Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The role of institutional elements in enterprise innovation resilience: Explainable machine learning evidence from China

Nan Li
Central University of Finance and Economics
Jianguo Zhang
yjs2025278@ahstu.edu.cn
Anhui Science and Technology University
Jianhui Yin
Northeast Forestry University
Wei Chen
Harbin Engineering University

Article

Keywords: Institutional elements, agriculture- and forestry-related enterprises, enterprise innovation resilience, machine learning, SHAP

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11129700/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

The role of institutional elements in enterprise innovation resilience: Explainable machine learning evidence from China

Abstract: The institutional environment is an important external condition shaping enterprise innovation resilience (*EIR*), particularly enterprises' capacity to sustain innovation and adapt to external shocks. Yet existing studies have largely focused on single institutional factors or linear average effects, leaving the relative importance, nonlinear patterns, and interactions of multidimensional institutional elements (*IE*) underexplored. Drawing on institutional theory and a sample of Chinese listed agriculture- and forestry-related enterprises, this study combines multiple machine-learning algorithms, including CatBoost, with SHAP to examine how regulative, normative, and cognitive *IE* predict one-period-ahead enterprise innovation resilience (EIR_{t+1}). The results show that current-period enterprise innovation resilience (EIR_t) is the strongest predictor of EIR_{t+1}, indicating pronounced dynamic persistence. Among the three institutional dimensions, environmental governance (*RG*), institutional investor shareholding (*INST*), and innovation culture (*EIC*) are the most important predictive features, while the Cognitive Institutional Index (*CII*) has the highest predictive importance among the three composite institutional indices. *IE* also exhibit nonlinear, stage-dependent, and cross-dimensional interaction patterns, with substantial heterogeneity across regions, ownership types, enterprise sizes, and agriculture- versus forestry-related enterprises. These findings extend institutional theory to *EIR* research and highlight the predictive relevance of multidimensional, nonlinear, and context-dependent institutional configurations.

Keywords: Institutional elements; agriculture- and forestry-related enterprises; enterprise innovation resilience; machine learning; SHAP

1. Introduction

Innovation-driven development is an important foundation for enterprises' sustained growth and their ability to respond to external uncertainty. Yet industries differ in their knowledge bases, technological opportunities, input structures, and institutional environments, giving rise to industry-specific innovation patterns and resource-allocation logics (Malerba, 2002). As uncertainty rises, enterprises' investment decisions tend to become more cautious, and resource allocation may become less responsive to external shocks (Bloom et al., 2007). In 2024, China's research and experimental development (R&D) expenditure reached 3.63 trillion yuan, of which enterprises' R&D expenditure amounted to 2.82 trillion yuan, accounting for 77.7% of the national total. This indicates that enterprises have become the primary contributors to innovation investment in China. The Global Innovation Index 2025, released by the World Intellectual Property Organization (WIPO), shows that China's ranking in the Global Innovation Index rose further to 10th in 2025. However, the R&D intensities of the agricultural and sideline food processing industry and the food manufacturing industry, which are closely related to agriculture-related enterprises, were only 0.66% and 0.94%, respectively, in 2024, well below the manufacturing average of 1.82%. Meanwhile, data released by the National Forestry and Grassland Administration show that the total output value of China's forestry and grassland industries reached 10.17 trillion yuan in 2024, representing a year-on-year increase of 9.6%. These facts indicate a noteworthy structural tension in agriculture- and forestry-related industries among scale expansion, innovation investment, and the enhancement of innovation capabilities. For agriculture- and forestry-related enterprises that are highly dependent on resources, have relatively long operating cycles, and are strongly influenced by the external institutional environment, sustained innovation depends not only on the scale of investment but also on their dynamic capabilities to perceive environmental changes, seize opportunities, and reallocate resources (Teece, 2007). More broadly, an enterprise's ability to maintain stability and recover in the face of external shocks is an important manifestation of organizational resilience (DesJardine et al., 2019). Enterprises therefore also need to continuously adjust and reconfigure their innovation activities under institutional constraints, resource changes, and external disturbances, thereby developing innovation resilience characterized by continuity and adaptability. The long-term accumulation of sustainable business practices and organizational capabilities helps enhance enterprises' stability and capacity for adjustment in the face of external shocks (Ortiz-de-Mandojana and Bansal, 2016). Organizational resilience

does not merely refer to recovery after a shock; it also reflects a dynamic process through which enterprises continuously identify changes, adjust resources, and develop adaptive capabilities in uncertain environments (Linnenluecke, 2017).

Existing research shows that the institutional environment can influence enterprises' innovation decisions and resource allocation through policy regulation and formal governance arrangements (Chen, 2024). Capital market governance can likewise affect enterprises' innovation investment through institutional investor monitoring, corporate governance, and managerial incentives (H. Zhang et al., 2024). For agricultural enterprises, digital transformation is reshaping knowledge and advisory networks and changing how enterprises connect informationally with knowledge providers, policy actors, and other innovation participants (Fielke et al., 2020). For forestry enterprises, innovation activities are also significantly constrained by characteristics such as resource dependence, production cycles, and industrial distribution (Tu et al., 2025). Thus, the innovation resilience of agriculture- and forestry-related enterprises depends not only on investment in innovation resources but also on formal institutional constraints, market governance, and organizational cognition and capability foundations. Although existing studies have examined the relationship between the institutional environment and enterprise innovation from multiple perspectives, including policy regulation, capital market governance, stakeholder pressure, and organizational cognition, several areas warrant further investigation. First, existing studies have largely focused on a single institutional factor and have rarely systematically compared the relative effects of the regulative, normative, and cognitive dimensions of *IE* within a unified framework. Second, existing research has focused more on average effects and paid insufficient attention to nonlinear changes, cross-dimensional interactions, and configurations among different *IE* variables. When multiple institutional requirements coexist or even conflict, organizations may develop differentiated response patterns, while internal actors and power relations may also shape the outcomes of specific responses (Pache and Santos, 2010). Moreover, institutional complexity may be shaped by the continually changing configurations of multiple institutional logics (Bohn and Gümüsay, 2024). Third, systematic evidence remains limited for agriculture- and forestry-related enterprises, which are highly resource-dependent, operate over relatively long cycles, and are simultaneously subject to constraints arising from industrial policies, natural resources, and local governance. Fourth, although traditional regression and structural models offer advantages for parameter estimation and theoretical interpretation, data-driven methods such as machine learning can provide complementary analysis in the presence of high-dimensional features, complex nonlinearities, and interactions.

Building on these research gaps, this study examines listed enterprises engaged in agriculture and forestry in China. It develops a multidimensional indicator system of *IE* across three dimensions–regulative, normative, and cognitive–and combines multiple machine-learning models with SHAP-based explainability techniques to identify the relative predictive contributions of *IE* to *EIR*, as well as their nonlinear patterns and interaction characteristics. It then constructs three composite indices of *IE* and examines their group heterogeneity across regions, ownership types, enterprise sizes, and types of agriculture- and forestry-related enterprises. The study's marginal contributions are threefold. First, it extends research on enterprises' innovation resilience beyond internal resources, organizational capabilities, and individual external institutional factor to the multidimensional institutional environment in which enterprises are embedded. By examining the regulative, normative, and cognitive dimensions of *IE* within a unified framework, it helps explain the formation of sustained innovation capabilities in agriculture- and forestry-related enterprises from the perspective of institutional configurations. Second, it compares the relative predictive importance of different institutional dimensions and specific institutional characteristics, while also identifying their nonlinear and stage-dependent patterns, interactions, and group differences. In doing so, it addresses the limitations of analyses based on single institutional variables and average relationships in characterizing complex institutional environments. Third, it introduces a framework combining multiple machine-learning models with SHAP-based explainability into research on *IE* and *EIR*, thereby providing, alongside traditional statistical analysis, a complementary approach to identifying high-dimensional features, predicting nonlinear relationships, and analyzing feature interactions.

2. Literature review

2.1 *IE* and *EIR*

Keywords provide a relatively concentrated reflection of the knowledge base, focal concerns, and thematic evolution of a research field. In terms of theoretical development, resilience research has gradually shifted from its early emphasis on organizations' capacity to recover after shocks to the dynamic process through which organizations anticipate risks, respond to disruptions, adapt to change, and reconstruct their capabilities in uncertain environments. Williams et al. (2017) noted that organizational resilience is not merely a matter of post-crisis recovery, but a continuous process encompassing preparation before a shock, response during the shock, and adjustment after the shock. Duchek (2020) further conceptualized organizational resilience as comprising three interconnected capability stages—anticipation, coping, and adaptation-and emphasized that organizations develop future-oriented resilience capabilities through continuous learning and resource reconfiguration. Hillmann and Guenther (2021) likewise argued that the core of organizational resilience lies in enterprises' ability to survive and continue developing in adverse and turbulent environments. From a dynamic perspective on enterprise resilience, Conz and Magnani (2020) further noted that enterprises may follow different resilience pathways, including absorption, adaptation, and adjustment, when confronted with shocks, and that these pathways evolve over time. Building on this line of research, this study conceptualizes enterprise innovation resilience as the dynamic capacity of enterprises to maintain, adjust, and reconstruct their innovation activities in response to external disruptions and changes in the institutional environment, thereby sustaining innovation continuity and adaptive capacity. Against this background, CiteSpace, proposed by Chen (2006), identifies knowledge structures and research frontiers through network-visualization methods such as keyword co-occurrence, clustering, and temporal evolution, and has therefore been widely used in scientometric and interdisciplinary literature analysis. In a study of the intersection between artificial intelligence and the low-altitude economy, Yin et al. (2026) likewise used CiteSpace to map the research structure through keyword co-occurrence, keyword clustering, and timeline analysis, thereby identifying research hotspots and frontiers. Drawing on this approach, this study uses CiteSpace to conduct keyword co-occurrence, clustering, and timeline analyses of the literature related to institutional elements and innovation resilience. Because "enterprise" is used primarily to define the micro-level sample in the subsequent empirical analysis, making enterprise or firm mandatory search terms could unduly narrow the coverage of representative literature on institutional theory, organizational studies, and innovation resilience. Accordingly, at the knowledge-mapping retrieval stage, "institutional elements" and "innovation resilience" are used as the core thematic concepts. This preserves thematic relevance while expanding coverage of the representative knowledge base; the subjects of the subsequent empirical analysis remain limited to agriculture- and forestry-related enterprises.

The keyword co-occurrence network in Figure 1 comprises 519 nodes and 1363 links, with a network density of 0.0101. The keywords systems, organizations, knowledge, strategy, performance, management, governance, policy, institutional theory, and innovation are relatively prominent and form comparatively dense associations. This indicates that relevant research is no longer confined to a single institutional rule or innovation outcome; instead, it has increasingly linked the institutional environment with organizational management, the knowledge base, strategic behavior, governance structures, and innovation performance. Meanwhile, the emergence of the keywords dynamics and climate change suggests that dynamic environmental change and how organizations respond to external disturbances are becoming new research foci. Innovation resilience is thus gradually taking shape as an important cross-cutting topic within this knowledge network.

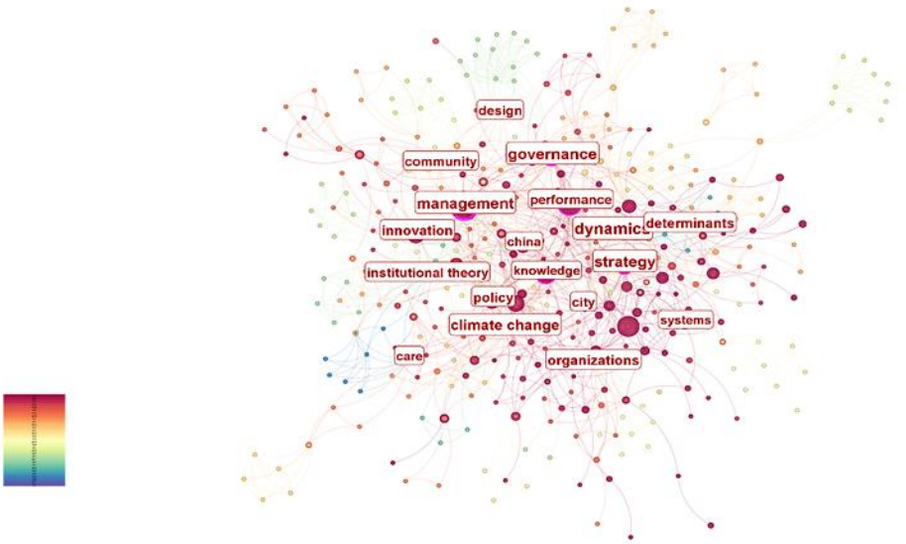


Fig. 1. Co-occurrence map of keywords in research on *IE* and innovation resilience, 1995-2026.

