from different periods, so as to avoid the model only fitting the local year characteristics. Secondly, cluster robust standard error correction was performed on the enterprise dimension to reduce the influence of the internal correlation of the panel sample on the estimation results. Third, the core explanatory variable DEit is tested by quantile truncation and alternative caliber to investigate whether the net effect estimate is driven by extreme observations. After the above processing, the final model results can not only depict the complex nonlinear relationship between digital empowerment and internationalization performance, but also identify the net impact direction, impact strength and heterogeneity characteristics of digital empowerment on internationalization performance.

3.5 Description of variables and model setting

In this paper, the enterprise-year is taken as the basic observation unit, the variables generated by text mining and multi-source data fusion are uniformly coded, and the analysis matrix for machine learning identification and robustness comparison is constructed. The explained variable is the internationalization performance composite index (IPit), which is weighted by multi-dimensional indicators such as the proportion of overseas revenue, the coverage of operation along the routes, the duration of export, the stability of export, the intensity of OFDI, the performance of cross-border M&A, the efficiency of overseas operation, the growth rate of overseas revenue, the recovery index of overseas business and the risk ratio of overseas revenue. The core explanatory variable is the digitization empowerment index DEit, which is obtained from the texts of enterprise annual reports, ESG reports and announcements after dictionary matching, semantic recognition, chapter weighting and entropy aggregation. The control variables include firm size, asset-liability ratio, R&D investment intensity, capital intensity, firm age, ownership nature, industry competition degree, regional digital infrastructure and host country institutional environment. The definitions, symbols and measures of each variable are shown in Table 4.

Table 4: Description of variable definitions and measures

Variable Type	Variable Name	Symbol	Measurement Method
Dependent Variable	Composite Internationalization Performance Index	IP_{it}	Calculated by integrating multi-source performance indicators using the entropy method and TOPSIS
Core Independent Variable	Digital Empowerment Index	DE_{it}	Constructed through text mining, semantic recognition, and chapter-weighted aggregation
Control Variable	Firm Size	$Size_{it}$	Natural logarithm of total assets
Control Variable	Leverage Ratio	Lev_{it}	Total liabilities / total assets
Control Variable	R&D Intensity	RD_{it}	R&D expenditure / operating revenue
Control Variable	Capital Intensity	Cap_{it}	Fixed assets / number of employees
Control Variable	Firm Age	Age_{it}	Current year minus year of establishment
Control Variable	Digital Infrastructure	Dig_{rt}	Digital infrastructure index of the region where the firm is located
Control Variable	Institutional Environment	Ins_{ct}	Host-country institutional environment index
Moderating Variable	Interaction Term	$DE_{it} \times H_{it}$	Product term of digital empowerment and heterogeneity variables

The model was set in the way of "baseline regression + interaction expansion". The baseline model is used to examine the overall impact of digitization empowerment on internationalization performance:

$$IP_{it} = \alpha + \beta DE_{it} + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{32}$$

Here, x t is a vector of control variables, and μi and λt denote firm fixed effects and year fixed effects, respectively. In order to identify marginal changes under context differences, the interaction model is further set as follows:

$$IP_{it} = \alpha + \beta_1 DE_{it} + \beta_2 H_{it} + \beta_3 (DE_{it} \times H_{it}) + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{33}$$

Among them, Hit represents heterogeneous variables such as firm size, technology intensity, coverage breadth along the route or institutional environment. All variables were corrected for missing values, tailed, and standardized before entering the model, and were consistent with the machine learning input matrix in Section 3.4.

