Ain Shams Engineering Journal 16 (2025) 103454

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Full Length Article

Navigating risk horizons: A proactive framework for early warning in construction companies' outward foreign direct investments

Ruisi Jing [a], Pengcheng Xiang [a,*], Xinran Hu [b]

[a] Faculty of Construction Management and Real Estate, Chongqing Univ., No. 174, Shazheng St., Shapingba District, Chongqing 400044, PR China
[b] Bartlett School of Environment, Energy and Resources, Niv. College London, London WC1H 0NN, UK

ARTICLE INFO

Keywords:
Outward Foreign Direct Investment (OFDI)
Machine learning
Risk warning
Construction enterprises

ABSTRACT

Outward Foreign Direct Investment (OFDI) provides construction enterprises with opportunities for market expansion and resource generation. However, effective risk prevention and control are crucial for maximizing returns, presenting unique challenges in the construction industry. This study explores these complexities and underscores the importance of robust risk management. It introduces an innovative risk warning model that integrates cluster analysis, Bayesian optimization, and extreme gradient boosting (XGBoost) to establish a macro-medium-micro OFDI risk indicator system. Unsupervised K-means++ clustering enables three-dimensional risk level classification, while the BO-XGBoost machine learning model enhances accuracy, stability, and interpretability. Validation using a dataset of 154 samples from a decade of OFDI by Chinese listed construction companies demonstrates the model's superior performance compared to existing machine learning approaches. This research advances OFDI risk management by addressing the specific challenges faced by construction enterprises worldwide.

1. Introduction

Outward Foreign Direct Investment is a crucial factor for global enterprises to expand their markets and promote economic growth in various countries. According to the CEIC database, the global OFDI funds reached US$ 1,573.65 billion in 2022. In all industries, construction enterprises are crucial in China's "going out" strategy [1]. According to the Statistical Bulletin of OFDI, China's OFDI in the construction industry in 2021 amounts to US$ 55.07 billion, despite the impact of the pandemic. Furthermore, the value added of the construction industry in 2021 is US$ 1150.2 billion. On the one hand, architectural enterprises can leverage their cost advantages in human resources and specific technical capabilities to expand their market presence. This strategy is imperative considering the domestic market's sluggish growth and the challenges posed by the unsustainable traditional profit model of engineering contracting.

On the other hand, architectural enterprises in specialized industries can overcome the adverse effects of stringent domestic tax or environmental policies through OFDI. This allows them to mitigate profit losses and pursue resources or strategic assets overseas [2,3]. OFDI not only promotes the host country's economic development through capital accumulation, technology transfer, and improved employment [4,5], but also through the investing enterprise; investment promotes economic growth and innovation development in the home country in terms of industrial structure, technological spillover, and balance of payments [6,7].

Given that almost all investments involve risks [8,9], enterprises must assess these risks in relation to potential returns when making investment decisions.OFDI risk is the uncertainty that firms face when engaging in OFDI. These risks may originate from the host country's political regulations, economic policies, social culture, and market constraints [10,11]. Risk may also originate from the enterprise's own internal financial and non-financial problems [12,13]. Whether these risks can be identified early, assessed, and accurately predicted is the key to the investment success of an enterprise. To address this issue, existing studies focus on the construction of national risk frameworks and assessment systems [14,15] or the choice of enterprise OFDI location [16], which is not the only way to identify and assess these risks in the early stages. This limits the research methodology; the current research on OFDI risk and cross-border mergers and acquisitions is still dominated by theoretical and empirical analyses [17,18]. In fact, machine learning and deep learning algorithms are widely used in risk

* Corresponding author.
E-mail addresses: 202203021058t@cqu.edu.cn (R. Jing), pcxiang@cqu.edu.cn (P. Xiang), xinran.hu.23@ucl.ac.uk (X. Hu).

https://doi.org/10.1016/j.asej.2025.103454
Received 15 April 2024; Received in revised form 24 March 2025; Accepted 27 April 2025
Available online 20 May 2025

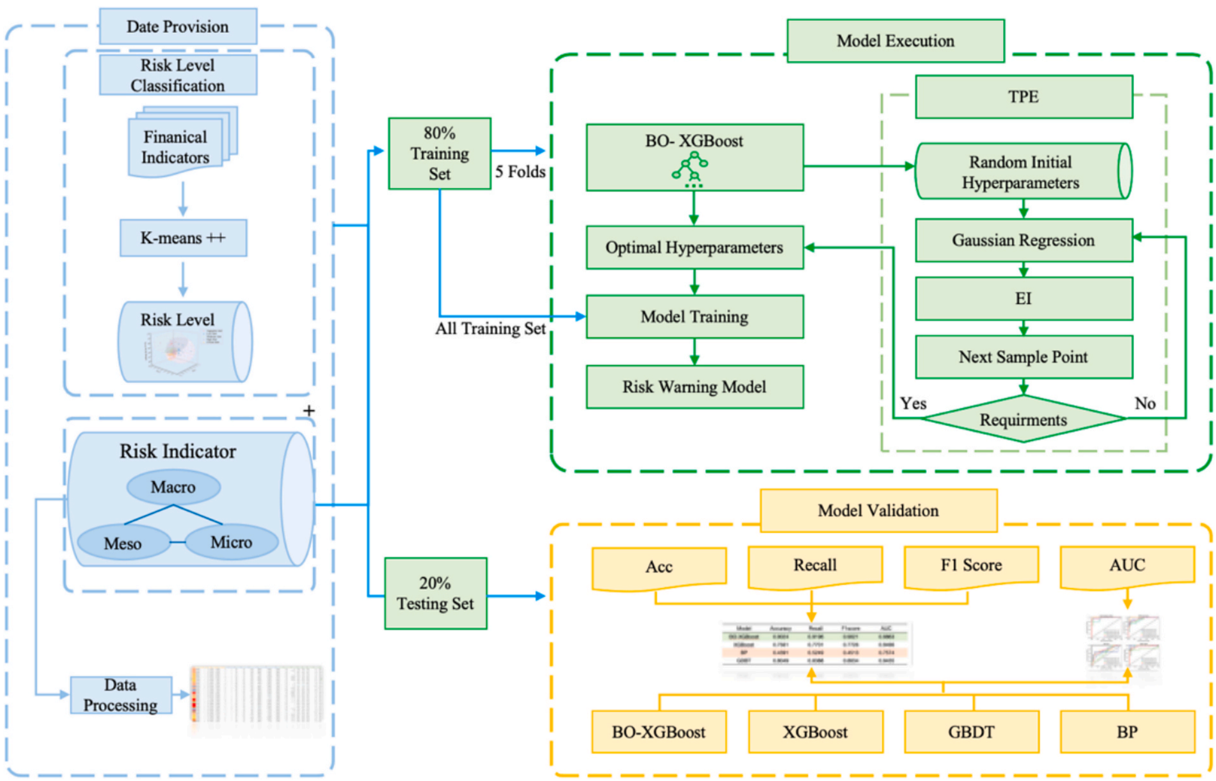

Fig. 1. Research Framework Diagram.

management research in various fields, such as default risk and online lending risk [19]. Regarding the main body of literature, the focus remains on investments in energy companies, financial institutions, and overall listed companies [20]; research on the risk of OFDI in construction enterprises remains limited. However, considering that enterprises pursuing different investment purposes usually face different risk causes and paths [21], relying on existing literature does not completely show the risks faced by OFDI in the construction industry.

Compared with general investment projects, the OFDI of construction enterprises has a larger investment scale, longer implementation cycle, more complex engineering technology, wider scope, and higher social concern. The purpose of this paper is to develop a complete framework for OFDI risk warning of construction enterprises, aiming at helping construction enterprises avoid investment risks, reduce unnecessary losses and improve investment success rate. The objective of this study is to develop a proactive risk warning framework for OFDI by construction enterprises. This framework is designed to help firms to anticipate, mitigate, and adapt to potential risks through early intervention. In this context, "proactive" refers to a systematic approach that emphasises the preemptive identification and dynamic management of risks before they escalate, rather than relying on reactive measures after a crisis [22]. This study's innovations are as follows: First, OFDI decision is the comprehensive result of the enterprise's full measurement of the external environment and the internal risk-resistant ability. We construct a macro-medium-micro multi-dimensional construction enterprise OFDI risk indicator system, which can help enterprises identify and avoid exogenous and endogenous risks. Second, using the unsupervised K-means++ algorithm in conjunction with multidimensional financial indicators, we classify risk levels in a three-dimensional space. However, compared to the average numerical risk level thresholds of traditional risk warning systems, the accuracy of the results requires further investigation. Third, we construct the BO-XGBoost machine learning model and validate its accuracy and stability, aiding enterprises identify the investment risk level. We utilize the model's interpretability to determine the significance of risk indicators and analyze the inherent risk mechanisms.