Keyword clustering was further conducted using CiteSpace, as shown in Figure 2. The results yielded a modularity of Q=0.7439 and an average silhouette value of S=0.8945, indicating a relatively clear modular structure and high clustering consistency. Ten major clusters were identified: #0 knowledge, #1 adaptive capacity, #2 innovation resilience, #3 evolution, #4 artificial intelligence, #5 digital divide, #6 institutional theory, #7 institutional pressures, #8 green innovation resilience, and #9 policy. Thematically, these clusters can be broadly organized into three interrelated research directions. First, research on organizational capabilities and resilience, represented by knowledge, adaptive capacity, and innovation resilience, focuses on knowledge accumulation, adaptive capacity, and the persistence of innovation activities under shocks. Second, research on the institutional environment, represented by institutional theory, institutional pressures, and policy, suggests that formal rules, policy instruments, and institutional constraints have become important external conditions for explaining organizational innovation and resilience. Third, emerging research directions represented by artificial intelligence, digital divide, and green innovation resilience indicate that digital technologies, artificial intelligence, and green transformation are increasingly entering the innovation resilience research agenda. It should be noted that "#7 institutional pressures" in the figure is merely an original clustering label automatically generated by CiteSpace based on the sample literature. It represents a specific theme in existing research and does not mean that this study defines the core explanatory variable as institutional pressures. This study consistently examines multiple institutional elements within a three-dimensional institutional framework comprising regulative, normative, and cognitive dimensions.

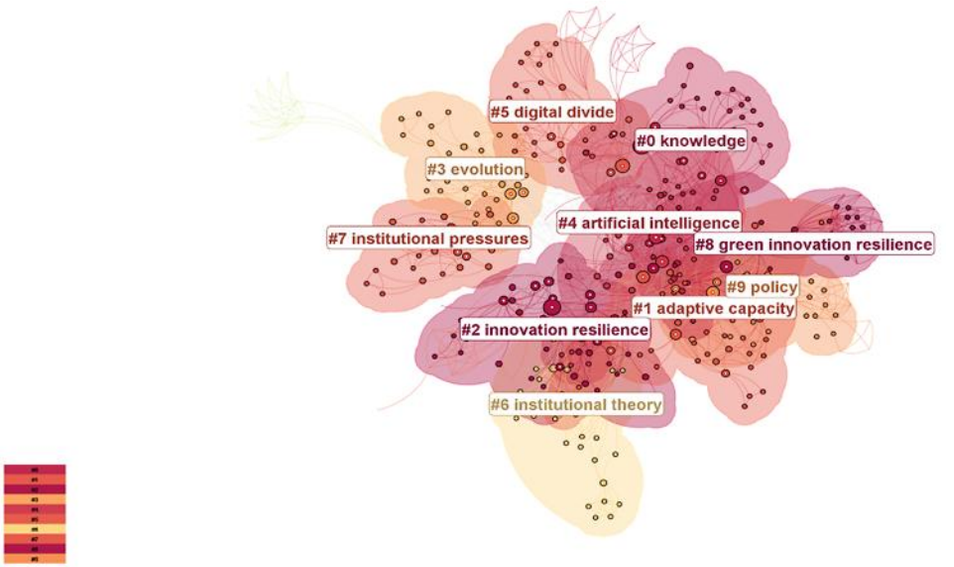


Fig. 2. Keyword clustering map of research on *IE* and innovation resilience, 1995-2026

Fig. 3 further reveals the temporal evolution of the related research themes. Early studies focused primarily on broad issues of technological innovation and organizational evolution, with keywords such as technology, innovation, and evolution featuring prominently. As institutional theory became increasingly integrated into enterprise and organizational research, themes such as governance, policy, and institutional theory gradually gained prominence. In recent years, keywords such as knowledge, adaptive capacity, innovation resilience, artificial intelligence, and green innovation resilience have become increasingly prominent, indicating a gradual shift in research focus from innovation performance in general toward adaptive capacity, sustained innovation, digital technologies, and institutional governance under conditions of complex disruption. Enterprise innovation resilience is therefore emerging as an important interdisciplinary research avenue linking institutional environment, organizational adaptation, and sustained innovation.

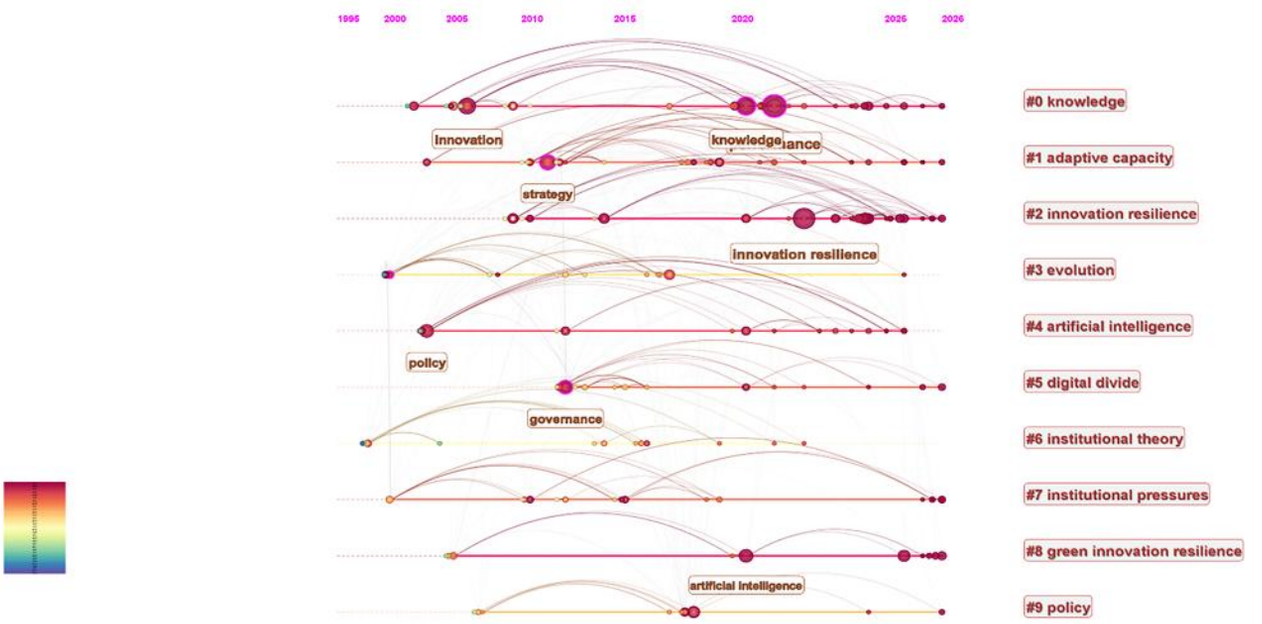


Fig. 3. Timeline knowledge graph of keywords in research on *IE* and innovation resilience, 1995-2026

The knowledge graph above indicates close research linkages among the institutional environment, knowledge base, adaptive capacity, and innovation resilience. However, few existing studies have compared the relative effects of multidimensional *IE* on the *EIR* of agriculture- and forestry-related enterprises within a unified framework. From a strategic research perspective, the institution-based view emphasizes that enterprises' strategic choices are shaped not only

by industry competition and internal resource constraints but also by the institutional environment in which they are embedded (Peng et al., 2009). Institutional research typically characterizes the environment in which organizations operate along three dimensions—regulative, normative, and cognitive—and shows that these dimensions can influence organizational behavior through rule constraints, social norms, and cognitive frameworks (Busenitz et al., 2000). Building on this framework, this study further examines the relative predictive effects of the regulative, normative, and cognitive dimensions of *IE* on the *EIR* of agriculture- and forestry-related enterprises.

2.2 Application of machine learning to EIR_{t+1}

Machine learning provides a complementary approach to conventional statistical modeling for identifying nonlinear relationships, complex interactions, and out-of-sample predictive patterns. Given the prediction-oriented objective of this study, multiple machine-learning models and SHAP are employed to examine the predictive contributions of multidimensional *IE* in predicting EIR_{t+1}. A detailed methodological comparison and review of related applications are provided in the Supplementary Information (Supplementary Note and Supplementary Tables S1 and S2).

3. Data and algorithms

Figure 4 summarizes the research workflow. A multidimensional *IE* indicator system is developed from the regulative, normative, and cognitive institutional framework and CiteSpace-based literature analysis. Data from CNRDS, corporate annual reports, and other public sources are screened, matched, treated for missing values, and standardized. The sample is then split into training and test sets to compare machine-learning models using fit and prediction-error metrics. Finally, SHAP is applied to the optimal model to assess the relative predictive importance, nonlinear patterns, and interactions of *IE* in predicting EIR_{t+1}.

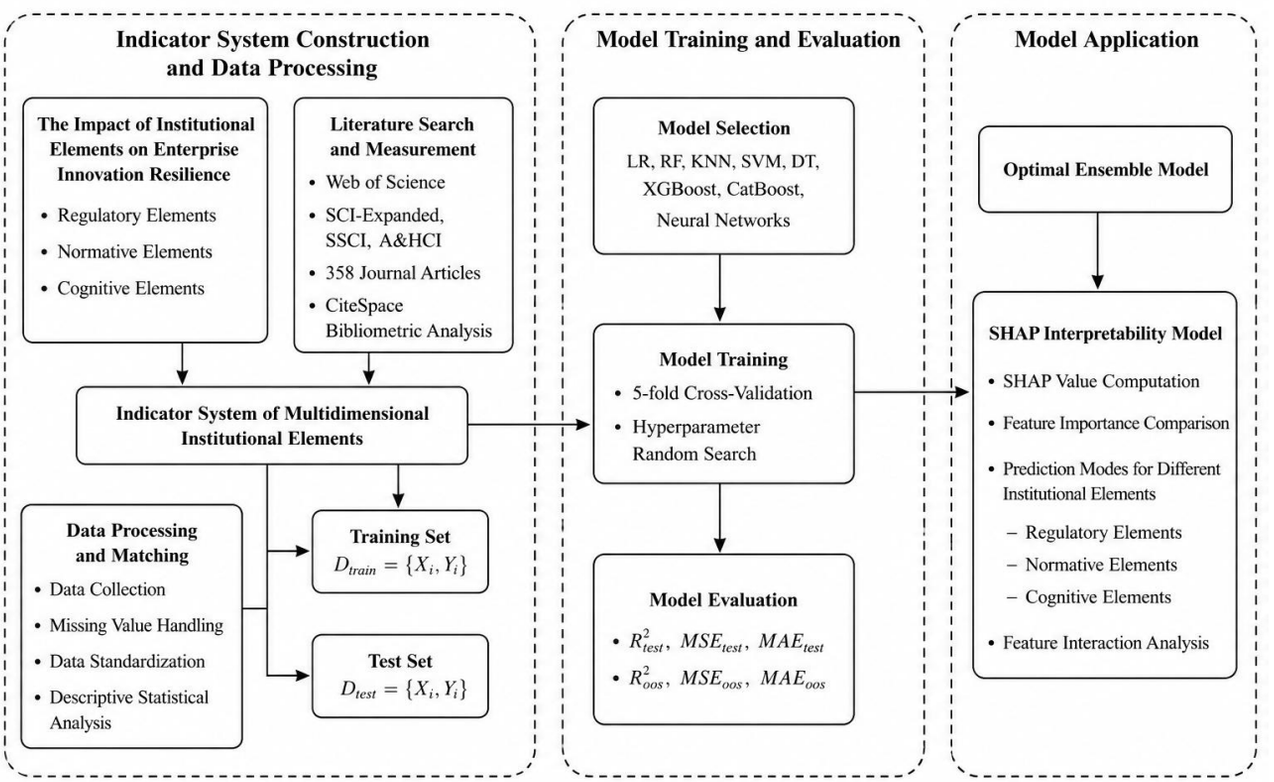


Fig. 4. Research methods and data workflow

3.1 Construction of the indicator system and data processing

3.1.1 Development of the indicator system

Drawing on the three-pillar institutional framework, this study develops a multidimensional indicator system for *IE* across the regulative, normative, and cognitive dimensions and incorporates the target variable and control variables (Kostova et al., 2020). Based on the preceding bibliometric and theoretical analyses, and considering the indicators'

theoretical relevance, operability, and data availability, we develop the indicator system presented in Supplementary Table S3.

First, *EIR* captures an enterprise's ability to maintain and adjust its innovation output in response to changes in the external environment. Compared with R&D investment and revenue from new product sales, the number of patent applications provides better intertemporal comparability and more effectively reflects innovation output. Following the relative innovation resilience measurement approach proposed by Wang et al. (2026), this study measures EIR_t as the deviation between the actual change in an enterprise's patent applications and the expected change implied by the overall trend in patent applications in the city where the enterprise is located. The specific calculations are presented in Eqs. (1) (2) (3).

$$EIR_t = \frac{\Delta P - \Delta E}{|\Delta E|} \tag{1}$$

$$\Delta P = P_t - P_{t-1} \tag{2}$$

$$\Delta \mathrm{E} = \left(\frac{P_{c,t} - P_{c,t-1}}{P_{c,t-1}}\right) P_{t-1} \tag{3}$$

where EIR_t denotes the relative innovation resilience of an enterprise in year t; ΔP denotes the actual change in the enterprise's number of patent applications; ΔE denotes the expected change estimated from the overall trend in patent applications in the city where the enterprise is located; P_t and P_{t-1} denote the numbers of patent applications filed by the enterprise in years t and $t-1$, respectively; and $P_{c,t}$ and $P_{c,t-1}$ denote the corresponding numbers of patent applications in the city where the enterprise is located. A larger EIR_t indicates a more favorable deviation of the enterprise's actual change in patent applications from the expected change implied by the citywide trend, reflecting stronger innovation resilience. Given that the effects of the institutional environment on enterprise innovation behavior may operate with a time lag, this study further uses EIR_{t+1} as the target variable in the machine-learning model.