4 An empirical analysis of digitization empowerment to enhance internationalization performance

4.1 Descriptive statistics and correlation analysis

In order to investigate the distribution characteristics of the main variables of the sample enterprises and their interrelations, this paper firstly carries out descriptive statistics on the comprehensive index of internationalization performance, the digital empowerment index and the control variables, and the results are shown in Table 5. In general, the mean value of IPIPIP is 0.468, and the standard deviation is 0.176, indicating that there are obvious differences in the business performance of the sample enterprises in the markets along the "Belt and Road". The mean value of the digital empowerment index DE is 0.514, and the standard deviation is 0.193, indicating that the investment degree of enterprises in data governance, platform collaboration, intelligent analysis and cross-border business digitization is not consistent. The variables such as enterprise scale, R&D intensity, digital infrastructure and institutional environment also show strong dispersion, indicating that the sample has a good degree of discrimination in resource endowment and external conditions, which provides a basis for the subsequent identification of the influence effect of digital empowerment.

Table 5: Results of the descriptive statistics

Variable	Symbol	Sample Size	Mean	Standard Deviation	Minimum	Maximum
Composite Internationalization Performance Index	IP	1824	0.468	0.176	0.102	0.891
Digital Empowerment Index	DE	1824	0.514	0.193	0.087	0.936
Firm Size	Size	1824	22.417	1.284	19.306	26.108
Leverage Ratio	Lev	1824	0.462	0.187	0.071	0.882
R&D Intensity	RD	1824	0.051	0.038	0.000	0.214
Capital Intensity	Cap	1824	0.387	0.214	0.041	1.126
Firm Age	Age	1824	14.263	6.518	2.000	34.000
Digital Infrastructure	Dig	1824	0.536	0.158	0.183	0.901
Institutional Environment	Ins	1824	0.587	0.142	0.241	0.873

Based on the descriptive statistics, this paper further conducted Pearson correlation analysis on the main variables, and the results are shown in Table 6. The digitalization empowerment index DE is significantly positively correlated with the internationalization performance composite index IP at the 1% level, which preliminarily indicates that the higher the digitalization level of an enterprise, the better its internationalization operation performance. R&d intensity, firm size, digital infrastructure and institutional environment also show a significant positive correlation with IP, indicating that resource base, technology investment and external environment may all support the internationalization performance of enterprises. The correlation coefficient between asset-liability ratio and IP is negative, but the absolute value is small, indicating that the impact of financial leverage is relatively limited. On the whole, the correlation coefficient between each explanatory variable did not reach a high level, and no obvious high correlation appeared, indicating that there was no serious multicollinearity problem in the sample, and the variable setting had good identifability.

Table 6: Results of correlation analysis of main variables

variables	IP	DE	Size	Lev	RD	Cap	Age	Dig	Ins
IP	1.000								
DE	0.472***	1.000							
Size	0.318***	0.287***	1.000						
Lev	-0.086*	-0.041	0.214***	1.000					
RD	0.295***	0.366***	0.198***	-0.072*	1.000				
Cap	0.124***	0.109**	0.241***	0.157***	0.083*	1.000			
Age	0.091**	0.118***	0.267***	0.063	0.044	0.139***	1.000		
Dig	0.341***	0.394***	0.176***	-0.028	0.201***	0.072*	0.058	1.000	
Ins	0.286***	0.221***	0.094**	-0.035	0.117***	0.048	0.021	0.316***	1.000

Note: *,, * denote significant at the 1%, 5%, and 10% levels, respectively.

4.2 Examination of the overall impact of digital empowerment on internationalization performance

After completing the variable construction and correlation analysis, this paper further examines the overall impact of digital empowerment on internationalization performance. The baseline model uses a combination of firm fixed effects and year fixed effects to control for unobservable individual characteristics of firms and macro annual shocks. The regression results are shown in Table 7. Column (1) only includes the digitization empowerment index, and the results show that the regression coefficient of DE is 0.286, which is significant at the 1% level, indicating that digitization empowerment has a significant positive impact on internationalization performance. Column (2) On this basis, after adding control variables such as enterprise size, asset liability ratio, R&D intensity, capital intensity and enterprise age, the coefficient of DE decreases to 0.241, but it is still significant at the 1% level, indicating that the promotion effect of digital empowerment is not caused by the difference in the basic characteristics of enterprises. In column (3), after the regional digital infrastructure and the institutional environment of host country are further included, the coefficient of DE is 0.219, and the significance remains unchanged, indicating that digital empowerment is still an important factor affecting the internationalization performance of enterprises after considering the external environmental constraints.