2. Literature review

OFDI generally refers to the economic activities of multinational enterprises from a home country seeking effective control in a host country by exporting capital, equipment, technology, and intangible assets such as management. Based on the entry modes of multinational enterprises into host countries, these investments can be classified into greenfield investment and cross-border mergers and acquisitions (M&A). The investment strategies of certain construction enterprises are notably intricate, and their categorization as such is contingent upon the specific investment policies of each country.

From the perspective of nations and governments, these investments serve as a significant source of revenue for the host country, reflecting the level of mutual investment dependence between the two countries [23]. For business operators, engaging in such foreign ventures enables companies, particularly those in emerging economies, to navigate potential regulatory constraints imposed by their home country policies during operations. Furthermore, it facilitates market expansion, empowering them to gain or offset the first-mover advantage of leading enterprises and enhance overall business performance [24,25]. Through a *meta*-analysis of 36 studies, Bausch and Krist [26] found that these foreign business activities have a significant and positive impact on corporate performance.

Numerous studies in the realm of OFDI focus on the analysis of risks faced by enterprises during the investment process. At the host country level, policy risk is regarded as one of the most critical risk types [27]. Butler and Joaquin [28] defined political risk, faced by investors, as the risk of the host country's government unexpectedly changing the rules governing business operations. Existing research indicates that the level and type of political risk in the host country, institutional distance between the home and host countries, and even leaders' risk preferences can influence the behaviours and outcomes of companies engaged in OFDI[29]. Investors are typically attracted to stable and robust

Table 1
An overview of the Indicator System.

Risk Dimension	Risk Type	Risk Indicator	Explanation	Data Source	Code
Macro Risks	Economic & Financial Risks	Economic Freedom	Measures the degree of economic freedom for individuals, enterprises, and markets in a country/region. Higher values indicate greater investment freedom.	EFI	H1
		Inflation Rate	GDP deflator, calculated as the ratio of nominal GDP to real GDP. Reflects price level changes in GDP. Higher values indicate higher inflation.	WDI	H2
		Investment Environment	Evaluate factors affecting investment risks, including contract feasibility, profit, repatriation, and payment delays. Higher values indicate better conditions.	ICRG	H3
		Tax Risk	Tax revenue as a percentage of GDP. Higher percentages indicate heavier tax burdens for enterprises and investors.	WDI	H4
	Social & Cultural Risks	Public Security Level	Reciprocal of the homicide rate per 100,000 people. Higher values indicate better public security.	WDI	H5
		Socioeconomic Development Level	Human Development Index (HDI). A composite metric for health, education, and living standards. Higher values indicate better development.	UNDP	H6
		Socio-cultural Distance	Measures cultural differences between countries/regions. Higher values indicate greater divergence in cultural norms.	Hofstede	H7
		Average Education Level	Reflects the population's educational attainment. Higher values indicate better-educated citizens.	WDI	H8
	Political & Legal Risks	Rule of Law	Assesses legal quality, contract enforcement, property rights, and judicial effectiveness. Scale: −0.25 to 2.5. Higher scores indicate stronger rule of law.	WGI	H9
		Corruption Control	Measures the extent to which public power is used for private gain—scale: −0.25 to 2.5. Higher scores indicate better governance.	WGI	*H*10
		Political Stability	Evaluates the government's ability to implement policies and maintain power. Scale: 0–12. Higher scores indicate greater stability.	ICRG	*H*11
Meso Risks	Market & Trade Risks	Environmental Regulation	Measured by the Environmental Performance Index (EPI), reflecting the stringency of environmental policies (air quality, water management, biodiversity, etc.).	EPI	H12
		Construction Industry Growth Prospect	Evaluated by the quarter-on-quarter GDP growth rate of the construction sector, indicating market demand and industry vitality.	UNSD	H13
	Human Resource Risks	Construction Labor Supply	Measured by the number of workers in the construction sector, reflecting labor availability and cost control for projects.	ILO	H14
		Construction Employment Attractiveness	Assesses the proportion of construction workers in the total labor market, indicating the sector's priority in the local economy.		H15
	Construction Resource Risks	Raw Material Cost Volatility (PPI)	Measures price fluctuations of construction raw materials, reflecting supply chain stability and cost management challenges.	IMF	H16
		Supplier Qualification	Evaluates the credibility, certifications, and capabilities of local material suppliers in the host country.	WDI	H17
		Raw Material Competitive Advantage	Assesses the host country's advantages in material quality, cost, and supply stability compared to other nations.	UN Comtrade	H18
Micro Risks	Corporate Capability & Governance Risks	Internal Control Capability Index	Measures the effectiveness of internal governance mechanisms, including process management, risk control, and compliance. Higher values indicate stronger control capabilities.	DBI	H19
		Executive International Background	Indicates whether senior executives have overseas education or work experience, reflecting their international expertise and market understanding.	CSMAR	H20
		Internationalization Level	Evaluates the enterprise's global engagement, including the proportion of overseas assets and revenue. Higher values indicate deeper international market involvement.		H21
		External Institutional Attention	Reflects the level of attention from external professional institutions (e.g., rating agencies, research firms). Higher values indicate greater recognition.		H22
	Financial & Operational Risks	Current Ratio	Measures the ability to cover short-term liabilities with short-term assets. Higher values indicate stronger short-term solvency.	CSMAR	H23
		Quick Ratio	Assesses liquidity by excluding inventory from current assets. Reflects stricter short-term solvency.		H24
		Inventory Turnover Ratio	Indicates how efficiently inventory is managed over a period. Higher values reflect better inventory utilization.		H25
		Debt-to-Asset Ratio	Measures the proportion of assets financed by debt. Higher values imply greater financial leverage and risk exposure.		H26
		Return on Equity (ROE)	Evaluates profitability by comparing net profit to shareholders' equity. Higher values indicate stronger profit generation.		H27
		Net Profit Growth Rate	Tracks the growth rate of net profit over a period. Higher values reflect faster improvement in profitability.		H28

investment conditions provided by the government. Due to factors such as global financial market integration and technological advancements, differences in geography, economics, and socio-cultural aspects can impact the performance of overseas subsidiaries [30]. Particularly for emerging economies, changes in economic risk directly affect the inflow of OFDI [31]. Taking China as an example, Kang and Jiang [32] suggest that both institutional and economic factors influence the choices of Chinese multinational corporations regarding OFDI. In addition, some scholars argue that factors such as climate risk, environmental regulation risk, and labour force can also impact a company's multinational investment [33–35]. However, in the aforementioned studies, risks at the industry and company have not been incorporated into the research framework of these OFDI. Paul and Feliciano-Cestero [36], through a review of nearly five decades of OFDI literature, assert that these foreign business activities have evolved into the most crucial field in international business. The next steps in research should integrate perspectives from various aspects such as national, industry, and company levels, using company-level data to assist business professionals in making more informed decisions in practice. Against this backdrop, this paper aims to construct a relatively comprehensive risk warning framework for foreign business activities in the construction industry. It seeks to introduce machine learning technology for intelligent decision-making, aiding the practical application of theories related to these business activities.

3. Materials and methodology

3.1. Research framework

First, we identify the endogenous"vulnerability" risk factors faced by construction enterprises during the investment process and the exogenous "threat" risk sources of investment target countries; we construct the risk assessment system of macro-, meso-, and micro-dimensions. Second, based on the K-means++ cluster analysis, we classify the risk grade of construction enterprises' OFDI in a three-dimensional space. Finally, we establish an OFDI risk-warning model based on BO-XGBoost. Combined with the interpretability of the tree model, this study analyzes the degree of importance of different types and source risk characteristics in the investment process of construction enterprises. This study helps enterprises or project teams identify, evaluate, and handle potential risks in different countries to ensure that the investment brings the expected benefits to enterprises and prevents investment losses or merger and acquisition failures. The research framework is illustrated in Fig. 1.