Second, the regulative dimension of *IE*. Regulative institutions shape enterprise behavior through government regulation, administrative rules, policy incentives, and reward-and-punishment mechanisms, combining constraining and incentivizing functions. Regulatory requirements may increase short-term compliance and resource-adjustment costs, whereas stable policy expectations, technical standards, and innovation incentives can redirect technology choices and innovation resources. Relevant studies have shown that policy instruments can influence the direction of enterprises' technological innovation (Acemoglu et al., 2012) and that regulatory pressure can promote enterprises' environmental innovation, although this effect is conditioned by factors such as enterprises' resource endowments (Berrone et al., 2013). Research based on Chinese enterprises has likewise found that environmental regulation promotes green innovation, but that local political incentives and growth pressures significantly alter the strength of this effect (Xu and Sheng, 2023). For resource-dependent agriculture- and forestry-related enterprises, regulative institutions may therefore operate through a context-dependent "constraint–incentive" mechanism. Accordingly, this study uses central inspections (*CI*), environmental governance (*RG*), ownership type (*SOE*), and green subsidies (*GS*) to measure regulative institutional elements.

Third, the normative dimension of *IE*. Normative institutions influence enterprise behavior through social expectations, stakeholder monitoring, professional norms, industry standards, and judgments of organizational legitimacy. Their constraints mainly arise from reputational evaluations and governance pressures rather than formal legal requirements. Capital-market and stakeholder monitoring can affect managers' innovation incentives, risk-taking, and resource allocation. Institutional investors, in particular, may curb managerial short-termism through governance oversight and influence enterprises' innovation decisions (Aghion et al., 2013). Further research shows that institutional investor shareholding is significantly but nonlinearly related to enterprises' R&D investment, with effects varying across types of institutional investors (Mishra, 2022). For agriculture- and forestry-related enterprises, social expectations regarding green production, product quality, ecological responsibility, and food safety may translate into reputational and market pressures, thereby affecting innovation-related resource allocation and risk-taking. However, mismatches between external normative

requirements and enterprise capabilities may increase short-term adaptation costs. Therefore, this study uses media attention (*MA*), public scrutiny (*PS*), institutional investor shareholding (*INST*), and internal control quality (*IC*) to measure normative institutional elements.

Fourth, the cognitive dimension of *IE*. Cognitive institutions are reflected primarily in shared beliefs, value judgments, knowledge frameworks, and widely accepted behavioral logics that evolve over time within regions and organizations. They shape how enterprises understand the external environment, identify innovation opportunities, and assess the appropriateness and feasibility of different innovation activities. Unlike formal regulations and external supervision, cognitive institutions mainly operate through organizational innovation culture, integrity orientation, and environmental awareness, thereby influencing long-term decision-making and resource allocation. Existing research shows that organizational innovation culture can significantly affect enterprises' new product performance, with this effect moderated by the external institutional environment (Xie et al., 2021). Enterprises' integrity culture is also closely associated with sustainability performance and may exert influence through channels such as environmental and social responsibility (Bao et al., 2024). For agriculture- and forestry-related enterprises, openness, innovation culture, and green awareness may help transform external environmental changes into opportunities for technological upgrading and product innovation, whereas entrenched cognitive structures may reinforce path dependence and reduce enterprises' capacity to reallocate innovation resources under shocks. Accordingly, this paper measures cognitive institutional elements along two dimensions, regional cognitive environment and internal corporate cognition, using openness (*OPEN*), integrity orientation (*HON*), innovation culture (*EIC*), and green cognition (*EGC*).

Fifth, we include control variables. Following the literature on corporate innovation, which considers enterprise characteristics such as size, age, capital structure, growth opportunities, profitability, and cash flow (Babina et al., 2024), and based on data availability for agriculture- and forestry-related enterprises, we include enterprise size (*SIZE*), enterprise age (*FAGE*), asset-liability ratio (*LEV*), Tobin's Q (*TOBINQ*), operating revenue growth rate (*GROWTH*), operating cash flow ratio (*CFO*), largest shareholder's ownership stake (*TOP1*), board size (*BOARD*), employee compensation level (*WAGE*), capital intensity (*CAPINT*), management fee ratio (*MFR*), return on assets (*ROA*), current ratio (*CR*), cash flow volatility (*CFV*), and inventory turnover ratio (*ITR*) as baseline control variables. Definitions and measurements of these variables are provided in Supplementary Table S3.

3.1.2 Data collection and optimization of the indicator system

This study examines A-share listed enterprises engaged in agriculture and forestry. Data on enterprises' financial conditions, corporate governance, and institutional characteristics were obtained primarily from the CNRDS database and supplemented and matched with enterprises' annual reports, local government work reports, and other relevant publicly disclosed information. To improve data quality and consistency across variables, the raw data were screened and cleaned. Observations with abnormal listing statuses, such as ST and *ST, financial-sector enterprises, and observations with missing or abnormal values for key variables were excluded. Data from different sources were then matched, processed for missing values, and standardized. Because the study uses EIR_{t+1} as the target variable, the data-matching process also ensured temporal correspondence between institutional characteristics and enterprise characteristics in period t and innovation resilience in period t+1. EIR_{t+1} varies substantially across enterprises, with a mean of 3.5681 and a standard deviation of 25.2072, while the institutional elements and enterprise characteristics also show considerable variation, supporting subsequent machine-learning analysis. Descriptive statistics are reported in Supplementary Table S4

To assess linear correlations and potential information redundancy among the model input features, we conducted a Pearson correlation analysis of all variables (Dormann et al., 2013). As shown in Figure 5, the absolute correlation coefficients among the explanatory features were generally low, with a maximum of approximately 0.64, and no clear evidence of strong linear correlations was found. Accordingly, we did not further exclude variables on the basis of pairwise correlation coefficients and included all the features in the subsequent machine-learning model training and prediction analyses.

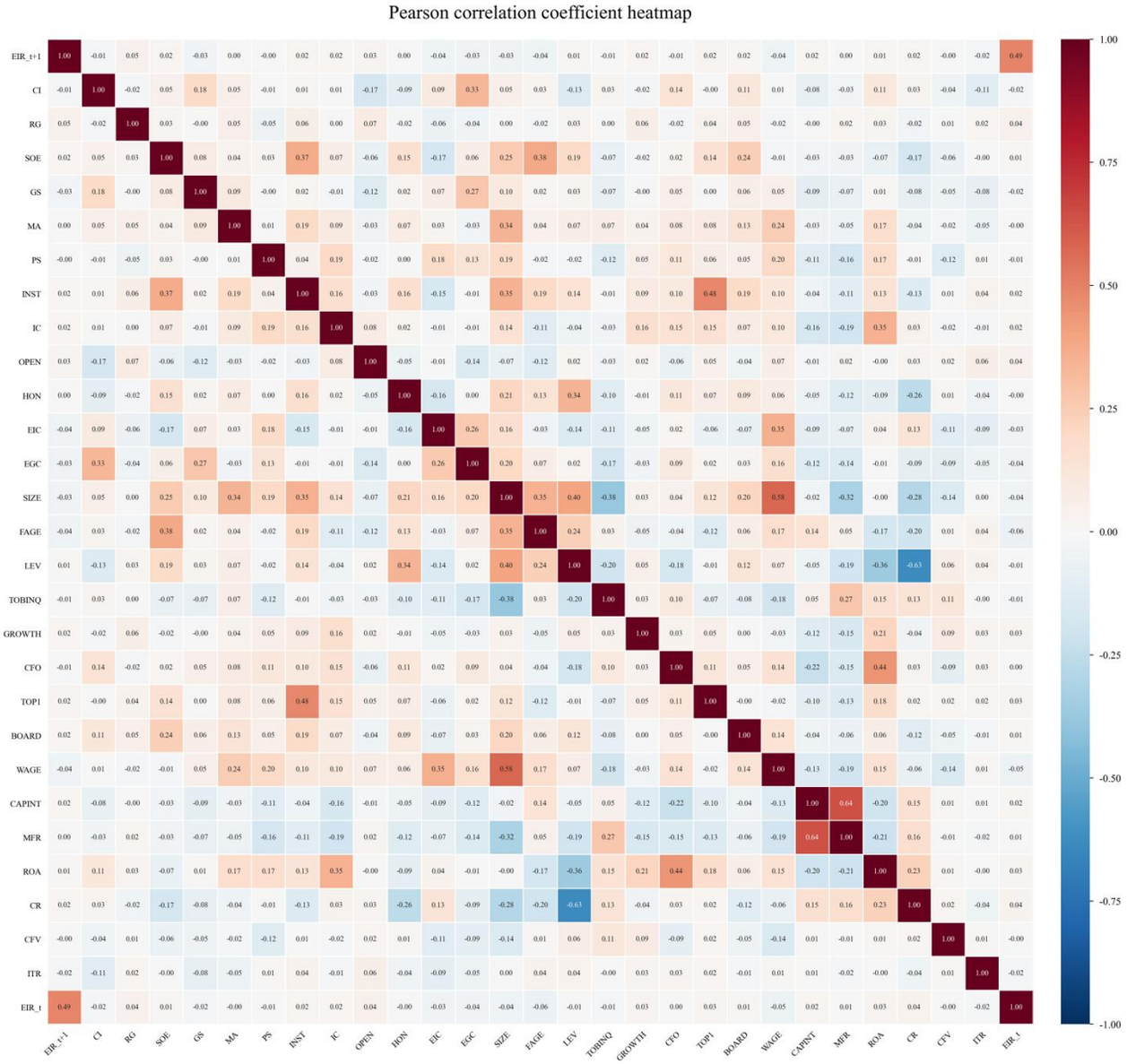


Fig. 5. Correlation analysis

3.2 CatBoost-based model for predicting EIR_{t+1}

CatBoost is employed because it can flexibly capture nonlinear relationships and interactions while improving generalizability through ordered boosting and categorical feature processing (Bentéjac et al., 2021). The model is compared with alternative machine-learning algorithms under cross-validation. Detailed formulations, encoding procedures, and evaluation metrics are provided in the Supplementary Information, while descriptions of the candidate prediction algorithms are provided in Supplementary Appendix A.

3.3 Model interpretation

SHAP is used to interpret the optimal model by quantifying feature contributions and identifying nonlinear predictive and interaction patterns (Chen et al., 2023; Lundberg et al., 2020). Detailed SHAP formulations and calculation procedures are provided in the Supplementary Information.

3.4 Model performance evaluation

Following standard machine-learning training and evaluation procedures, we divide the full sample into a training set comprising 80% of the observations and a test set comprising the remaining 20%. The training set is used to fit the models and optimize their parameters, while the test set assesses their out-of-sample predictive performance. All candidate models are trained exclusively on the training set and evaluated on observations not used in training. To further improve the

predictive performance of complex models, we conduct 5-fold cross-validation combined with hyperparameter search during training, thereby reducing the risk of overfitting and enhancing the robustness of the predictions (Mullainathan and Spiess, 2017).

Following the model specifications described above, this study constructs LR, RF, KNN, SVM, DT, NN, XGBoost, and CatBoost models. Because EIR_{t+1} is continuous, the prediction task is formulated as a regression problem. Model performance is evaluated along three dimensions: in-sample fit, out-of-sample generalization, and prediction error. Specifically, the evaluation metrics include the in-sample coefficient of determination (R_{is}^2), out-of-sample coefficient of determination (R_{oos}^2), in-sample mean squared error (MSE_{is}), in-sample mean absolute error (MAE_{is}), out-of-sample mean squared error (MSE_{oos}), and out-of-sample mean absolute error (MAE_{oos}). Higher R_{is}^2 and R_{oos}^2 values indicate stronger in-sample fit and out-of-sample generalization, respectively, whereas lower *MSE* and *MAE* values indicate smaller prediction errors.

Table 1 reports the model fit and predictive performance across different models. Both KNN and DT achieve an in-sample R_{is}^2 of 100%, with in-sample *MSE* and *MAE* values of 0. However, their out-of-sample R_{oos}^2 values are only 25.7679% and -77.4296%, respectively, indicating pronounced overfitting. By contrast, CatBoost achieves the highest out-of-sample R_{oos}^2, at 43.6747%, followed by XGBoost at 43.0946% and RF at 42.4256%. In terms of out-of-sample prediction errors, CatBoost has the lowest MSE_{oos}, at 317.8345, while its MAE_{oos} is 7.0670, which is also relatively low.

Considering both out-of-sample fit and prediction error, CatBoost achieved a favorable overall balance across the evaluation metrics. Accordingly, we ultimately selected CatBoost as the benchmark predictive model for enterprises' innovation resilience and used it in the subsequent analysis to estimate feature importance and conduct SHAP-based interpretation. Detailed model-training and hyperparameter settings are provided in Supplementary Appendix B.