From the results of control variables, both enterprise size and R&D intensity show significant positive effects, indicating that enterprises with stronger resource base and higher technology investment are more likely to form stable international business advantages in the B&R markets. The coefficients of digital infrastructure and institutional environment are also positive, indicating that external digital and institutional conditions can provide support for enterprises' cross-border operations. The coefficient of the asset-liability ratio is negative but weakly significant, indicating that financial leverage does not have a stable effect on internationalization performance. In general, the benchmark regression results support the core hypothesis proposed in this paper, that is, digital empowerment can significantly improve the internationalization performance of Chinese enterprises along the Belt and Road.

In order to further test whether the estimated results still hold in the nonlinear case, this paper uses a dual machine learning method to supplement the identification of the net effect of digital empowerment. The results show that the average net effect estimate of digitization empowerment is 0.207 and significant at the 1% level, consistent with the fixed effects regression results in both direction and order of magnitude. This indicates that even when controlling for the complex nonlinear effects of high-dimensional covariates, the promoting effect of digitization empowerment on internationalization performance is still robust. Combined with the benchmark regression and machine learning identification results, it can be concluded that digital empowerment is not only statistically related to internationalization performance, but also has a substantial role in promoting the quality of operation, expansion ability and stability of enterprises in the Belt and Road markets. It indicates that enterprises through data governance, platform collaboration, intelligent analysis and digital construction of cross-border business, Can effectively improve international business performance.

Table 7: Test results of the overall impact of digitization empowerment on internationalization performance

Variables	(1) Baseline Model	(2) With Firm Characteristics	(3) With External Environment	(4) DML Net Effect
Digital empowerment index (DE)	0.286***	0.241***	0.219***	0.207***
	(6.84)	(5.97)	(5.42)	(4.96)
Firm size (Size)		0.053***	0.047***	
		(3.88)	(3.41)	
Leverage (Lev)		-0.029*	-0.021	
		(-1.78)	(-1.31)	
R&D intensity (RD)		0.118***	0.102***	
		(3.96)	(3.42)	
Capital intensity (Cap)		0.026*	0.019	
		(1.73)	(1.26)	
Firm age (Age)		0.008	0.006	
		(1.12)	(0.86)	
Digital infrastructure (Dig)			0.071**	
			(2.31)	
Institutional environment (Ins)			0.064**	
			(2.08)	
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	1824	1824	1824	1824
(R^2)	0.231	0.318	0.347	—

Note: t value in parentheses; *,, * denote significant at the 1%, 5%, and 10% levels, respectively.

4.3 Identifying Mechanisms of Action based on Computational models

Based on the overall effect test, this paper further uses the dual machine learning and mediation effect decomposition method to identify the internal transmission path of digital empowerment affecting internationalization performance. The results of the model show that digital empowerment has a significant role in promoting the efficiency of resource allocation, organizational collaboration ability and the level of responsibility governance, and after the above variables are included in the regression, the direct effect coefficient of digital empowerment decreases, indicating that its influence is not completely realized through direct effects, but has obvious mechanism transmission characteristics. Specifically, digitalization enables enterprises to improve the identification accuracy of overseas market demand, customer structure and supply chain status, and enhance the allocation efficiency of investment, export and M&A resources in cross-border scenarios. At the same time, the embedding of platform system, data interface and intelligent analysis tools improves the information sharing and process connection between headquarters and overseas branches, suppliers and customers, compresses the decision-making delay, and enhances the level of international business collaboration. The results based on SHAP value ranking and marginal contribution decomposition show that resource allocation efficiency and organizational collaboration ability are the two main channels of digital empowerment affecting internationalization performance, and the level of responsibility governance is reflected in the auxiliary intermediary mechanism. It shows that digital empowerment does not directly improve business results by relying solely on technological input, but continuously promotes the improvement of international performance by reshaping the information processing, collaborative governance and risk response capabilities in the process of cross-border business.