3.2. Construction of the indicator system

OFDI risk is the uncertainty that enterprises face when engaging in OFDI. This uncertainty manifests as the potential for losses during various stages of OFDI projects, including investment, construction, management, and operation. It directly impacts the profitability of businesses and exhibits characteristics such as objectivity, loss potential, systematic nature, and universality. At the macro level, assessing investment country risk is prominent in international economics. Existing studies categorize these risks into political, economic, legal, and social dimensions. Scholars analyze the risks associated with enterprises' OFDI from a macro perspective, focusing on national institutional risk [37]. The sources of risks associated with enterprises' OFDI are intricate and diverse, impacting their investment returns. In addition to risks originating from the host country, current research indicates that the risks associated with OFDI are also linked to the intrinsic financial and resource capabilities of the enterprise [38,39]. To conduct more precise analyses in the construction engineering sector, including civil engineering, housing construction, building installation, and the ACE domain, we utilize the popular OFDI dataset as a foundation. By manually selecting cases from the mentioned industries, this approach significantly reduces subtle industry variations compared to previous studies, thereby enhancing the accuracy of predictions and better serving construction engineering investors. This targeted approach enables the development of a more practical model for anticipating OFDI risks in construction. Our methodology integrates literature reviews, and expert interviews, contextualizing construction market dynamics. Following established practices [40,41], we construct a *meso*-level index system to identify macro-, meso-, and micro-level risks, forming the basis for a comprehensive risk evaluation system. Table 1 presents the indicator system.

Macro risks stem from factors tied to foreign countries or regions, often beyond investors' control. Meso-level risks focus on industry-specific and market-related factors, exhibiting greater self-organization. Compared with the other two levels of risk, *meso*-level risk has the characteristics of more stakeholders and complex operating mechanism. Micro-level risks involve internal factors within investing enterprises, such as financial indicators and internationalization extent, offering more intuitive manageability [42].

3.3. Methodology

3.3.1. K-means++

Following Wang et al. [43], we employ K-means++ to categorize the risk levels associated with investment samples. The K-means clustering algorithm, proposed by MacQueen in 1957, suits the cluster analysis of samples where the number of categories is known and small and the samples are unlabeled, such as identifying and classifying types of stochastic processes with different labels. The core concept is to divide n individuals in a set of multivariate data into k clusters, with individual in the dataset fully assigned to a specific cluster [44]. However, the K-means algorithm randomly selects the initial centres, and the good or bad clustering results are characterized by randomness. Specifically, K-means++ algorithm can sequentially select the point with the largest distance as the initial center, ensuring that the initial center does not fall into the high-density data area simultaneously, thereby allowing differences in the risk level of different samples to be effectively differentiated. The specific steps of the K-means++ algorithm are as follows:

Step 1. Normalize the data by randomly selecting a centroid from the data set.

Step 2. Calculate the Euclidean distance $D(x)$ between each sample and the existing clustering center.

$$D(x) = \sqrt{\sum_{l=1}^{m} (x_l - \mu_{il})^2}$$

x is a sample point in a cluster; μ_i is the clustering center of the cluster; m is the number of clustering features; l is each feature composing point x; x_l is the *lth* feature of the sample point x, and μ_{il} is the *lth feature of the clustering* center μ_i of the cluster.

Step 3. According to the roulette method, select the next sample point as the new centroid with probability P. Enter the loop until K cluster centroids are selected.

$$P = \frac{D(x)^2}{\sum_{x\in X} D(x)^2}$$

where x is the sample data set.

Step 4. Calculate the distance from each sample in the dataset to the K clustering centers and assign that sample to the category corresponding to the clustering center with the smallest distance.

Step 5. For each category, recalculate the center of mass, that is, the new cluster center, for all samples belonging to that category.

$$\mu_i = \frac{1}{|C_i|}\sum_{x\in C_i} x$$

where μ_i is the clustering center of C_i, C_i is the ith cluster set, and $|C_i|$ is the number of samples in the C_i cluster set.

Step 6. Repeat the above two steps until the center of the cluster no longer changes, and the algorithm converges to obtain K clusters.

K is an important parameter in the clustering analysis that can directly affect the accuracy of the clustering results. To better warn enterprises of investment risk, we select the optimal value of cluster division K by combining SSE and manual preclassification as the evaluation index of the clustering algorithm [45].

$$SSE = \sum_{i=1}^{K} \sum_{x\in C_i} |x_i - \mu_i|$$

3.3.2. EXtreme gradient boosting

The extreme gradient boosting (XGBoost) algorithm solves classification and regression problems through integration and effectively implements gradient-enhanced decision trees [46]. Compared to the traditional gradient-enhanced decision tree, the XGBoost algorithm improves the loss function through Taylor expansion, and its core idea is to determine the optimal tree structure. Specifically, the XGBoost algorithm utilizes the greedy algorithm to learn each base tree and learns new functions by continuously forming new decision trees, thus fitting the residuals of previous predictions. It continuously reduces the residual error between the predicted value and the true value, hence improving the prediction accuracy [47], and the final predicted value is the weighting of the calculated value of each leaf node. The XGBoost is selected for its strong performance, interpretability, and ability to record feature importance. enhances model generalization, mitigates overfitting, and operates efficiently with small data volumes.

Step 1. Construct a dataset with n samples and m features; x_i and y_i are the input variables and corresponding variables, respectively, and $\hat{y}_i$ represents the predicted values of a model with K trees.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F$$

F is the set space of the regression trees and $f(x)$ is the regression tree. Notably,$f_k(x_i)$ can be denoted as $^{\omega}q(x_i), q \in \{1,2,...,T\}$, where T denotes the number of leaf nodes, q denotes the decision rule of the tree, and ω denotes the sample weight (i.e., leaf score) of the leaf nodes.

Step 2. Define the XGBoost objective function as:

$$\mathscr{L}(t) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{K} \Omega(f_k)$$

The objective function consists of two parts: the true prediction deviation and regularization term. l is the loss function, which indicates the difference between the predicted and true values; $\Omega(f_k)$ indicates the model complexity, which is the regularization term to prevent overfitting and is given by $\Omega(f_k) = yT + \frac{1}{2}\sum_{j=1}^{T} w_j^2$. The formula for λ denotes the penalty factor, which is a fixed coefficient; γ denotes the complexity of each leaf, and ω is the vector of scores on the leaves.

As the new tree must fit the last prediction deviation, XGBoost is constrained at step t in the form of a summation.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$$

The objective function can be expressed as:

$$\mathscr{L}(t) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \sum_{k=1}^{K} \Omega(f_k)$$

The XGBoost algorithm utilizes this at $f_t = 0$, which can be obtained using Taylor's second-order expansion and simplification:.

$$\mathscr{L}(t) \approx \sum_{i=1}^{n} l\left[y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i g_i f_t^2(x_i)\right] + \sum_{k=1}^{K} \Omega(f_k)$$

where g_i is the first order derivative and h_i is the second order derivative.

$$g_i = \partial_{\hat{y}(t-1)} l\left(y_i, \hat{y}_i^{(t-1)}\right), h_i = \partial^2_{\hat{y}(t-1)} l\left(y_i, \hat{y}_i^{(t-1)}\right)$$

By combining the above equations, the objective function is rewritten as a one-quadratic function of the fraction of leaf nodes to compute the optimal leaf weights for each leaf node w_j^* and the extremes of $L(t)$ and $L^*(t)$.

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

$$\mathscr{L}^*(t) = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda}$$

In practice, directly enumerating all possible tree structures for computation using the above formula is difficult. XGBoost starts from the root node. It first proposes a partition threshold based on the percentile of the feature distribution and then iteratively matches the feature thresholds. For each node in the tree, XGBoost tries to add a split, and its computational process is expressed as

$$\mathscr{L}_{split} = \frac{1}{2}\left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} + \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda}\right] - \gamma$$

where $I = I_L \cup I_R$; I_L and I_R are the split left- and right-hand sample sets, respectively. L_{split} consists of left branch score, right branch score, no-split score, and complexity; its maximum value is considered as the optimal split on that node.