Table 1. Comparison of model performance

Model	R_{is}^2(%)	MSE_{is}	MAE_{is}	R_{oos}^2(%)	MSE_{oos}	MAE_{oos}
LR	23.1481	501.8776	8.4419	28.6976	402.3476	8.1311
RF	59.9547	261.5138	5.9930	42.4256	324.8829	7.0866
KNN	100.0000	0.0000	0.0000	25.7679	418.8796	8.6467
SVM	23.9721	496.4967	7.1967	29.1469	399.8120	7.0245
DT	100.0000	0.0000	0.0000	-77.4296	1001.2055	11.4744
XGBoost	62.7949	242.9662	5.7999	43.0946	321.1076	7.1403
NN	33.3530	435.2354	7.8965	35.2134	365.5801	7.8054
CatBoost	52.0597	313.0719	6.5234	43.6747	317.8345	7.0670

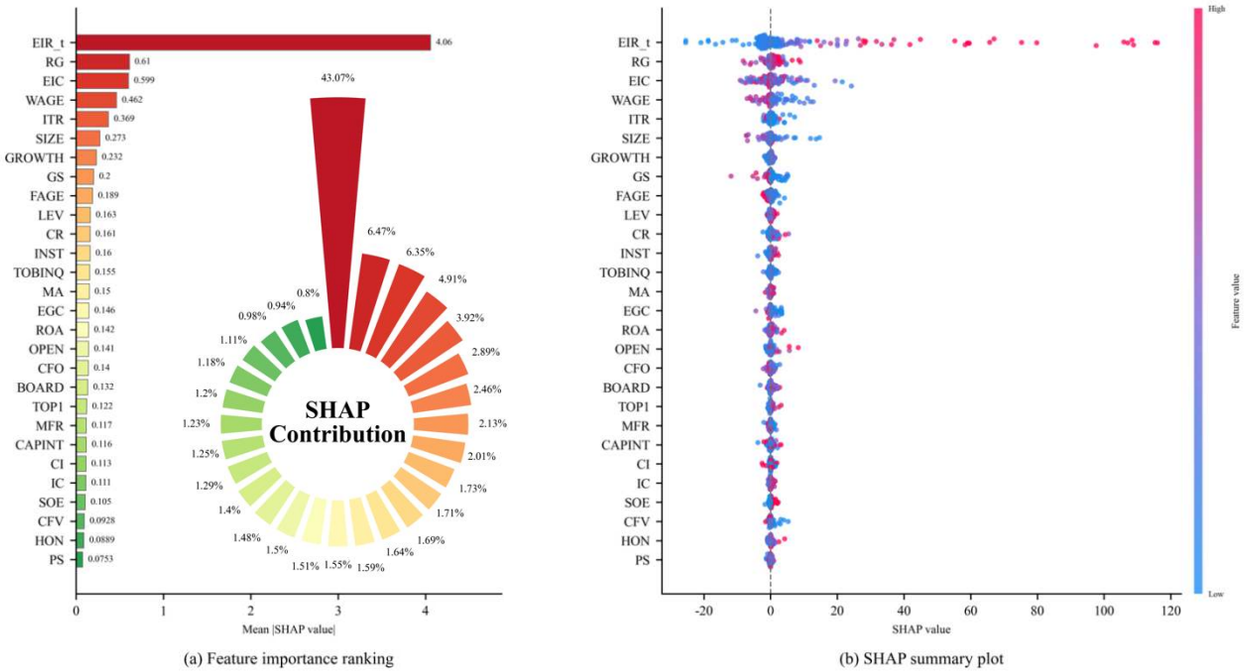
4 Analysis of empirical results

4.1 Comparative analysis of feature importance

As shown in Fig. 6(a), the predictive contributions of different features to EIR_{t+1} exhibit pronounced differences in magnitude. EIR_t is the most important feature for predicting EIR_{t+1}, with a SHAP importance value of 43.0710%, indicating strong dynamic persistence in *EIR*. Among the *IE* variables, *RG* and *EIC* have the highest SHAP importance values, at 6.4730% and 6.3546%, respectively. Among the control variables, *WAGE*, *ITR*, *SIZE*, and *GROWTH* have SHAP importance values of 4.9077%, 3.9171%, 2.8943%, and 2.4625%, respectively. Fig. 6(b) further shows that EIR_t exhibits the widest range of SHAP values, while *RG*, *EIC*, and *WAGE* also display a certain degree of heterogeneity in their predictive contributions. By contrast, the SHAP values of most other features are relatively concentrated, indicating that the model's predictive information is primarily concentrated in a small number of key features. The specific nonlinear patterns of these features are examined further below.

Table 2 further compares feature importance within the three categories of *IE*. Among the regulative elements, *RG* has the highest SHAP importance value (6.4730%), followed by *GS* (2.1280%), *CI* (1.1983%), and *SOE* (1.1110%). Among the normative elements, the SHAP importance values of *INST* and *MA* are 1.6944% and 1.5908%, respectively. Among the cognitive elements, *EIC* has the highest SHAP importance value (6.3546%), substantially exceeding those of

EGC (1.5470%), *OPEN* (1.4984%), and *HON* (0.9436%). When aggregated by institutional dimension, the SHAP importance values of the regulative, normative, and cognitive elements are 10.9103%, 5.2626%, and 10.3436%, respectively. These results indicate that the institutional predictive information for EIR_{t+1} is more concentrated in the regulative and cognitive dimensions, while substantial differences in feature importance also exist within the individual institutional dimensions.



(a) Feature importance ranking (b) SHAP summary plot

Fig. 6. Variable importance ranking based on SHAP values

Table 2. Top 5 variables by SHAP importance within each category

No.	Regulative elements	Normative elements	Cognitive elements	Other predictors
1	*RG*(6.4730%)	*INST*(1.6944%)	*EIC*(6.3546%)	*EIR*$_{t}$(43.0710%)
2	*GS*(2.1280%)	*MA*(1.5908%)	*EGC*(1.5470%)	*WAGE*(4.9077%)
3	*CI*(1.1983%)	*IC*(1.1784%)	*OPEN*(1.4984%)	*ITR*(3.9171%)
4	*SOE*(1.1110%)	*PS*(0.7990%)	*HON*(0.9436%)	*SIZE*(2.8943%)
5	—	—	—	*GROWTH*(2.4625%)

4.2 Feature prediction patterns

To further examine the nonlinear predictive patterns of *IE* in predicting EIR_{t+1}, we plot SHAP dependence plots for each institutional variable based on the SHAP feature-importance analysis. The horizontal axis shows the feature value, and the vertical axis shows its corresponding SHAP value. The fitted curves trace changes in each feature's predictive contribution across value ranges, allowing potential stages, turning points, and marginal changes to be identified. The turning-point ranges discussed below are inferred from the fitted trends in the SHAP dependence plots and should not be interpreted as statistically estimated threshold parameters. Figure 7 presents the SHAP dependence relationships for the predictor variables.

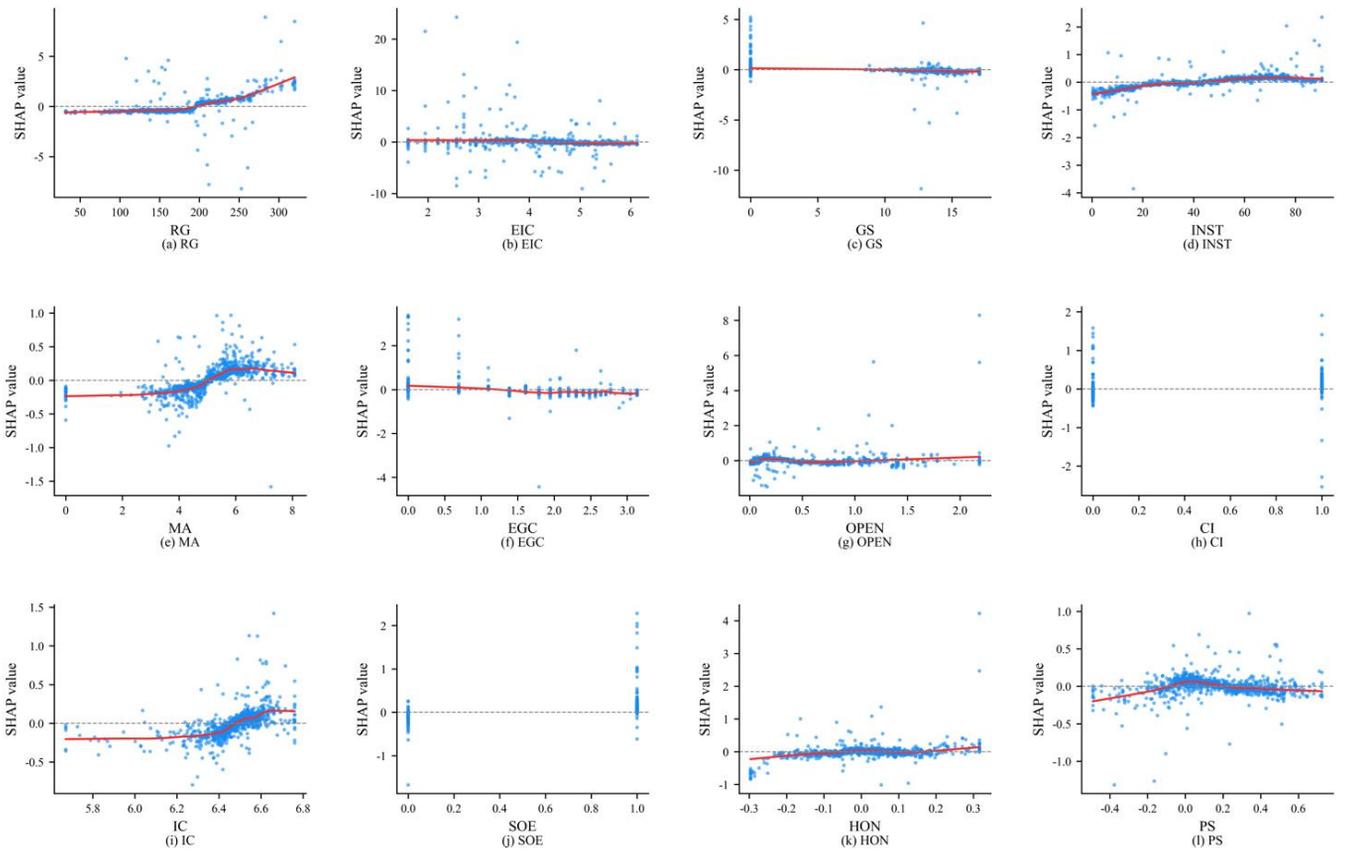


Fig. 7. SHAP dependence plots of predictor variables ranked by feature importance

Fig. 7 shows clear nonlinear and joint predictive patterns between *IE* and EIR_{t+1}. Among the regulative elements, *RG* exhibits the most pronounced variation. When *RG* is below approximately 180, SHAP values generally remain near zero or in the negative range. Within approximately [180, 220], the fitted curve rises markedly, with SHAP values gradually shifting from negative to positive. Above approximately 220, the positive predictive contribution strengthens further, with SHAP values reaching approximately 2–3 at high levels of *RG*. The fitted curve for *GS* fluctuates around zero and exhibits limited overall variation. *CI* and *SOE* are discrete variables; although the distribution of observations differs across categories, neither exhibits a continuous monotonic relationship. Among the normative elements, *INST* shows a gradual upward trend as its value increases. *MA* makes a relatively weak contribution at low values, but its SHAP values rise markedly within approximately the 4–6 range, with the increase slowing beyond approximately 6. SHAP values for *IC* remain generally low below approximately 6.2, begin to increase within approximately [6.2, 6.5], and gradually become positive at higher values. *PS* exhibits relatively weak nonlinear fluctuations overall. Among the cognitive elements, the SHAP values of *OPEN* and *HON* gradually shift from negative values toward zero or positive values as the variables increase. The fitted curves for *EIC* and *EGC* are relatively flat, indicating limited variation in their predictive contributions. Overall, *RG*, *MA*, and *IC* exhibit comparatively clear stage-specific changes, whereas the other institutional variables display varying degrees of flat, discrete, or weakly nonlinear patterns.

4.3 Feature interaction analysis

To further examine the joint predictive patterns of different *IE* variables, we selected three representative variables based on their SHAP importance rankings: *RG*, *EIC*, and *GS*. We then plotted SHAP interaction dependence plots for *RG*×*EIC*, *RG*×*GS*, and *EIC*×*GS*. In each plot, the horizontal axis indicates the value of the target feature, the vertical axis shows its corresponding SHAP value, and the color gradient from low to high represents the value of the other interacting feature. These plots reveal how predictive contributions vary across different combinations of institutional elements, as shown in Figure 8.

Fig. 8 shows marked differences in the joint predictive patterns across the three pairs of *IE* variables. For *RG* × *EIC*, when *RG* is below approximately 180, SHAP values are concentrated around zero or in the negative range. Within approximately [180, 220], the fitted curve rises rapidly and gradually turns positive. Above approximately 220, the positive predictive contribution continues to strengthen, reaching approximately 2–3 at high values. Samples with different *EIC* levels still overlap substantially in the medium-to-high *RG* range, indicating that the predictive contribution of *RG* mainly differs in magnitude across *EIC* levels rather than exhibiting a stable unidirectional stratification. *RG* × *GS* exhibits a similar stage-specific pattern: the predictive contribution of *RG* is relatively weak below approximately 180, increases rapidly within approximately [180, 220], and remains clearly positive above approximately 220. Samples with different *GS* levels are interspersed throughout the main value range, suggesting that *GS* does not exhibit a clear monotonic reinforcement pattern. By contrast, the fitted curve for *EIC* × *GS* remains generally close to zero and shows a slight downward trend. Samples are relatively dispersed when *EIC* is approximately 2–4 but gradually converge at higher values. The color distributions associated with different *GS* levels also overlap substantially, indicating that the joint predictive relationship between *EIC* and *GS* is generally weaker than those of the other two pairs. Overall, *RG* is the core variable exhibiting the clearest stage-specific pattern across the three joint predictive relationships, whereas *EIC* and *GS* primarily exhibit conditional and heterogeneous differences.