4.4 Heterogeneity, robustness and comparative analysis

In order to further test the difference and robustness of the results of the impact of digital empowerment on internationalization performance, this paper carries out an extended analysis from three levels: enterprise characteristics, business layout along the line and estimation methods, and the results are shown in Figure 4 and Figure 5.

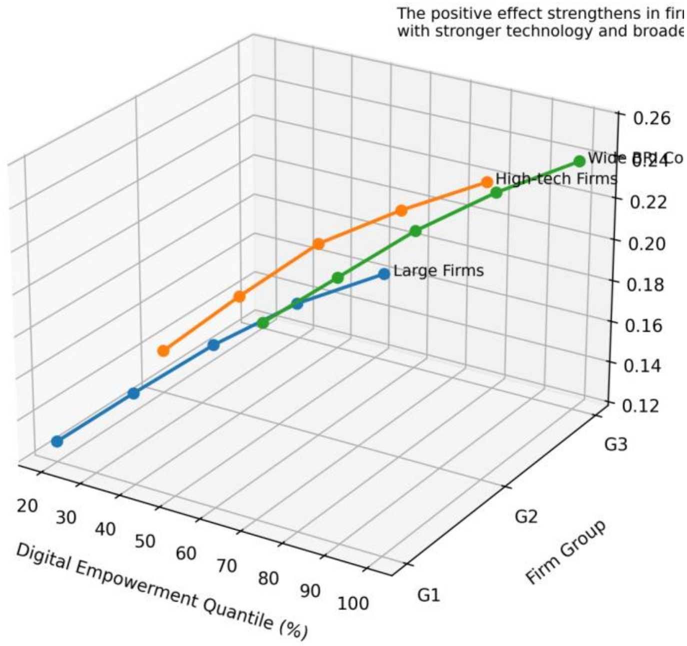

Figure 4: Heterogeneous Effects of Digital Empowerment

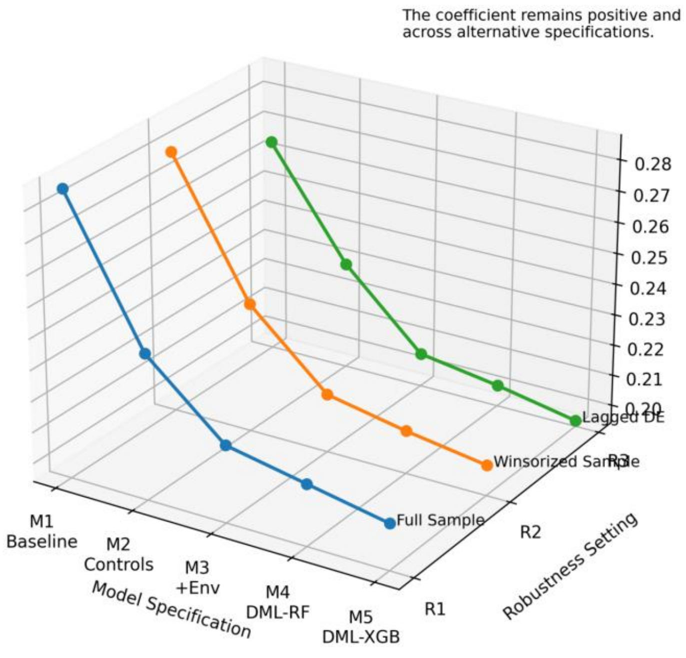

Figure 5: Robustness and Model Comparison Results

The heterogeneity analysis shows that there are obvious differences in the strength of the role of digital empowerment in different types of enterprises. Taking the quantile of digitalization empowerment as an example, the estimated coefficient of large-scale enterprise group increased from 0.128 to 0.247, with an increase of 0.119. The group of high-tech enterprises increased from 0.139 to 0.258, with an increase of 0.119; The group of enterprises with wider coverage of B&R markets increased from 0.121 to 0.239, an increase of 0.118. The three groups of curves show a continuous upward trend, but the coefficients of high-tech enterprises in each quantile are higher than the other two groups as a whole. For example, at the 80% quantile, the coefficient of high-tech enterprises group is 0.236, the coefficient of large-scale enterprises group is 0.224, and the coefficient of enterprises with wide coverage along the line is 0.215. It shows that digital empowerment is easier to translate into performance advantages in enterprises with strong technical foundation and perfect international business network. The three-dimensional curve in Figure 4 also shows that although different types of enterprises all benefit from the improvement of digital empowerment, the marginal effects are not the same, and there is a relatively obvious differentiation between groups.

Robustness and method comparison analysis further verify the reliability of the above results. Figure 5 shows that under different model Settings and sample processing conditions, the estimated coefficient of digitization empowerment always remains positive, and the fluctuation range is relatively limited. Taking the results of the full sample as an example, the coefficient decreased from 0.286 of the benchmark model to 0.241 after adding the control variables, and then to 0.219 after adding the external environment variables, with a change of 0.067, indicating that the core conclusion was still stable after the inclusion of the control variables. After tailing the samples, the coefficients were 0.279, 0.236 and 0.213, respectively, and the differences were all controlled at about 0.01 compared with the benchmark results. After lagging the digitization empowerment index by one period, the coefficients are 0.264, 0.229 and 0.205, which decrease slightly but still maintain a significant positive relationship. After further adopting the dual machine learning method, the estimated