3.3.3. Bayesian optimization algorithm

The XGBoost algorithm has numerous hyperparameters that significantly affect prediction accuracy. Optimizing these parameters often depends on experience, trial and error, and subjective judgment. Efficient optimization methods are necessary to determine the optimal hyperparameter values for the model. Hyperparameter selection is akin to finding the optimal solution, represented by a performance function with optimal hyperparameter values as independent variables [48]. The tree-structured Parzen estimator (TPE) in this study is an iterative probability distribution-based approach known as Sequential Model-Based Global Optimization, which employs a categorical approach to agent modeling. The acquisition function, such as Thompson sampling, entropy search, and expected improvement (EI), is a heuristic that efficiently explores, weighs, and exploits new and known regions in the objective space to obtain a balance between the optimal hyperparameters. EI is chosen as a criterion because it performs well and is intuitive for incorporating the degree of improvement of the function into the judging criteria [49].

$$EL_{y^*}(c) = \int_{-\infty}^{+\infty} \max(y^* - y, 0) p_M(y|c) dy$$

EI Under model M, $f(c)$ exceeds (negatively) the expectation of some threshold y^* such that the hyperparameter c that maximizes its value is the locally optimal hyperparameter, the algorithm stores $f(c)$ and c from each iteration into the search history, fitting a new model M^t until it returns a globally optimal solution, the exact solution process is uneconomical, and to address this problem, TPE provides simple approximate solutions to determine the model M and find locally optimal superparameter settings [50]. To approximate model M, TPE is indirectly modeled with $p_M(y|c)$ and $p_M(y)$.

$$p_M(c|y) = \begin{cases} l(c) & \text{if} \quad y < y^* \\ q(c) & \text{if} \quad y \geq y^* \end{cases}$$

where $l(c)$ and $q(c)$ are the density estimates formed from the observations and the remaining observations, respectively, and typically y^* is set by the algorithm to be the γ quantile of the observation y, where the default value is $\gamma = 0.15$ [51]. Combining the above formulas, the optimization of EL in the TPE algorithm is as follows:

$$EL_{y^*}(c) = \frac{\gamma y^* l(c) - l(c) \int_{-\infty}^{y^*} P_M(y) \mathrm{d}y}{\gamma l(c) + (1-\gamma) q(c)} \propto \left(\gamma + \frac{q(c)}{l(c)}(1-\gamma)\right)^{-1}$$

4. Experimental design

4.1. Sample selection and sources

China's construction industry investment methods are complex and diverse, usually involving large capital; however, there are problems such as missing data and incomplete data disclosure for some micro-enterprises. Listed construction enterprises have relatively better and standardized information in terms of year statistics and investment projects, which enhances comparability among sample cases.

Table2
Feature Values with High Missing Rates.

Feature Name	Code	Missing Rate
Tax Risk	H4	0.3471
Average Education Level	H8	0.3021
Supplier Qualification	H17	0.5367

Simultaneously, they account for a great share of China's OFDI, which can partially represent the overall Chinese construction enterprises.

Through web crawler technology combined with manual collation, we synthesize multiple databases and official websites to select OFDI cases of Chinese construction enterprises from 2010 to 2021; the selected companies are primarily involved in civil engineering, housing construction, building installation, and other industries. The original sample data is systematically processed in accordance with the following criteria: (1) Samples are meticulously curated to exclusively encompass instances wherein the investing entities are listed companies; (2) Samples associated with unsuccessful or terminated M&A transactions, as well as those that have not materialized, are systematically excluded; (3) Instances featuring special treatment indicators such as ST (special treatment), *ST, etc., are systematically excluded from the sample set; (4) Recognizing the typically substantial scale of listed construction companies, transactions with amounts less than US$ 50 million and projects with an equity transfer ratio below 30 % are methodically excluded, given their perceived limited measurable impact on the companies; (5) In scenarios where a single enterprise engages in investments across multiple host countries within the same fiscal year, this study meticulously selects the most representative investment amount corresponding to the largest investment transaction. We obtained 154 valid samples, comprising 43 listed enterprises and investments across 54 countries and regions, from various databases such as the China Global Investment Tracker, Wind, ORBIS, Thomson Reuters, and FDI markets. We obtain enterprise-level financial data from the CSMAR and annual reports of listed companies.

4.2. Data preprocessing

Data preprocessing is a crucial step in the machine learning process, involving the cleaning, transformation, and organization of raw data to ensure its quality and format meet the requirements for model training and prediction. As shown in Table 2, this study removes features with a missing rate exceeding 30 % from the original dataset and appropriately imputes missing warning indicators to mitigate the impact of data loss on the model, thereby enhancing data quality and model reliability. This study adopts the Min-Max scaling method for standardization. Normalization scales data proportionally within the range of [0,1], eliminating the impact of magnitude differences among different indicators. This ensures that all indicators have equal weight when calculating distances and determining risk levels.

4.3. Risk level assessment

The main characteristic of risk is that the potential economic losses it encompasses are ultimately visible in a firm's financial performance. A high-risk level implies a relatively higher probability of investment failure and thus may lead to a more significant difference between expected and actual returns [52]. Therefore, a firm's risk level could be judged by assessing its financial performance after an investment [53]. Financial performance is a valuable indicator of investment returns and profitability, facilitating a nuanced assessment of the risk profile associated with an investment [54]. This study selects three financial indicators that can represent the performance of enterprises for cluster analysis: return on total assets (ROA, which measures the ability of all the assets of the enterprise to obtain profits), return on capital (ROIC, which integrates the consideration of equity and debt and measures the efficiency of the investment), and the gross margin of overseas profits [55,56]. Unlike in other industries, investment in the construction industry usually has a long construction and operation payback period. Therefore, the financial indicator data is set at a two-year window. This study divides the degree of risk alert into five categories by K-means++ clustering method combined with SSE. This is because the three financial indicators are positive indicators; thus, the size of their risk level is determined according to the Euclidean distance from the center point of the clusters to the origin d_{μ}, and the five levels of alerts are named, from small to large, Negligible Alert, Low Alert, Moderate Alert, High Alert, and Critical Alert. Finally, we test the significance of the classification to determine whether the classification method is effective.

4.4. Machine learning model experiments

The XGBoost algorithm can improve the warning accuracy and robustness while shortening the running time of the model. The development of a reliable machine learning model requires separate training and testing phases, typically conducted on distinct datasets. Given this study's small-scale, high-dimensional, and imbalanced dataset, a nested validation approach is employed to improve generalization. The dataset is initially divided, with 80 % allocated for training and 20 % for testing. The training set is then randomly partitioned using 5-fold cross-validation, where one subset serves as the validation set while the remaining four act as the training set. This process is repeated five times to generate five models, and the optimal hyperparameter combination is selected based on their average performance. Finally, a new model is trained on the entire training set to produce the final model. In terms of hyperparameter optimization and selection, we take the accuracy of the XGBoost model as the goal of model parameter optimization and adopt TPE for parameter optimization. We identify four crucial parameters for optimization in model training and effectiveness. The learning rate governs the model's learning pace, with values too large causing convergence issues and too small values slowing down loss gradient descent. The min_child_weight parameter impacts the sum of sample weights in the smallest leaf node, preventing learning outliers when set larger but risking compromised performance when set too small. Conversely, max_depth sets the tree's maximum depth, with larger values enhancing model specificity to prevent overfitting. The n_estimators parameter, representing the number of base learners, is adjustable within the range defined by these parameters for optimal configuration. To enhance the reliability of this study, the optimized BO-XGBoost model is compared with the unoptimized XGBoost, GBDT, and backpropagation neural network (BPNN). XGBoost and GBDT represent tree-based ensemble learning methods, while BPNN represents a neural network-based approach, allowing for a comprehensive comparison between these two dominant modeling paradigms. This selection ensures that the evaluation encompasses both structured, rule-based learning and representation-based learning, highlighting the advantages of XGBoost over a wide range of alternative methods.

4.5. Performance indicators

This study uses the confusion matrix to calculate the accuracy and F1 score to evaluate the prediction results of the model and make comparisons between the models [57].

$$recall = \frac{TP}{TP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

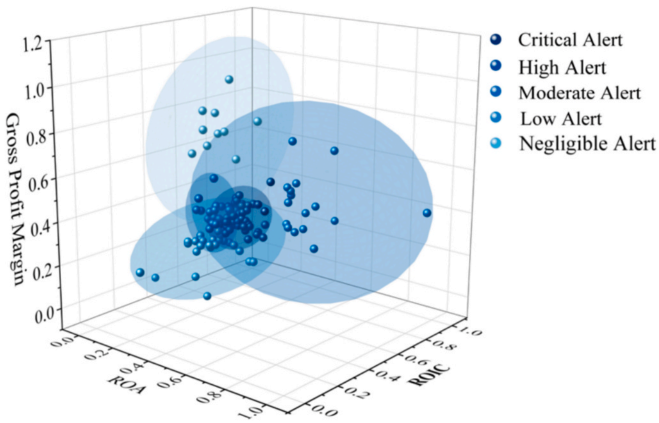

Fig. 2. Results of clustering.

Table 3
Risk Level Classification Results.

Risk level	Negligible Alert	Low Alert	Moderate Alert	High Alert	Critical Alert
Number	16	20	32	59	27
d_{μ}	1.081	1.004	0.807	0.712	0.053

$$F1Score = 2 \times \frac{precision \times recall}{precision + recall}$$

True positive (TP) is the number of samples in which a positive outcome is predicted correctly; false positive (FP) is the number of samples in which a negative outcome is predicted correctly; false negative (FN) is the number of samples in which a negative outcome is predicted incorrectly, and true negative (TN) is the number of samples in which a negative outcome is predicted incorrectly.