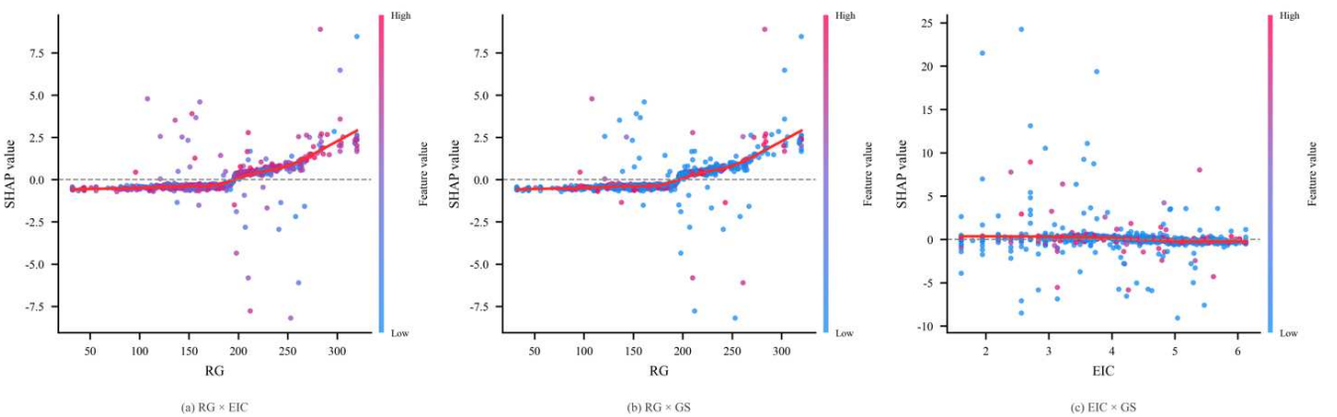


Fig. 8. Interaction patterns among key institutional elements

5 Predictive analysis of multidimensional interactions from a global perspective

5.1 Construction of the *IE* indices

The preceding analysis examined the predictive roles of individual *IE* variables in predicting EIR_{t+1} for agriculture- and forestry-related enterprises. However, a single indicator cannot fully capture the combined information conveyed by multiple characteristics within the same institutional dimension. Based on Shapley values, SHAP decomposes model predictions into feature-level contributions, enabling global comparisons of the relative predictive contributions of different features and providing a methodological basis for integrating multidimensional institutional information (Chen et al., 2023). Accordingly, to further examine the predictive role of the multifaceted institutional environment at the overall institutional-dimension level, this study constructed the Regulative Institutional Index (*RII*), Normative Institutional Index (*NII*), and Cognitive Institutional Index (*CII*), respectively.

Specifically, using the optimal machine-learning model identified above, this study measures the relative predictive importance of each individual institutional characteristic by its mean absolute SHAP value and uses this importance to determine the weight of each indicator within its corresponding institutional dimension. First, the institutional indicators are standardized. Second, they are weighted and aggregated according to their relative SHAP importance to construct *RII*, *NII*, and *CII*. Finally, the three composite indices are normalized to the [0,1] interval using the Min-Max method to improve comparability across indices. These weights represent relative information contributions within the specified predictive model rather than structural causal effects. As shown in Table 3, the mean values of *RII*, *NII*, and *CII* are 0.4721, 0.5921,

and 0.5573, respectively, with corresponding standard deviations of 0.1539, 0.1315, and 0.1674. These results indicate that the three composite *IE* indices exhibit some degree of variation across the sampled enterprises, providing a basis for further analysis.

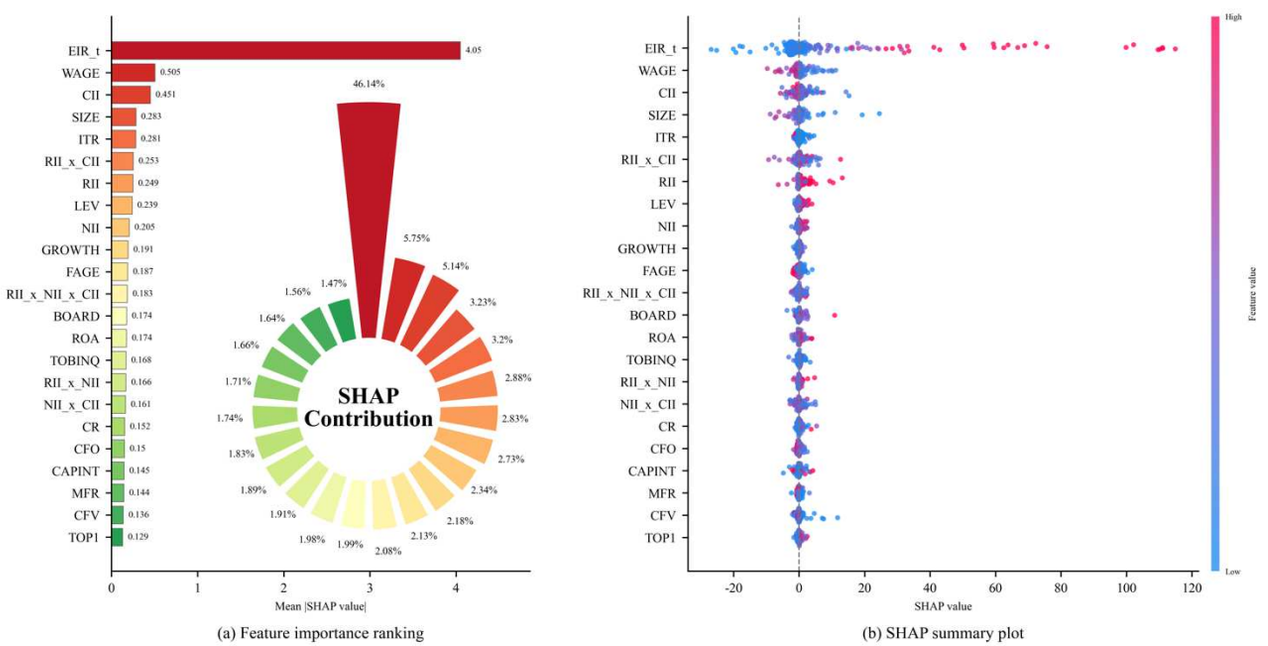
Table 3. Descriptive statistics of the three composite *IE* indices

Composite *IE* index	Variable	count	mean	std	min	max
Regulative *IE* index	*RII*	14854	0.4721	0.1539	0	1
Normative *IE* index	*NII*	14854	0.5921	0.1315	0	1
Cognitive *IE* index	*CII*	14854	0.5573	0.1674	0	1

5.2 Full-sample analysis of *IE*

Building on the composite indices for the three *IE* dimensions, this study further uses *RII*, *NII*, and *CII*, their pairwise interaction terms (*RII* × *NII*, *RII* × *CII*, and *NII* × *CII*), the three-way interaction term (*RII* × *NII* × *CII*), and the control variables as input features, and retrains the models with hyperparameter optimization and performance evaluation. A comparison of the models' out-of-sample fit and prediction errors shows that CatBoost performs relatively well, with an R^2_{oos} of 42.6236% and an MSE_{oos} of 323.7657. Accordingly, the following analysis is based on the CatBoost results to further examine the joint predictive characteristics of the three *IE* dimensions, as shown in Fig. 9. Detailed model-training and performance results are provided in Supplementary Appendix C.

(1) Relative importance analysis. Figure 9 shows that EIR_t remains the most important feature for predicting EIR_{t+1}, with a SHAP importance of 46.14%. This finding indicates strong dynamic persistence in *EIR*. Among the three composite institutional indices, *CII* has the highest importance at 5.14%, followed by *RII* and *NII* at 2.83% and 2.34%, respectively. Notably, joint information across institutional dimensions also has considerable predictive value. *RII*×*CII* has an importance of 2.88%, higher than any other interaction term, while *RII*×*NII*×*CII*, *RII*×*NII*, and *NII*×*CII* have importance values of 2.08%, 1.89%, and 1.83%, respectively. Together, the three institutional indices and their interaction terms contribute approximately 18.99%. These results suggest that information predictive of innovation resilience among agriculture- and forestry-related enterprises derives not only from individual institutional dimensions but also partly from the combinations of different institutional dimensions.



(a) Feature importance ranking (b) SHAP summary plot

Fig. 9. Variable importance ranking based on SHAP values

(2) Nonlinear and interactive predictive patterns. Figure 10 further shows that all three composite institutional indices exhibit nonlinear patterns to varying degrees. *RII*'s SHAP values remain generally close to zero or slightly negative at low

and lower-middle levels, begin to increase at approximately 0.5–0.6, and make a markedly stronger positive contribution to predictions above approximately 0.75, indicating a relatively clear stagewise upward pattern. *NII*'s SHAP values are generally close to zero or negative below approximately 0.5, show a relatively distinct turning point at approximately 0.55–0.65, and then rapidly become positive, reaching a relatively high level around 0.7 before leveling off. By contrast, *CII* varies relatively little across the low-to-middle range. Its fitted curve gradually declines above approximately 0.5, with high values more often corresponding to negative SHAP values. These ranges are identified from the smoothed trends in the SHAP dependence plots and describe stagewise changes in the predictive relationships; they should not be interpreted as statistically estimated threshold parameters.

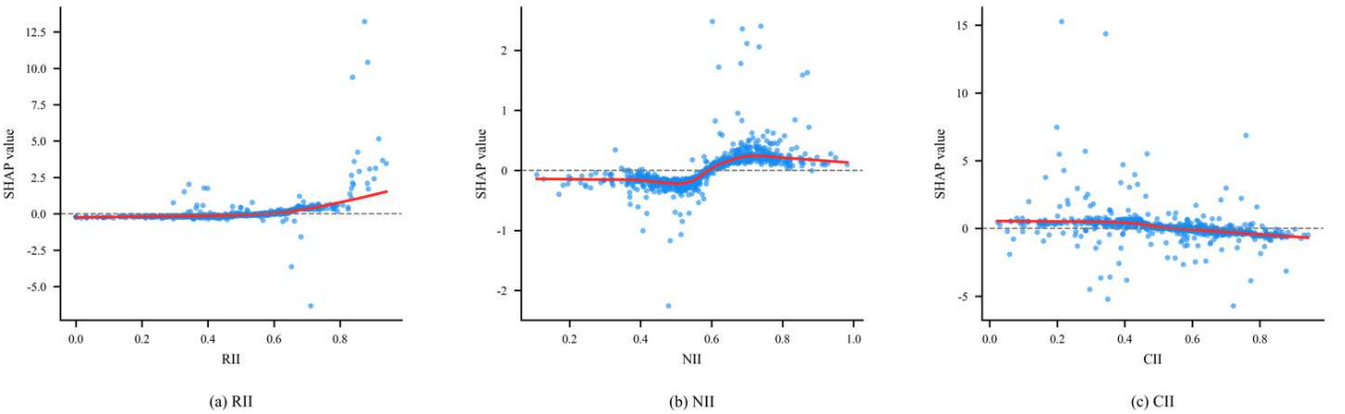


Fig. 10. SHAP dependence plots for the three categories of *IE* composite indices

(3) Feature interaction analysis. Figure 11 shows that the joint predictive relationships among the three institutional indices are also markedly nonlinear. The overall predictive contributions of both *RII*×*NII* and *RII*×*CII* increase primarily with *RII*. At low levels of *RII*, SHAP values are generally close to zero or negative; above approximately 0.6, they gradually become positive and rise markedly at high levels of *RII*. However, the color distributions associated with different *NII* and *CII* values still overlap substantially, suggesting that these interactions primarily capture conditional differences in magnitude rather than stable unidirectional stratification. By contrast, *NII*×*CII* exhibits a relatively clear transition across stages. Its predictive contribution is relatively low when *NII* is around 0.5, rapidly becomes positive once *NII* reaches approximately 0.55–0.65, and stabilizes at higher levels of *NII*. These findings further suggest that the multidimensional institutional environment is not simply the sum of several independent institutional factors but may exhibit distinct joint predictive patterns across institutional dimensions.