coefficients of DML-RF, DML-LightGBM and DML-XGBoost are 0.201, 0.205 and 0.209, respectively, and the maximum difference is only 0.008, indicating that different algorithms have strong consistency in judging the performance of digitization empowerment promoting internationalization. On the whole, the heterogeneity, robustness and method comparison results jointly show that digital empowerment can not only significantly improve the internationalization performance of Chinese enterprises along the "Belt and Road", but also this role is more prominent in high-tech enterprises, large-scale enterprises and enterprises with wider layout along the route.

5 Conclusion and Prospect

The research shows that digital empowerment has a significant role in promoting the internationalization performance of Chinese enterprises along the "Belt and Road". The benchmark regression results show that the estimated coefficient of the digital empowerment index is 0.286, and after the enterprise characteristics and external environment variables are included in turn, the coefficient still remains at 0.219, and the net effect of the dual machine learning estimation is 0.207, indicating that the conclusion has strong robustness. The heterogeneity analysis further shows that the promotion effect of digital empowerment is more obvious in high-tech enterprises, large-scale enterprises and enterprises with wider coverage along the routes. For example, the estimated coefficient of high-tech enterprises group increases from 0.139 to 0.258, with an increase of 0.119, indicating that there is a significant synergistic effect between digital technology, organizational foundation and international business network. The theoretical contribution of this paper is that the digital empowerment, internationalization performance and computer technology methods are integrated into a unified analysis framework, and the identification path based on text mining, multi-source data fusion and machine learning is constructed, which expands the analysis methods and measurement methods of enterprise internationalization research. In practice, we should pay attention to data governance, platform collaboration and intelligent analysis capacity building, improve information acquisition efficiency, resource allocation ability and risk response ability in cross-border operations, and promote internationalization from scale expansion to quality improvement. There are still some limitations in the research. The samples mainly focus on Chinese listed enterprises and have not fully covered small and medium-sized enterprises and non-listed enterprises. Some internationalization business process variables and high-frequency dynamic data are also difficult to obtain completely. In the future, the sample scope can be further expanded, and more fine-grained cross-border transaction data, overseas operation data and time series tracking data can be introduced to enhance the model's ability to identify the dynamic mechanism and context differences of digital empowerment.