The proposed model in this study is a multiclassification problem, and the dataset is characterized by an imbalance and small size. Therefore, we introduce the area under the Receiver Operating Characteristic (ROC) curve (AUC) to measure the model performance. The AUC can comprehensively and effectively assess the actual prediction performance of a model for unbalanced classifications and small-sample data [58].

5. Results and discussion

5.1. Risk classification results

In this study, we cluster and analyze the ROA, ROIC, and gross profit margin of overseas profits of the selected enterprises after two years of OFDI. We divide risk level according to the Euclidean distance from the center of the clusters to the origin in the three-dimensional space. After standardizing the selected three metrics, as they all represent positive indicators, the farther the distance from the cluster centroid to the origin, the better the financial performance of the enterprise post-investment. This implies a lower likelihood of investment loss and, consequently, a lower risk level.

Fig. 2 visually depicts the distribution of analysis results in three-dimensional space. Table 2 provides a summary of the distances from the origin to the clustering centers for different categories, along with the number of investment samples corresponding to different risk levels, and Table 3 presents the results. The p-values of all three indicators are zero, indicating that this classification is effective.

By assigning risk levels from 1 to 5 in ascending order based on the clustering results, the average risk level of Chinese construction enterprises' OFDI projects in host countries over the years has been calculated, as shown in Fig. 3. The spatial distribution of OFDI risks exhibits a distinct east–west differentiation. In Asia, where project volume is the highest, risk distribution appears relatively balanced, with a lower proportion of low-risk investments. The region's geographical proximity and cultural similarities make it the preferred destination for overseas expansion. In Africa, risk levels display a polarised pattern, with significantly higher risks in the western region compared to the East, reflecting challenges such as political instability, weak economic foundations, and complex security conditions. In Europe, high-risk projects

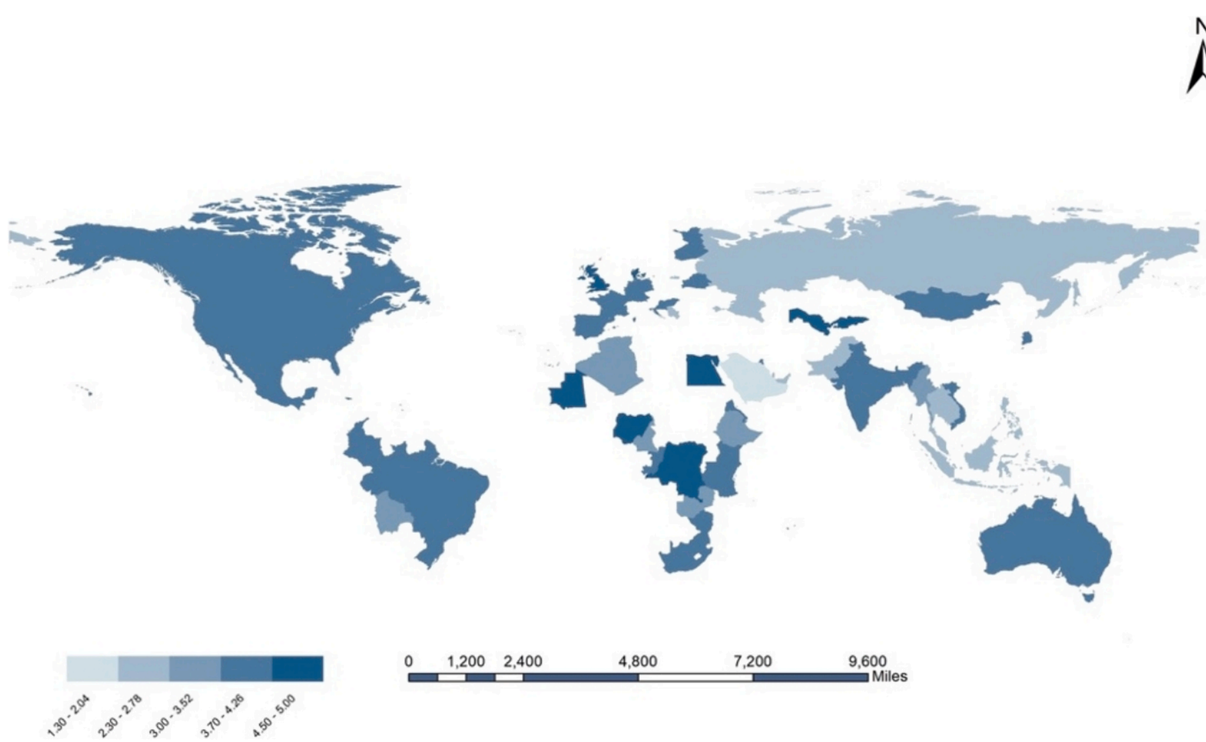

Fig. 3. Distribution of risk.

Table 4
Test of significance.

Metric	ROA	ROIC	Gross Profit Margin
F	48.341	39.041	63.294
P-value(p)	0.000	0.001	0.000

Table 5
Optimization of parameters.

Parameter	Low limit	Upper limit	Optimization
min_child_weight	1	10	3
max_depth	1	20	6
learning_rate	0.01	0.5	0.03
n_estimators	100	600	127

dominate, which can be attributed to the region's stringent technical standards, complex regulatory framework, and high market entry barriers. In the Americas and Oceania, most projects fall into the high-risk or relatively high-risk categories, with few low-risk cases, largely due to high labor costs and strict environmental regulations.

5.2. Bayesian parameter optimization results

This study uses TPE to optimize the important hyperparameters and parameters of the XGBoost model. We select four parameters that significantly influence model training and model effects as the parameters to be optimized. Table 4 shows the adjustment ranges of the above parameters and the optimal parameter results.

5.3. Xgboost model alert performance

OFDI early risk warning is a typical multidimensional, multilayered, and multicoupled system model, and machine learning can address the unclear relationship between variables and expected outputs. To verify the effectiveness of the proposed model for construction companies, this study compares it with the unoptimized model and two other advanced machine learning models: BP and GDBT. To ensure a fair comparison, we train and test all models using the same training and test sets. Table 5 shows the recall, accuracy, F1 score, and AUC of these models, while Fig. 4 and Fig. 5 show the ROC curves and confusion matrix.

Among the four models, BO-XGBoost performs the best on this task with a high accuracy, recall, F1 score and AUC value (Table 6). This suggests that for the OFDI early risk warning study with small samples and imbalanced categorization, the tree model may perform better compared to the neural network. This is because the classification tree is a discriminative model; the mapping function corresponding to the classification tree is a segmented linear division of the multi-dimensional space, which naturally supports multi-classification problems. The tree model is a supervised nonparametric model with nonlinear expressive ability, which has higher robustness in small sample problems. For the neural network model, a higher feature dimension affects the design of the neural network, and smaller number of samples leads to incomplete network training or overfitting, both of which limit its warning accuracy. Under the default hyperparameters, the optimized model significantly improves its prediction performance compared to the XGBoost model, and the optimized model performance is significantly improved by almost 15 % for accuracy, recall, and F1 score when compared to the pre-optimized model performance.

5.4. Xgboost model warning explained

One major advantage of tree modeling is its interpretability, which disambiguates the machine learning process. For early-warning risk problems, interpretable machine learning models can facilitate decision-makers in identifying potential causes of risks.