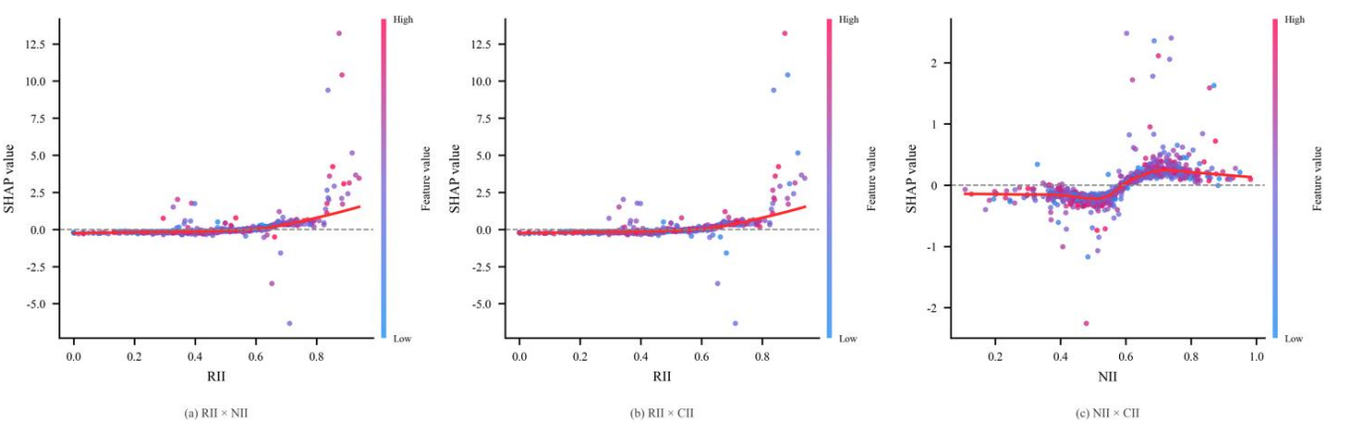
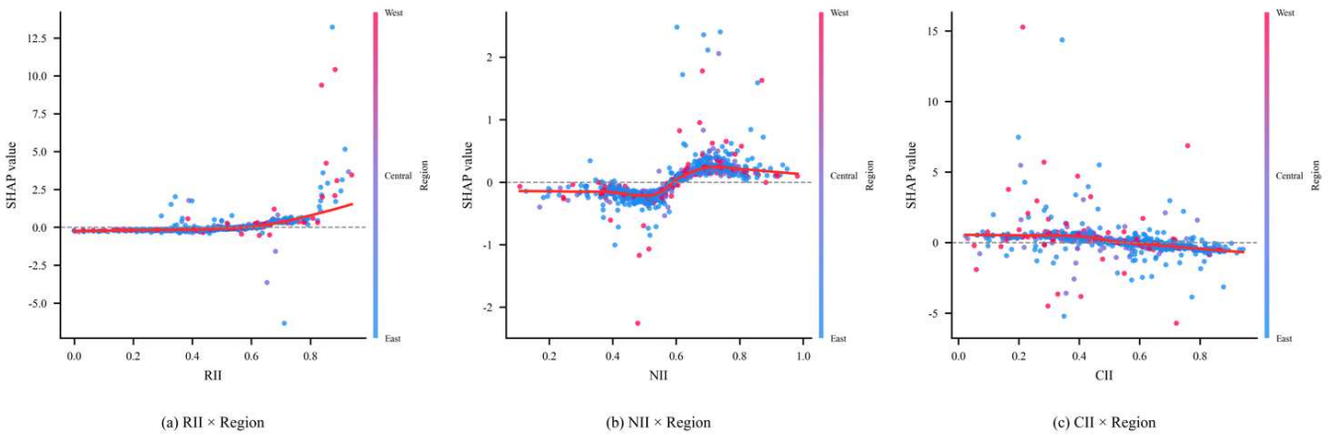


Fig. 11. Interaction patterns among the three categories of *IE* composite indices

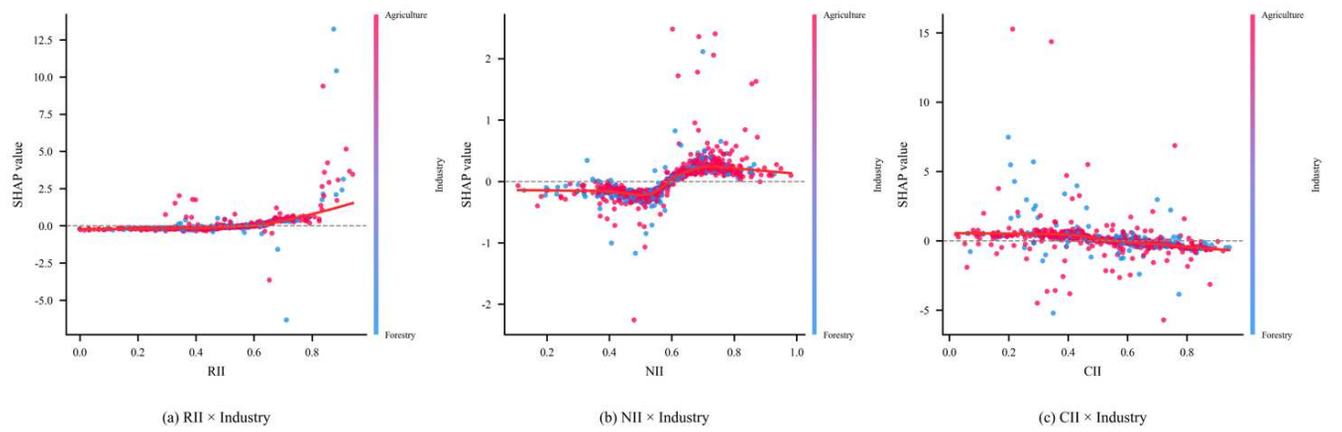
Figures 12-14 further illustrate the conditional predictive patterns of the three composite institutional indices across regions, enterprise types, and enterprise sizes. Overall, *RII* shows a relatively consistent pattern of stronger contributions at higher values across all three contextual dimensions: its contribution gradually becomes positive after approximately

0.6 and rises markedly in the high-value range. *NII* exhibits a stage-dependent pattern, shifting from a weakly negative or near-zero contribution at low to medium values to a positive contribution at medium to high values. By contrast, the contribution of *CII* generally weakens as the index increases. The sample distributions also differ to some extent across regions, enterprise types, and enterprise sizes, indicating that the predictive role of the composite institutional environment is context-dependent. Recent research has similarly found that the effects of different types of institutional pressure on enterprise innovation vary with regional conditions for innovation support (Lee et al., 2024). Figures 12–14 therefore highlight potential conditional differences within the full-sample framework, while further between-group differences are identified in the next section by modeling the subsamples separately.



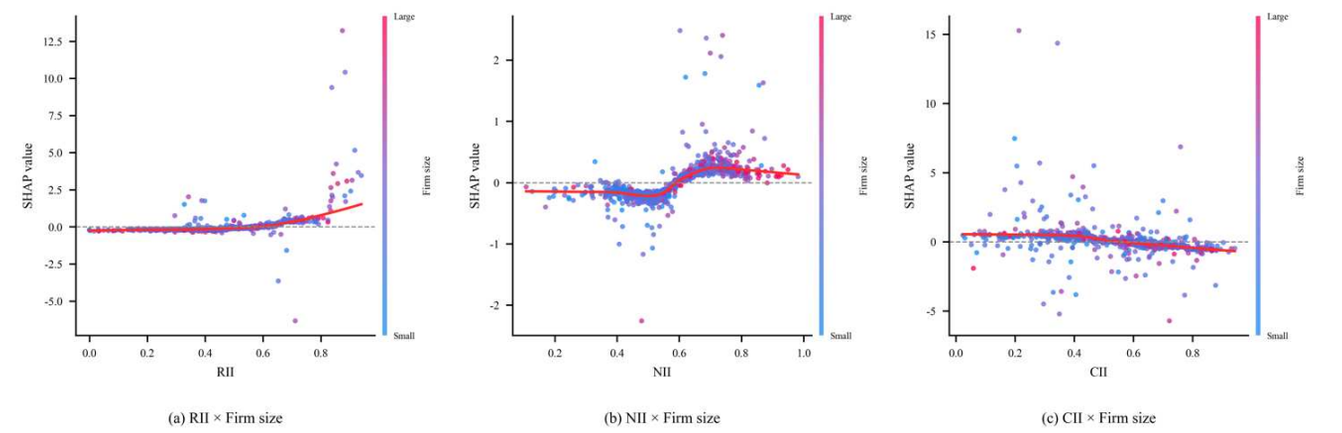
(a) RII × Region (b) NII × Region (c) CII × Region

Fig. 12. Interaction patterns between the three categories of *IE* composite indices and region



(a) RII × Industry (b) NII × Industry (c) CII × Industry

Fig. 13. Interaction patterns between the three categories of IE composite indices and enterprise type



(a) RII × Firm size (b) NII × Firm size (c) CII × Firm size

Fig. 14. Interaction patterns between the three categories of IE composite indices and enterprise size

5.3 Heterogeneity analysis across groups

Building on the full-sample analysis, this study further examines group differences in how *RII*, *NII*, *CII*, and their interaction terms predict EIR_{t+1} for agriculture- and forestry-related enterprises. Unlike the preceding full-sample interaction plots, this section estimates separate subsample models by region, ownership, enterprise size, and enterprise type. For each group, the models are retrained, their parameters are optimized, and their out-of-sample performance is evaluated. The better-performing model is then selected for each subsample, and SHAP is used to compare the relative importance of the institutional composite indices, interaction terms, and enterprise characteristics. Regions are classified as eastern, central, or western; enterprises as state-owned or non-state-owned; enterprises by size as large or small; and enterprise types as agriculture-related or forestry-related. Table 4 reports the feature-importance rankings for each group. Detailed subgroup model-training and hyperparameter settings are provided in Supplementary Appendix D.

Table 4 first shows that EIR_t ranks first across all nine subsamples, with SHAP importance values ranging from 35.3016% to 47.7834%. This indicates that the dynamic persistence of *EIR* is not confined to any particular group. Regarding regional heterogeneity, *CII* is relatively prominent among the institutional elements in both the eastern and central regions, with importance values of 4.6240% and 5.2475%, respectively. In the western region, institutional configurations are more pronounced: *NII* × *CII* has an importance value of 6.9302%, ranking second among all explanatory features; *CII* and *RII* have importance values of 6.0735% and 4.8322%, respectively; while *RII* × *NII*, *RII* × *NII* × *CII*, and *RII* × *CII* have importance values of 4.0449%, 3.7533%, and 3.6847%, respectively. These results indicate that, compared with the eastern and central regions, the predictive information for EIR_{t+1} in the western region is more strongly associated with multidimensional *IE* and their configurations.

By ownership type, *RII* has an importance of 4.1196% among state-owned enterprises, exceeding the 3.9147% importance of *CII* and indicating that the formal governance environment makes a relatively greater predictive contribution in this subsample. Among non-state-owned enterprises, the importance of *CII* rises to 5.9663%, while that of *RII*×*CII* reaches 4.4204%, suggesting that cognitive institutions and the combined information captured by the regulative–cognitive interaction are more prominent. By enterprise size, the importance of *CII*, *NII*, and the three-way interaction *RII*×*NII*×*CII* among large enterprises is 7.5862%, 5.0368%, and 4.7494%, respectively, indicating that their predicted innovation resilience is more strongly associated with the overall institutional environment. Although the SHAP importance of *CII* remains high at 6.3842% among small enterprises, that of EIR_t reaches 47.7834%, markedly exceeding the 36.5739% observed among large enterprises and indicating stronger state persistence. Differences in resource endowments and organizational attributes may affect enterprises' capacity to absorb and respond to external institutional signals. Recent research has also found substantial heterogeneity across ownership type and enterprise size in the effect of the intellectual property judicial system on enterprise innovation, with private and small enterprises exhibiting more pronounced responses (Y. Zhang et al., 2024).

Heterogeneity across enterprise types further shows that, among agriculture-related enterprises, the importance of *RII*×*CII* is 4.5059%, while those of *CII* and *RII* are 4.0879% and 3.7480%, respectively. This indicates that the combined information conveyed by regulative and cognitive institutions is relatively prominent in predicting the innovation resilience of agriculture-related enterprises. By contrast, among forestry-related enterprises, the importance of *CII* reaches 7.6225%, while those of enterprise size, employee compensation level, and capital intensity are 7.4771%, 6.1575%, and 6.0672%, respectively. This suggests that the predictive information for the innovation resilience of forestry-related enterprises is more concentrated in the cognitive institutional environment and enterprises' own resources and operational characteristics. Overall, the importance of the three institutional composite indices and their interaction terms varies markedly across regions, ownership types, enterprise sizes, and enterprise types. This indicates that the predictive relationship between the multidimensional institutional environment and the innovation resilience of agriculture- and forestry-related enterprises is strongly context-dependent and cannot be adequately characterized by a single institutional factor or a uniform linear relationship.