References

[1] Dong Y, He X, Blut M. How and when does digitalization influence export performance? A meta-analysis of its consequences and contingencies[J]. International Marketing Review, 2024, 41(6): 1388-1413.

[2] Fan L, Ou J, Yang G, et al. Digitalization and outward foreign direct investment of Chinese listed firms[J]. Review of International Economics, 2024, 32(2): 604-634.

[3] Zhang Y, Chen Y, Zhang R, et al. Outward foreign direct investment productivity along the belt and road[J]. Economics & Politics, 2024, 36(2): 959-988.

[4] Hongliang L, Zeren Z, Haoran M. The impact of digital transformation on enterprise export resilience: Evidence from China[J]. International Review of Economics & Finance, 2024, 95: 103500.

[5] Zhang C, Qiu P, Zhang L, et al. The impact of digital transformation on enterprises' export stability: Evidence from listed companies in China[J]. International Review of Financial Analysis, 2024, 96: 103582.

[6] Khurram M U, Abbassi W, Chen Y, et al. Outward foreign investment performance, digital transformation, and ESG performance: Evidence from China[J]. Global Finance Journal, 2024, 60: 100963.

[7] Guo L, Zhong Q, Wang H. Digital transformation, ESG responsibility and corporate's export performance[J]. Finance Research Letters, 2024, 69: 106106.

[8] Wu H, Qiao Y, Luo C. Cross-border e-commerce, trade digitisation and enterprise export resilience[J]. Finance Research Letters, 2024, 65: 105513.

[9] Zhu L. The influence of digital transformation on Chinese firms' outward foreign direct investment[J]. Finance Research Letters, 2025, 72: 106567.

[10] Xuan Z, Yan Q. Digital transformation of enterprises, degree of internationalization, and performance in cross-border mergers and acquisitions[J]. Finance Research Letters, 2025: 108445.

[11] Xu X, Chen X, Yang J, et al. Digital transformation and export duration: Implications for firm financial performance[J]. Pacific-Basin Finance Journal, 2025, 92: 102780.

[12] Wu Y, Xu W. Digital innovation and firm internationalization: Insights from publicly listed firms in China[J]. Economic Modelling, 2025, 151: 107212.

[13] Appiah E K, Gabrielsson P, Criado A R. The role of digitalization in early internationalization Research: A systematic literature review and future research directions[J]. Journal of Business Research, 2025, 186: 115043.

[14] Yang B, Bai W, Chen Y, et al. Internationalization of digital firms: A systematic review and research agenda[J]. Journal of Business Research, 2025, 189: 115124.

[15] Li J, Surana A, Chavan M, et al. The internationalization of digital platform-based firms: A systematic literature review and directions for future research[J]. Journal of World Business, 2025, 60(4): 101629.

[16] Cahen F, Borini F, Dhanaraj C, et al. Unpacking global digital competence in the contemporary international venture[J]. International Business Review, 2025, 34(4): 102414.

[17] Zhang C, Bai T, Zhou A J, et al. Digital platforms, internal digitalization, and internationalization of SMES[J]. Long Range Planning, 2025: 102588.

[18] Capelleras J L, Martin-Sanchez V, Zhang C. The curvilinear relationship between digitalization and export propensity: The role of home country corruption in emerging economies[J]. Technological Forecasting and Social Change, 2025, 214: 124043.

[19] Wen Z, Zhang L, Fang X. Belt and road initiative and sustainable development performance of Chinese enterprises[J]. International Review of Economics & Finance, 2025: 104598.

[20] Wu J, Yin W. Initiative as Institution: How the Belt and Road Initiative Influences Chinese Firms' Participation in Global Value Chains[J]. The Chinese Journal of International Politics, 2025, 18(1): 63-84.