The risks associated with Chinese construction enterprises' OFDI are most significantly influenced by micro-level factors, followed by meso-level risks, while macro-level factors have the least impact (Fig. 6). This aligns with the practical experiences of many construction firms expanding overseas [54]. Although macro-level policies and economic conditions are important, the success of investment projects often hinges on firms' execution capabilities, management efficiency, and internal resource allocation. Moreover, this finding partially confirms that

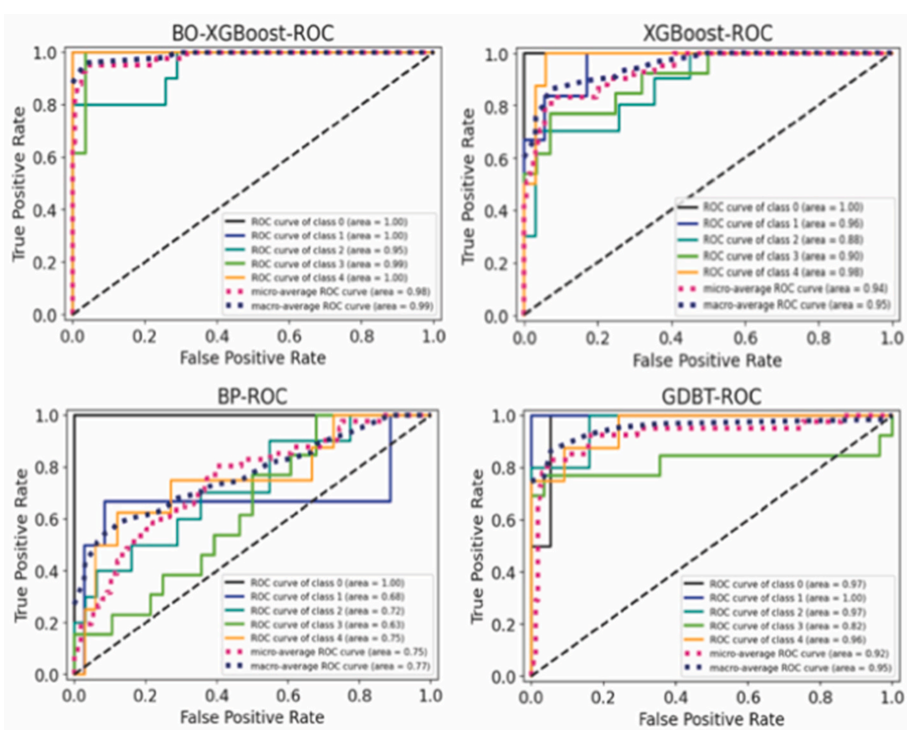

Fig. 4. ROC curve comparison of four machine learning models.

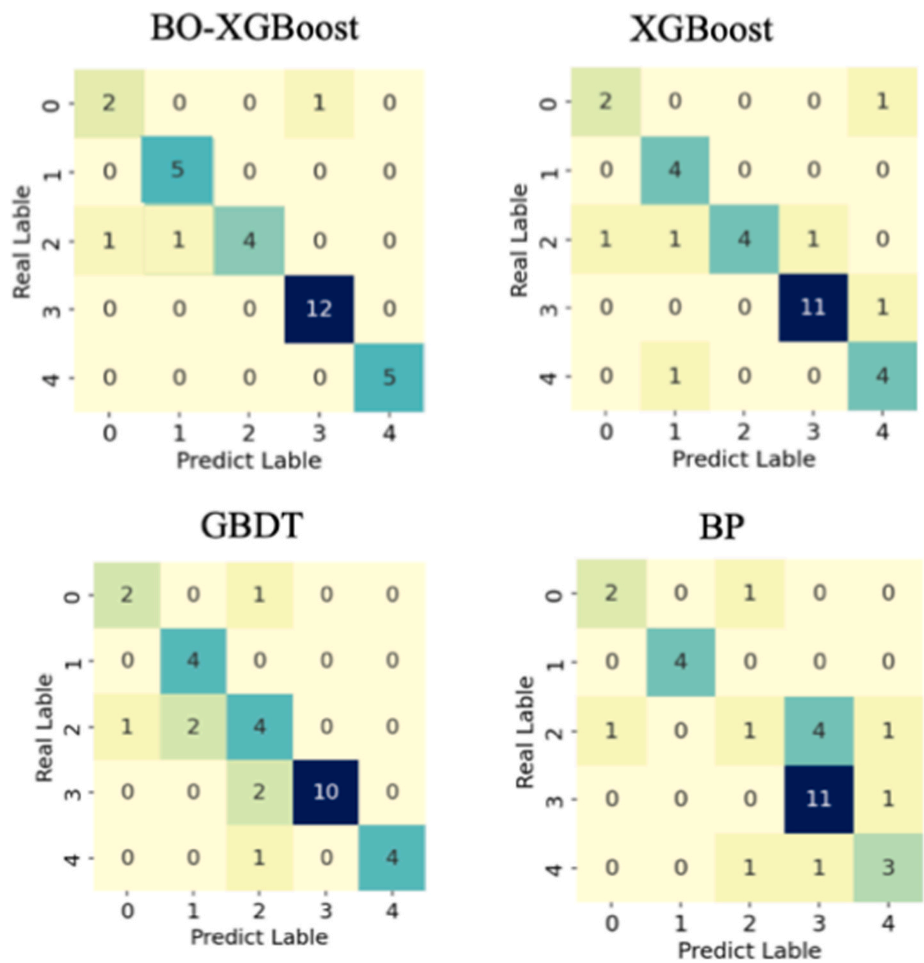

Fig. 5. Confusion matrix.

Table 6
Performance of the models.

Model	Accuracy	Recall	F1score	AUC
BO-XGBoost	0.9032	0.9034	0.9143	0.9863
XGBoost	0.8387	0.8065	0.8041	0.9637
BP	0.6317	0.6774	0.6369	0.7521
GDBT	0.8118	0.7742	0.7838	0.9235

policies such as the Belt and Road Initiative (BRI) have effectively provided support and protection for Chinese construction enterprises, reducing adverse impacts from host countries' institutional and economic environments.

Among all risk indicators, the asset-liability ratio (H26) stands out as the most influential micro-level factor. This ratio directly reflects a firm's financial health and operational efficiency, which play a crucial role in project execution, risk management, and profitability. It indicates the proportion of debt financing affecting financial stability, debt repayment capacity, and resilience to risk. For Chinese construction enterprises engaging in large-scale investments—particularly in infrastructure, transportation, and energy—the asset-liability ratio determines their ability to sustain financial burdens and respond to external risks such as policy changes or market downturns. At the macro risk level, the importance of risk factors in Chinese construction enterprises' OFDI is relatively low, with public security (H5) and political stability (H11) emerging as the most critical indicators. At the meso risk level, the significance of risk factors varies considerably. Market and trade risks are the most influential, followed by human resource risks, while construction resource risks have a comparatively minor impact.

5.5. Discussion

The construction sector constitutes a key driver of national economic development [59], necessitating the continued pursuit of optimal methodologies for mitigating investment risks within the industry. This study aligns with the theoretical foundations of OFDI risk assessment frameworks in prior literature; however, further contextual dialogue with domain-specific theories is needed. For instance, while institutional risks are often considered dominant predictors in resource-seeking OFDI [60], our findings highlight the heightened volatility of micro risks in the construction sector. This may reflect industry-specific dynamics, as construction projects involve intensive local stakeholder engagement [61]. Unlike dynamic early-warning frameworks that integrate big data analytics with geopolitical event tracking, our methodology specifically targets the pre-investment decision-making phase. By leveraging existing datasets to predict potential investment returns, this approach establishes a proactive decision-support framework with anticipatory capabilities [62].

This study is primarily based on OFDI data from Chinese listed construction enterprises. Although the dataset covers investment cases across various countries and regions, its relatively small sample size may limit the generalizability of the findings. Regarding model selection, this study employs BO-XGBoost for risk prediction but does not extensively compare it with deep learning or other advanced methods, which may influence the evaluation of different algorithmic applications. Future research could expand the dataset to include a broader range of

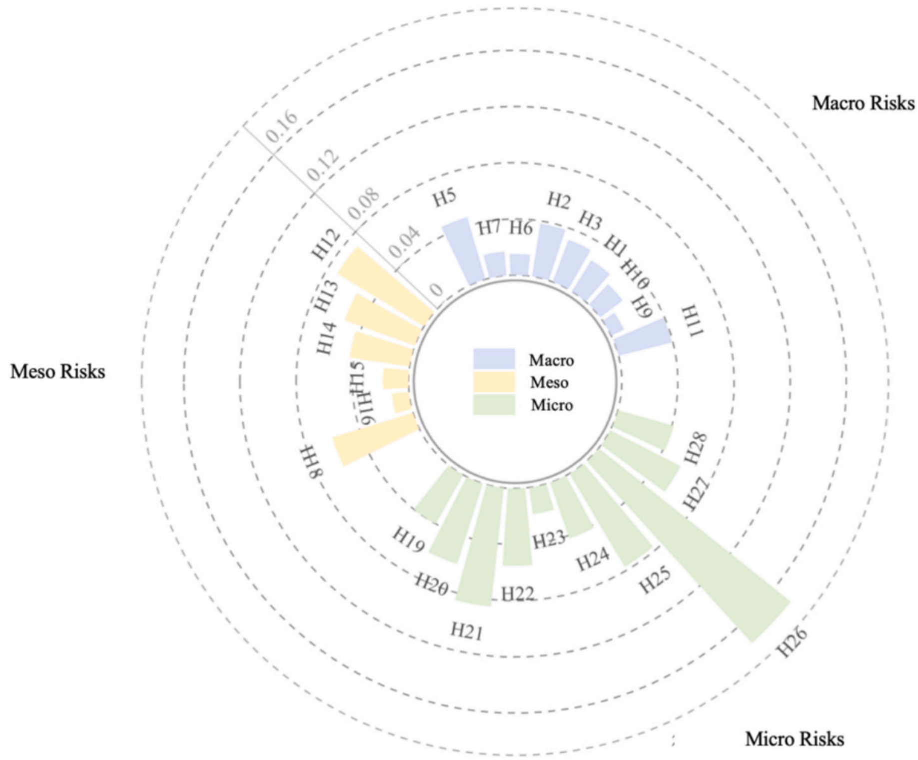

Fig. 6. Feature Importance Score.

industries and regions to enhance the model's applicability and generalizability. Additionally, incorporating deep learning, hybrid models, or causal inference approaches could further improve the predictive accuracy and interpretability of the risk assessment framework.