Table 4. Feature importance rankings across subgroups

Rank	Regional heterogeneity						Ownership heterogeneity				Enterprise size heterogeneity				Enterprise type heterogeneity			
	Eastern region		Central region		Western region		State-owned enterprises		Non-state-owned enterprises		Large enterprises		Small enterprises		Agriculture-related enterprises		Forestry-related enterprises	
No.	Predictor	Importance	Predictor	Importance	Predictor	Importance	Predictor	Importance	Predictor	Importance	Predictor	Importance	Predictor	Importance	Predictor	Importance	Predictor	Importance
1	*EIR_t*	37.0570%	*EIR_t*	38.6864%	*EIR_t*	39.8222%	*EIR_t*	41.8353%	*EIR_t*	43.2053%	*EIR_t*	36.5739%	*EIR_t*	47.7834%	*EIR_t*	35.3016%	*EIR_t*	35.4563%
2	*WAGE*	9.9767%	*SIZE*	6.3040%	*NII×CII*	6.9302%	*WAGE*	6.2795%	*CII*	5.9663%	*CII*	7.5862%	*CII*	6.3842%	*CFV*	4.7842%	*CII*	7.6225%
3	*CII*	4.6240%	*CII*	5.2475%	*CII*	6.0735%	*RII*	4.1196%	*RII×CII*	4.4204%	*WAGE*	5.5979%	*WAGE*	5.6001%	*RII×CII*	4.5059%	*SIZE*	7.4771%
4	*MFR*	4.0021%	*CFV*	4.6593%	*SIZE*	5.3588%	*CII*	3.9147%	*CR*	4.3275%	*NII*	5.0368%	*SIZE*	4.4411%	*SIZE*	4.2753%	*WAGE*	6.1575%
5	*CAPINT*	3.5286%	*CFO*	4.2399%	*RII*	4.8322%	*SIZE*	3.8410%	*WAGE*	4.0538%	*RII×NII×CII*	4.7494%	*ITR*	3.5957%	*CII*	4.0879%	*CAPINT*	6.0672%
6	*FAGE*	3.4052%	*FAGE*	4.1295%	*RII×NII*	4.0449%	*ROA*	3.5807%	*CAPINT*	3.2171%	*CAPINT*	3.6401%	*RII×CII*	3.1193%	*FAGE*	3.9624%	*CFO*	3.4954%
7	*TOP1*	2.9810%	*ROA*	3.6445%	*RII×NII×CII*	3.7533%	*LEV*	3.2806%	*SIZE*	3.0951%	*SIZE*	3.4863%	*GROWTH*	2.2639%	*ITR*	3.7657%	*MFR*	3.1963%
8	*LEV*	2.9711%	*CR*	3.2922%	*RII×CII*	3.6847%	*RII×NII×CII*	3.2325%	*FAGE*	2.7272%	*LEV*	3.3288%	*CAPINT*	2.2332%	*RII*	3.7480%	*LEV*	2.7671%
9	*ROA*	2.9180%	*NII×CII*	3.1671%	*ITR*	3.6807%	*TOBINQ*	2.8192%	*ROA*	2.7143%	*ITR*	3.2395%	*CR*	2.1501%	*MFR*	3.3467%	*GROWTH*	2.4452%
10	*SIZE*	2.7453%	*RII×CII*	3.0935%	*CAPINT*	3.1669%	*ITR*	2.6876%	*RII×NII×CII*	2.6730%	*RII*	3.2333%	*NII×CII*	2.0496%	*LEV*	3.3081%	*CR*	2.4323%

6 Robustness and sensitivity analysis

6.1 Replacing the GBDT implementation method

To examine whether the baseline results depend on a particular implementation of the gradient boosting decision tree (GBDT) algorithm, we replaced the baseline CatBoost model with XGBoost and retrained the model for prediction while keeping the sample split, input features, and model evaluation metrics unchanged. As shown in Supplementary Appendix E (Table E.1), the out-of-sample coefficient of determination (R^2_{oos}) of XGBoost is 43.0946%, which is close to the 43.6747% obtained from the baseline CatBoost model. Its out-of-sample mean squared error (MSE_{oos}) and mean absolute error (MAE_{oos}) are 321.1076 and 7.1403, respectively, compared with 317.8345 and 7.0670 for the baseline model, indicating only minor differences. Although the in-sample coefficient of determination (R^2_{is}) of XGBoost increases to 62.7949%, its out-of-sample predictive performance does not change substantially. Overall, replacing the GBDT implementation leaves the model's out-of-sample predictive performance largely unchanged, indicating that the main results are not sensitive to the specific algorithmic implementation.

6.2 Replacing core parameters

Model predictive performance may be sensitive to hyperparameter settings. To further assess the sensitivity of the baseline results to hyperparameter perturbations, this study varied the number of iterations and the learning rate of CatBoost while keeping the sample split, input features, and model structure unchanged. Specifically, the number of iterations was set to 200 or 1000, and the learning rate to 0.01 or 0.05, yielding four hyperparameter combinations for model retraining and evaluation. The results show that, under the baseline settings of 1000 iterations and a learning rate of 0.01, the out-of-sample coefficient of determination (R^2_{oos}) is 43.6747%, while the out-of-sample mean squared error (MSE_{oos}) and mean absolute error (MAE_{oos}) are 317.8345 and 7.0670, respectively, indicating the best overall predictive performance among the four specifications. When the number of iterations is 1000 and the learning rate is 0.05, R^2_{oos} is 43.5952%; when the number of iterations is 200 and the learning rate is 0.05, R^2_{oos} is 42.9925%. Both results are relatively close to the baseline result. By contrast, when the number of iterations is reduced to 200 and the learning rate is set to 0.01, R^2_{oos} declines to 35.5069%, accompanied by a corresponding increase in prediction errors. Overall, the model's out-of-sample predictive performance remains broadly stable within a reasonable range of hyperparameter perturbations. At the same time, the combination of fewer iterations and a lower learning rate is associated with weaker model-fitting capacity, further supporting the appropriateness of the baseline hyperparameter settings.

6.3 Alternative sample period

To examine whether the findings are sensitive to the choice of sample period, we extended the sample window to 2008-2025 and retrained the model while keeping the variable definitions, data-processing procedures, and model evaluation criteria unchanged. After extending the sample window, R^2_{is} is 51.4064%, broadly consistent with the baseline model's 52.0597%. In contrast, R^2_{oos} decreases to 30.9451%, below the baseline value of 43.6747%, while the corresponding MSE_{oos} and MAE_{oos} are 476.6978 and 7.6783, respectively. These results indicate that expanding the sample period increases the difficulty of out-of-sample prediction to some extent, although the model continues to retain positive and non-negligible out-of-sample predictive power. Taken together with the results of the algorithm replacement

and hyperparameter perturbation analyses, the main predictive conclusions remain broadly robust, while the predictive performance for enterprise innovation resilience shows some dependence on temporal context.

7. Discussion

The *EIR* of agriculture- and forestry-related enterprises exhibits strong dynamic persistence, while the predictive contributions of different institutional dimensions differ markedly. EIR_t has the highest SHAP importance (43.0710%) overall and across all subsamples, indicating strong path dependence in innovation resilience. This finding is consistent with recent research on organizational resilience, which suggests that innovation resources and strategies developed before a shock can enhance enterprises' stability and adaptability amid external disturbances (Engelen et al., 2024). Among the institutional elements, *RG*, *INST*, and *EIC* are important predictive features within the regulative, normative, and cognitive dimensions, respectively, with SHAP importance values of 6.4730% for *RG* and 6.3546% for *EIC*. This suggests that *EIR* depends on both the formal governance environment and enterprises' cognitive and cultural foundations. Peng and Kong (2024) show that enterprises' green innovation responses to environmental regulation are influenced by their environmental strategic behavior; Zhao et al. (2024) further find that ownership type can alter the relationship between environmental regulation and enterprises' green innovation. Innovation culture can influence enterprises' long-term performance by affecting resource integration and innovation processes (Li et al., 2025). Meanwhile, corporate innovation is driven not only by external institutional pressures: internal value orientations and external normative requirements may both shape enterprises' innovation choices (Nguyen and Vu, 2024), while the influence of institutional investors on corporate innovation is also constrained by social norms and the governance environment (Kwack and Kim, 2025). Furthermore, stakeholder pressures from different sources may elicit differentiated innovation responses, and enterprises' internal characteristics and external environments can affect how these responses take shape (Dong et al., 2026). Together, these findings suggest that *EIR* is associated with multidimensional institutional conditions as well as enterprises' cognitive and capability foundations.

More importantly, the predictive relationship between *IE* and EIR_{t+1} is nonlinear and configurational. For *RG*, SHAP values are near zero or negative below approximately 180, rise rapidly within the [180, 220] interval, and become more strongly positive above 220. *RG*×*EIC* and *RG*×*GS* show similar stage-dependent patterns, whereas *EIC*×*GS* exhibits a weaker joint predictive relationship. Among the composite institutional indices, *CII*, *RII*, and *NII* have importance values of 5.14%, 2.83%, and 2.34%, respectively, while *RII*×*CII* reaches 2.88%; together, the three indices and their interaction terms contribute approximately 18.99%, indicating that predictive information is partly embedded in multidimensional institutional configurations. Existing studies likewise show that enterprises' responses to pressure from external stakeholders are highly enterprise-specific and are influenced by internal conditions such as exploratory and exploitative innovation capabilities (Jakhar et al., 2020). In the context of environmental regulation, regulatory effects also vary with innovation type, enterprise characteristics, and governance conditions (Lian et al., 2022). Furthermore, Ye et al. (2024) find that enterprises adopt selective coupling strategies when confronted with competing institutional logics, suggesting that multiple institutional demands do not simply accumulate. Research on institutional complexity further indicates that when organizations face multiple and potentially conflicting institutional logics, their actual responses are jointly shaped by organizational characteristics and field conditions (Greenwood et al., 2011). From the perspectives of institutional complementarity and institutional substitution, Nayyar and Luiz (2024) further show that the joint effects of different institutional dimensions are subject to clear boundary conditions and interval-specific patterns. Thus, multidimensional institutional configurations and their stage-dependent changes provide greater explanatory value than the average association of a single institutional factor in isolation.

Finally, the predictive role of the institutional environment is strongly context-dependent. In the western region, *NII*×*CII* reaches 6.9302%, with several institutional indices and interactions showing relatively high importance. *RII* is more prominent among state-owned enterprises, whereas *CII* and *RII*×*CII* are more important among non-state-owned enterprises. Large enterprises show greater importance of *CII*, *NII*, and the three-way interaction, while small enterprises

depend more strongly on EIR_t. *RII*×*CII* is relatively prominent among agriculture-related enterprises, whereas *CII* reaches 7.6225% among forestry-related enterprises. This finding is consistent with the regional institutional literature, which holds that enterprise heterogeneity and local institutional conditions jointly shape enterprise behavior (Amendolagine et al., 2024). Further enterprise-level evidence indicates that the effects of institutional and governance factors on enterprise innovation vary with ownership structure, the competitive environment of the industry, and the degree of government intervention (Yu et al., 2022). The impact of environmental regulation on enterprises' green innovation likewise exhibits marked heterogeneity across enterprise types and regulatory environments (Qin et al., 2024). These studies further indicate that the predictive relationship between the institutional environment and innovation resilience varies substantially across enterprise and regional contexts. Accordingly, this study extends the institutional perspective from individual institutional factors to a framework of multidimensional institutional configurations, nonlinear prediction, and contextual heterogeneity. Importantly, SHAP values represent relative predictive contributions and conditional patterns within the specified model; the identified interval variations and interactions should not be interpreted as structural causal effects or statistically significant thresholds. Thus, this study identifies predictive patterns in complex institutional environments, while causal mechanisms require further rigorous validation.

8. Conclusions and policy recommendations

8.1 Conclusions

Drawing on institutional theory, this study uses multiple machine-learning models and SHAP to examine the relative predictive importance, nonlinear patterns, interactions, and group heterogeneity of multidimensional *IE* in predicting EIR_{t+1}, yielding four main conclusions.

First, EIR_t exhibits strong dynamic persistence. CatBoost achieves an out-of-sample fit of 43.6747%, while EIR_t has the highest SHAP importance (43.0710%) across all subsamples, confirming that current-period resilience is the strongest predictor of EIR_{t+1}.

Second, institutional dimensions contribute differently to prediction. *RG*, *INST*, and *EIC* are the leading features within the regulative, normative, and cognitive dimensions, respectively, while *CII* shows the greatest importance among the composite indices (5.14%), exceeding *RII* (2.83%) and *NII* (2.34%).

Third, *IE* display pronounced nonlinear and configurational predictive patterns for EIR_{t+1}. *RG* shows clear stage-dependent patterns, while *RII*×*CII* has the highest interaction importance (2.88%); together, the three institutional indices and their interaction terms contribute approximately 18.99%.

Fourth, the predictive role of *IE* for EIR_{t+1} is strongly context-dependent. Institutional indices and interactions vary markedly across regions, ownership types, enterprise sizes, and agriculture- versus forestry-related enterprises, reflecting differences in resource endowments, organizational attributes, and regional institutional environments.

8.2 Policy recommendations

Because this study identifies predictive associations between *IE* and *EIR* rather than structural causal effects, the following recommendations should be viewed as policy implications derived from these predictive patterns. They focus on four areas: innovation capacity building, multidimensional institutional coordination, policy instrument optimization, and differentiated governance.

First, strengthen enterprises' long-term capacity for innovation resilience. Policy support should shift from short-term project incentives toward sustained R&D, long-term capital, and industry–university–research collaboration. Enterprises should also strengthen rolling R&D plans, technology reserves, talent pipelines, and flexible innovation-resource allocation to enhance their capacity to sustain and restore innovation under external shocks.

Second, strengthen coordination among regulative, normative, and cognitive institutions. Rather than relying on a single regulatory or subsidy instrument, institutional arrangements should integrate environmental governance and policy incentives, capital-market and social oversight, and enterprise innovation culture and green cognition, thereby improving the conversion of external institutional support into sustained innovation capacity.

Third, establish a phased, portfolio-based, and dynamically adjustable policy system. Given the nonlinear and interactive patterns of *IE*, environmental governance, green subsidies, market supervision, and innovation support should be coordinated through pilot implementation, phased evaluation, and rolling adjustment. Cross-departmental coordination should also be strengthened where institutional synergies are evident.

Fourth, provide differentiated institutional support tailored to enterprise and regional characteristics. Regulatory oversight, financing, and innovation services should reflect local institutional conditions, while support measures should vary by ownership type, enterprise size, and sector. Agriculture- and forestry-related enterprises should also receive tailored policy packages consistent with their production cycles, resource dependence, and operational characteristics.

8.3 Limitations and future directions

Given the limitations of sample coverage, variable measurement, and methodological identification, the conclusions of this study remain open to further extension. Future research could expand the range of enterprise types and geographic coverage, integrate data from multiple sources, and employ methods with stronger causal identification to further examine the dynamic relationship between *IE* and *EIR*. It could also refine the dynamic measurement of *EIR* across the dimensions of innovation resistance, recovery, adaptation, and reconfiguration, thereby deepening our understanding of the mechanisms through which the institutional environment influences *EIR*.

References

Acemoglu, D., Aghion, P., Bursztyn, L., & Hemous, D. (2012). The environment and directed technical change. *American Economic Review, 102*(1), 131–166. https://doi.org/10.1257/aer.102.1.131

Aghion, P., Van Reenen, J., & Zingales, L. (2013). Innovation and institutional ownership. *American Economic Review, 103*(1), 277–304. https://doi.org/10.1257/aer.103.1.277

Amendolagine, V., Crescenzi, R., & Rabellotti, R. (2024). The geography of acquisitions and greenfield investments: Firm heterogeneity and regional institutional conditions. *Journal of Regional Science, 64*(4), 1476–1505. https://doi.org/10.1111/jors.12705

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics, 151*, 103745. https://doi.org/10.1016/j.jfineco.2023.103745

Bao, X., Sun, B., Han, M., Mai, Q., & Lin, H. (2024). Corporate integrity culture on environmental, social, and governance (ESG) performance. *Corporate Social Responsibility and Environmental Management, 31*(2), 1399–1417. https://doi.org/10.1002/csr.2637

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54(3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Berrone, P., Fosfuri, A., Gelabert, L., & Gomez-Mejia, L. R. (2013). Necessity as the mother of "green" inventions: Institutional pressures and environmental innovations. *Strategic Management Journal, 34*(8), 891–909. https://doi.org/10.1002/smj.2041

Bloom, N., Bond, S., & Van Reenen, J. (2007). Uncertainty and investment dynamics. *The Review of Economic Studies, 74*(2), 391–415. https://doi.org/10.1111/j.1467-937X.2007.00426.x

Bohn, S., & Gümüsay, A. A. (2024). Growing institutional complexity and field transition: Towards constellation complexity in the German energy field. *Journal of Management Studies, 61*(7), 3184–3225. https://doi.org/10.1111/joms.13004

Busenitz, L. W., Gómez, C., & Spencer, J. W. (2000). Country institutional profiles: Unlocking entrepreneurial phenomena. *Academy of Management Journal, 43*(5), 994–1003. https://doi.org/10.5465/1556423

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359–377. https://doi.org/10.1002/asi.20317

Chen, H., Covert, I. C., Lundberg, S. M., & Lee, S.-I. (2023). Algorithms to estimate Shapley value feature attributions.

Nature Machine Intelligence, 5(6), 590–601. https://doi.org/10.1038/s42256-023-00657-x

Chen, W. (2024). SMEs' strategic responses toward emerging technologies: An interpretive case study of the farming industry in the UAE and Taiwan. *Technological Forecasting and Social Change, 206*, 123572. https://doi.org/10.1016/j.techfore.2024.123572

Conz, E., & Magnani, G. (2020). A dynamic perspective on the resilience of firms: A systematic literature review and a framework for future research. *European Management Journal, 38*(3), 400–412. https://doi.org/10.1016/j.emj.2019.12.004

DesJardine, M., Bansal, P., & Yang, Y. (2019). Bouncing back: Building resilience through social and environmental practices in the context of the 2008 global financial crisis. *Journal of Management, 45*(4), 1434–1460. https://doi.org/10.1177/0149206317708854

Dong, Y., Sun, Z., & Zhou, Y. (2026). Not all voices are equal: Stakeholder environmental pressures and corporate green innovation strategies. *Journal of Business Research, 202*, 115793. https://doi.org/10.1016/j.jbusres.2025.115793

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Márquez, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Duchek, S. (2020). Organizational resilience: A capability-based conceptualization. *Business Research, 13*, 215–246. https://doi.org/10.1007/s40685-019-0085-7

Engelen, A., Huesker, C., Rieger, V., & Berg, V. (2024). Building a resilient organization through a pre-shock strategic emphasis on innovation. *Journal of Product Innovation Management, 41*(1), 36–61. https://doi.org/10.1111/jpim.12697

Fielke, S., Taylor, B., & Jakku, E. (2020). Digitalisation of agricultural knowledge and advice networks: A state-of-the-art review. *Agricultural Systems, 180*, 102763. https://doi.org/10.1016/j.agsy.2019.102763

Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E. R., & Lounsbury, M. (2011). Institutional complexity and organizational responses. *Academy of Management Annals, 5*(1), 317–371. https://doi.org/10.5465/19416520.2011.590299

Hillmann, J., & Guenther, E. (2021). Organizational resilience: A valuable construct for management research? *International Journal of Management Reviews, 23*(1), 7–44. https://doi.org/10.1111/ijmr.12239

Jakhar, S. K., Bhattacharya, A., Rathore, H., & Mangla, S. K. (2020). Stakeholder pressure for sustainability: Can 'innovative capabilities' explain the idiosyncratic response in the manufacturing firms? Business Strategy and the Environment, 29(6), 2635–2653. https://doi.org/10.1002/bse.2526

Kostova, T., Beugelsdijk, S., Scott, W. R., Kunst, V. E., Chua, C. H., & van Essen, M. (2020). The construct of institutional distance through the lens of different institutional perspectives: Review, analysis, and recommendations. Journal of International Business Studies, 51(4), 467–497. https://doi.org/10.1057/s41267-019-00294-w

Kwack, S. Y., & Kim, J. (2025). Social norm differences and innovation: Do institutional investors always spur corporate innovation? *Pacific-Basin Finance Journal, 90*, 102644. https://doi.org/10.1016/j.pacfin.2024.102644

Lee, M.-J., Choi, H., & Roh, T. (2024). Is institutional pressure the driver for green business model innovation of SMEs? Mediating and moderating roles of regional innovation intermediaries. Technological Forecasting and Social Change, 209, 123814. https://doi.org/10.1016/j.techfore.2024.123814

Li, W., Yu, J., & Chen, F. (2025). Linking firm performance with innovation culture: An algorithmic approach towards theory building. *Journal of Business Research, 187*, 115048. https://doi.org/10.1016/j.jbusres.2024.115048

Lian, G., Xu, A., & Zhu, Y. (2022). Substantive green innovation or symbolic green innovation? The impact of ER on enterprise green innovation based on the dual moderating effects. Journal of Innovation & Knowledge, 7(3), 100203. https://doi.org/10.1016/j.jik.2022.100203

Linnenluecke, M. K. (2017). Resilience in business and management research: A review of influential publications and a research agenda. *International Journal of Management Reviews, 19*(1), 4–30. https://doi.org/10.1111/ijmr.12076

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*, 56–67. https://doi.org/10.1038/s42256-019-0138-9

Malerba, F. (2002). Sectoral systems of innovation and production. *Research Policy, 31*(2), 247–264. https://doi.org/10.1016/S0048-7333(01)00139-1

Mishra, C. S. (2022). Does institutional ownership discourage investment in corporate R&D? *Technological Forecasting and Social Change, 182*, 121837. https://doi.org/10.1016/j.techfore.2022.121837

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Nayyar, R., & Luiz, J. M. (2024). Institutional complementarity and substitution in Indian multinational enterprises' cross-border investment decisions. *Management International Review, 64*(6), 1053–1089. https://doi.org/10.1007/s11575-024-00560-6

Nguyen, B., & Vu, N. (2024). Does intrinsic motivation or extrinsic pressure matter more? An exploratory study of small businesses going green and innovation. *Business Strategy and the Environment, 33*(5), 3855–3886. https://doi.org/10.1002/bse.3676

Ortiz-de-Mandojana, N., & Bansal, P. (2016). The long-term benefits of organizational resilience through sustainable business practices. *Strategic Management Journal, 37*(8), 1615–1631. https://doi.org/10.1002/smj.2410

Pache, A.-C., & Santos, F. (2010). When worlds collide: The internal dynamics of organizational responses to conflicting institutional demands. *Academy of Management Review, 35*(3), 455–476. https://doi.org/10.5465/amr.35.3.zok455

Peng, D., & Kong, Q. (2024). Corporate green innovation under environmental regulation: The role of ESG ratings and greenwashing. *Energy Economics, 140*, 107971. https://doi.org/10.1016/j.eneco.2024.107971

Peng, M. W., Sun, S. L., Pinkham, B., & Chen, H. (2009). The institution-based view as a third leg for a strategy tripod. *Academy of Management Perspectives, 23*(3), 63–81. https://doi.org/10.5465/amp.2009.43479264

Qin, K., Yao, C., & Niu, R. (2024). The impact of environmental regulation on enterprise green innovation. *Economic Analysis and Policy, 84*, 1933–1944. https://doi.org/10.1016/j.eap.2024.11.008

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal, 28*(13), 1319–1350. https://doi.org/10.1002/smj.640

Tu, B., Chen, Z., & Dang, J. (2025). The impact of foreign direct investment on innovation in China's forest products industry. *Forest Policy and Economics, 170*, 103371. https://doi.org/10.1016/j.forpol.2024.103371

Wang, C., Zhang, X.-E., Tao, Y., & Zhang, K. (2026). Artificial intelligence, dynamic capabilities, and innovation resilience: The contingent roles of market competition intensity and technological turbulence. *Journal of Innovation & Knowledge, 16*, 101004. https://doi.org/10.1016/j.jik.2026.101004

Williams, T. A., Gruber, D. A., Sutcliffe, K. M., Shepherd, D. A., & Zhao, E. Y. (2017). Organizational response to adversity: Fusing crisis management and resilience research streams. *Academy of Management Annals, 11*(2), 733–769. https://doi.org/10.5465/annals.2015.0134

Xie, X., Wu, Y., Xie, P., Yu, X., & Wang, H. (2021). Organizational innovation culture and firms' new product performance in two emerging markets: The moderating effects of institutional environments and organizational cohesion. *Journal of Management & Organization, 27*(5), 972–991. https://doi.org/10.1017/jmo.2019.8

Xu, J., & Sheng, Y. (2023). Regulations, politics, and firm green innovation. *Economic Analysis and Policy, 80*, 13–32. https://doi.org/10.1016/j.eap.2023.07.014

Ye, Q., Cai, Y., & Wang, J. (2024). Navigating institutional complexity: How firms respond to conflicting patenting demands from competing institutional logics. Technological Forecasting and Social Change, 204, 123431.

https://doi.org/10.1016/j.techfore.2024.123431

Yin, J., Chen, W., & Wang, F. (2026). A study of the impact of artificial intelligence on the low-altitude economy. *Humanities and Social Sciences Communications, 13*, 1163. https://doi.org/10.1057/s41599-026-07561-w

Yu, Z., Shen, Y., & Jiang, S. (2022). The effects of corporate governance uncertainty on state-owned enterprises' green innovation in China: Perspective from the participation of non-state-owned shareholders. *Energy Economics, 115*, 106402. https://doi.org/10.1016/j.eneco.2022.106402

Zhang, H., Zhang, X., Tan, H., & Tu, Y. (2024). Institutional investors' shareholding, corporate governance, and corporate innovation investment. International Review of Economics & Finance, 96, 103643. https://doi.org/10.1016/j.iref.2024.103643

Zhang, Y., Wang, L., & He, T. (2024). Can IP tribunals increase firm patent applications?—Empirical evidence from China. *China Economic Review, 87*, 102240. https://doi.org/10.1016/j.chieco.2024.102240

Zhao, Z., Zhao, Y., Lv, X., Li, X., Zheng, L., Fan, S., & Zuo, S. (2024). Environmental regulation and green innovation: Does state ownership matter? *Energy Economics, 136*, 107762. https://doi.org/10.1016/j.eneco.2024.107762

Data availability

The processed dataset supporting the findings of this study is provided as a separate Supplementary Data file accompanying this submission.

Author contributions

Author 1 contributed to the literature review, data collection, and manuscript preparation. Author 2 led the conceptualization and design of the study, developed the methodology, conducted the empirical analysis and machine-learning modeling, performed the SHAP-based interpretation, prepared the figures and tables, and drafted and substantially revised the manuscript. Author 3 contributed to the theoretical framework, interpretation of the results, and critical revision of the manuscript. Author 4 contributed to validation, interpretation of the findings, and manuscript review. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Ethics declaration

Not applicable. This study uses firm-level data on listed enterprises and does not involve human participants, personally identifiable human data, animals, or biological materials. Therefore, ethical approval and informed consent were not required.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx
- data.xlsx