6. Conclusions

This study introduced a machine learning method aimed at early warning of the risk level associated with OFDI in construction enterprises. The approach involved constructing a comprehensive risk early warning indicator system at the national, market, and enterprise levels, integrating various databases to form a robust dataset. The experimentation phase employed the clustering method to categorize risk levels, incorporating XGBoost as an enhancement technique. Model hyperparameters were optimized using TPE. The key conclusions are summarized as follows:

(1) The study conducted risk factor identification from diverse perspectives, resulting in the establishment of a multidimensional risk early warning system. This system encompassed indicators across macro, meso, and micro dimensions, drawing data from 12 global databases to ensure the scientificity and objectivity of data sources.
(2) Using K-means++ clustering and Sum of Squared Errors (SSE), the study classified risk levels, identifying five optimal categories. Findings indicate high OFDI risk for Chinese construction firms, with Asia as the primary investment region.
(3) The study utilized Bayesian optimization with TPE to expedite the identification of optimal hyperparameters for the XGBoost algorithm. This approach significantly enhanced the model's efficiency and accuracy. The resulting BO-XGBoost model demonstrated superior performance in risk early warning compared to alternative models, offering precise and robust support for the ex-ante management and early warning of OFDI risk in construction enterprises.

7. Data availability statement

Some or all data, models, or codes that support the study's findings are available upon reasonable request.

Author contributions:

Pengcheng Xiang: Data preservation, manuscript review, and communication coordination.

Ruisi Jing: Active contributions to the writing of the paper and program development.

Xinran Hu: Responsible for manuscript writing and analysis of visualizations.

CRediT authorship contribution statement

Ruisi Jing: Writing – original draft, Validation, Software, Resources, Methodology, Formal analysis. **Pengcheng Xiang:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Xinran Hu:** Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was also supported by the Chongqing Design Group Scientific Research Project (2023-C2).

References

[1] Duan F, Ji Q, Liu BY, Fan Y. Energy investment risk assessment for nations along China's Belt & Road Initiative. J Clean Prod 2018;170:535–47. https://doi.org/10.1016/j.jclepro.2017.09.152.

[2] Dunning JH, Lundan SM. Institutions and the OLI paradigm of the multinational enterprise. Asia Pac J Manag 2008;25(4):573–93. https://doi.org/10.1007/s10490-007-9074-z.
[3] Stoian C, Mohr A. Outward foreign direct investment from emerging economies: escaping home country regulative voids. Int Bus Rev 2016;25(5):1124–35. https://doi.org/10.1016/j.ibusrev.2016.02.004.
[4] Borensztein E, De Gregorio J, Lee JW. How does foreign direct investment affect economic growth. J Int Econ 1998;45(1):115–35. https://doi.org/10.1016/S0022-1996(97)00033-0.
[5] Lee CC, Chang CP. FDI, financial development, and economic growth: international evidence. J Appl Econ 2009;12(2):249–71. https://doi.org/10.1016/S1514-0326%2809%2960015-5.
[6] Helpman E. Trade, FDI, and the organization of firms. J Econ Lit 2006;44(3):589–630. https://doi.org/10.1257/jel.44.3.589.
[7] Ciesielska D, Koltuniak M. Outward foreign direct investments and home country's economic growth. Phys A 2017;482:127–46. https://doi.org/10.1016/j.physa.2017.04.057.
[8] Kellard NM, Kontonikas A, Lamla MJ, Maiani S, Wood G. Risk, financial stability and FDI. J Int Money Finance 2022;120:102232. https://doi.org/10.1016/j.jimonfin.2020.102232.
[9] Ti LL, Ng BK, Rasiah R. Motivations behind SME greenfield investment in emerging markets. Manage Decis 2023;61(6):1656–76. https://doi.org/10.1108/Md-04-2022-0482.
[10] Globerman S, Shapiro DM. The impact of government policies on foreign direct investment: the Canadian experience. J Int Bus Stud 1999;30(3):513–32. https://doi.org/10.1057/palgrave.jibs.8490081.
[11] Rienda L, Claver-Cortes E, Quer D, Andreu R. Greenfield investments or acquisitions? The influence of distance on emerging-market multinationals. Manage Decis 2018;57(5):1223–36. https://doi.org/10.1108/MD-02-2017-0154.
[12] Chen YS. Classifying credit ratings for Asian banks using integrating feature selection and the CPDA-based rough sets approach. Knowl-Based Syst 2012;26:259–70. https://doi.org/10.1016/j.knosys.2011.08.021.
[13] Alexandridis G, Antypas N, Gulnur A, Visvikis I. Corporate financial leverage and M&As choices: evidence from the shipping industry. Transp Res E Logist Transp Rev 2020;133:101828. https://doi.org/10.1016/j.tre.2019.101828.
[14] Dimitrova A, Triki D. Does state fragility matter for foreign direct investment? Evidence from Southern and Eastern Mediterranean countries. Manage Decis 2018;56(8):1787–803. https://doi.org/10.1108/MD-03-2017-0252.
[15] Wiessner YT, Giuliani E, Wijen F, Doh J. Towards a more comprehensive assessment of FDI's societal impact. J Int Bus Stud 2024;55(1):50–70. https://doi.org/10.1057/s41267-023-00636-9.
[16] Helpman E, Melitz MJ, Yeaple SR. Export versus FDI with heterogeneous firms. Am Econ Rev 2004;94(1):300–16. https://doi.org/10.1257/000282804322970814.
[17] Arslan A, Tarba SY, Larimo J. FDI entry strategies and the impacts of economic freedom distance: Evidence from Nordic FDIs in transitional periphery of CIS and SEE. Int Bus Rev 2015;24(6):997–1008. https://doi.org/10.1016/j.ibusrev.2015.03.004.
[18] Kanas A, Molyneux P. Do measures of systemic risk predict US corporate bond default rates. Int Rev Financ Anal 2020;71:101553. https://doi.org/10.1016/j.irfa.2020.101553.
[19] Zhang L, Hu S, Xu G. Pandemic uncertainty and firm investments. Financ Res Lett 2023;55:103924. https://doi.org/10.1016/j.frl.2023.103924.
[20] Kocaarslan B, Soytas U. The role of major markets in predicting the US municipal green bond market performance: New evidence from machine learning models. Technol Forecast Soc Change 2023;196:122820. https://doi.org/10.1016/j.techfore.2023.122820.
[21] Makino S, Beamish PW, Zhao NB. The characteristics and performance of Japanese FDI in less developed and developed countries. J World Bus 2004;39(4):377–92. https://doi.org/10.1016/j.jwb.2004.08.009.
[22] Elshater A, Abusaada H. Proactive insights into place management: spatiotemporal effects of street food activities in public spaces. J Place Manage Dev 2024;17(4):442–63. https://doi.org/10.1108/JPMD-10-2023-0103.
[23] Wang Y, Lee HW, Tang W, Whittington J, Qiang M. Structural equation modeling for the determinants of international infrastructure investment: evidence from Chinese contractors. J Manage Eng 2021;37(4):4021033. https://doi.org/10.1061/(Asce)Me.1943-5479.0000933.
[24] Witt MA, Lewin AY. Outward foreign direct investment as escape response to home country institutional constraints. J Int Bus Stud 2007;38(4):579–94. https://doi.org/10.1057/palgrave.jibs.8400285.
[25] Chang SJ, Rhee JH. Rapid FDI expansion and firm performance. J Int Bus Stud 2011;42:979–94. https://doi.org/10.1057/jibs.2011.30.
[26] Bausch A, Krist M. The effect of context-related moderators on the internationalization-performance relationship: evidence from meta-analysis. Manag Int Rev 2007;47:319–47. https://doi.org/10.1007/s11575-007-0019-z.
[27] Bussy A, Zheng H. Responses of FDI to geopolitical risks: The role of governance, information, and technology. Int Bus Rev 2023;32(4):102136. https://doi.org/10.1016/j.ibusrev.2023.102136.
[28] Butler KC, Joaquin DC. A note on political risk and the required return on foreign direct investment. J Int Bus Stud 1998;29:599–607. https://doi.org/10.1057/palgrave.jibs.8490009.
[29] Beazer QH, Blake DJ. The conditional nature of political risk: how home institutions influence the location of foreign direct investment. Am J Polit Sci 2018;62(2):470–85. https://doi.org/10.1111/ajps.12344.
[30] Andrews DS, Meyer KE. How much does host country matter, really? J World Bus 2023;58(2):101413. https://doi.org/10.1016/j.jwb.2022.101413.
[31] Gokmenoglu K, Kirikkaleli D, Eren BM. Time and frequency domain causality testing: the causal linkage between FDI and economic risk for the case of Turkey. J Int Trade Econ Dev 2019;28(6):649–67. https://doi.org/10.1080/09638199.2018.1561745.
[32] Kang Y, Jiang F. FDI location choice of Chinese multinationals in East and Southeast Asia: traditional economic factors and institutional perspective. J World Bus 2012;47(1):45–53. https://doi.org/10.1016/j.jwb.2010.10.019.
[33] Wheeler D, Mody A. International investment location decisions: the case of US firms. J Int Econ 1992;33(1–2):57–76. https://doi.org/10.1016/0022-1996(92)90050-T.
[34] Goerzen A, Sapp S, Delios A. Investor response to environmental risk in foreign direct investment. Manag Int Rev 2010;50(6):683–708. https://doi.org/10.1007/s11575-010-0058-8.
[35] Kottaridi C, Louloudi K, Karkalakos S. Human capital, skills and competencies: varying effects on inward FDI in the EU context. Int Bus Rev 2019;28(2):375–90. https://doi.org/10.1016/j.ibusrev.2018.10.008.
[36] Paul J, Feliciano-Cestero MM. Five decades of research on foreign direct investment by MNEs: an overview and research agenda. J Bus Res 2021;124:800–12. https://doi.org/10.1016/j.jbusres.2020.04.017.
[37] Deghetto K, Holmes Jr RM, Lamont BT, Lucianetti L, Kasapoglu MM. New ventures going global: cognition and context. J Small Bus Manage 2023;61(6):2732–61. https://doi.org/10.1080/00472778.2021.1934853.
[38] Rajabi R, Yakhchali SH. Optimizing cash flow in construction portfolios: a metaheuristic approach from the organization's perspective. Ain Shams Eng J 2025;16(2):103259. https://doi.org/10.1016/j.asej.2024.103259.
[39] Qabaja M, Tenekeci G. Influence of inflation on the construction sector and economic growth in selected countries: a continental comparison. Ain Shams Eng J 2024;15(11):103013. https://doi.org/10.1016/j.asej.2024.103013.
[40] Azadegan A, Dooley K. A typology of supply network resilience strategies: complex collaborations in a complex world. J Supply Chain Manage 2021;57(1):17–26. https://doi.org/10.1111/jscm.12256.
[41] Laud G, Chou CY, Leo WWC. Service system well-being: scale development and validation. J Serv Manage 2023;34(3):368–402. https://doi.org/10.1108/JOSM-06-2021-0224.
[42] Stefano G, Denicol J, Broyd T, Davies A. What are the strategies to manage megaproject supply chains? A systematic literature review and research agenda. Int J Project Manage 2023;41(3):102457. https://doi.org/10.1016/j.ijproman.2023.102457.
[43] Wang D, Tong X, Wang Y. An early risk warning system for outward foreign direct investment in mineral resource-based enterprises using multi-classifiers fusion. Resour Policy 2020;66:101593. https://doi.org/10.1016/j.resourpol.2020.101593.
[44] Goicovich I, Olivares P, Román C, Vazquez A, Poupon C, Mangin JF, et al. Fiber clustering acceleration with a modified kmeans++ algorithm using data parallelism. Front Neuroinform 2021;15:727859. https://doi.org/10.3389/fninf.2021.727859.
[45] Yan T, Shen SL, Zhou A, Chen X. Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm. J Rock Mech Geotech Eng 2022;14(4):1292–303. https://doi.org/10.1016/j.jrmge.2022.03.002.
[46] Parsa AB, Movahedi A, Taghipour H, Derrible S, Mohammadian AK. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accident Anal Prev 2020;136:105405. https://doi.org/10.1016/j.aap.2019.105405.
[47] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining San Francisco, California, USA, 2016, pp. 785–794.
[48] Yang E, Yang Q, Li J, Zhang H, Di H, Qiu Y. Establishment of icing prediction model of asphalt pavement based on support vector regression algorithm and Bayesian optimization. Constr Build Mate 2022;351:128955. https://doi.org/10.1016/j.conbuildmat.2022.128955.
[49] Ghanbari-Adivi F, Mosleh M. Text emotion detection in social networks using a novel ensemble classifier based on Parzen Tree Estimator (TPE). Neural Comput Appl 2019;31(12):8971–83. https://doi.org/10.1007/s00521-019-04230-9.
[50] C. Thornton, F. Hutter, H.H. Hoos, K. Leyton-Brown, Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, Illinois, USA, 2013, pp. 847–855.
[51] Tao S, Peng P, Li Y, Sun H, Li Q, Wang H. Supervised contrastive representation learning with tree-structured parzen estimator Bayesian optimization for imbalanced tabular data. Expert Syst Appl 2024;237:121294. https://doi.org/10.1016/j.eswa.2023.121294.
[52] Wang D, Tong X, Wang Y. An early risk warning system for Outward Foreign Direct Investment in Mineral Resource-based enterprises using multi-classifiers fusion[J]. Resour Policy 2020;66. https://doi.org/10.1016/j.resourpol.2020.101593.
[53] Laeven L, Levine R. Bank governance, regulation and risk taking. J Financ Econ 2009;93(2):259–75. https://doi.org/10.1016/j.jfineco.2008.09.003.
[54] Musarat MA, Maqsoom A, Naeem MH, Ullah F, Salman A, Alaloul WS, et al. Evaluating the correlation between project selection criteria and organizational performance within the construction industry. Ain Shams Eng J 2024;15(7):102794. https://doi.org/10.1016/j.asej.2024.102794.
[55] Rocha K, Camacho F, Bragança G. Return on capital of Brazilian electricity distributors: a comparative analysis. Energy Policy 2007;35(4):2526–37. https://doi.org/10.1016/j.enpol.2006.09.012.

[56] Seo K, Soh J. Asset-light business model: an examination of investment-cash flow sensitivities and return on invested capital. Int J Hosp Manag 2019;78:169–78. https://doi.org/10.1016/j.ijhm.2018.12.003.

[57] Stehman SV. Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ 1997;62(1):77–89. https://doi.org/10.1016/S0034-4257(97)00083-7.

[58] Gao W, Jin R, Zhu S, Zhou ZH. One-pass AUC optimization. Artif Intell 2016;236:1–29. https://doi.org/10.1016/j.artint.2016.03.003.

[59] Althoey F, Houda M, Waqar A, Nisar S. Comprehensive framework for Integrated Construction Platforms (ICP) implementation and its effect on project success. Ain Shams Eng J 2024;15(12):103152. https://doi.org/10.1016/j.asej.2024.103152.

[60] I. Kolstad, A. Wiig, What determines Chinese outward FDI? J. World Bus. 47 (1) (2012) 26–34. Doi: 10.1016/j.jwb.2010.10.017.Y. Qiu, Labor adjustment costs and risk management, J. Financ. Quant. Anal. 54(3) (2019) 1447–1468. Doi: 10.1017/S0022109018000960.

[61] Manickavasagam J, Visalakshmi S, Apergis N. A novel hybrid approach to forecast crude oil futures using intraday data. Technol Forecast Soc Change 2020;158:120126. https://doi.org/10.1016/j.techfore.2020.120126.

[62] Arikan I, Shenkar O. Neglected elements: What we should cover more of in international business research. J Int Bus Stud 2022;53:1484–507. https://doi.org/10.1057/s41267-021-00472-9.

PengchengXiang . Vice President, Professor, Doctoral Supervisor, School of Management Science and Real Estate, Chongqing University; Secretary General of Chongqing Management Science and Engineering Professional Steering Committee, director of Construction Economy Branch of Architectural Society of China, Executive director of the second Council of Chongqing Law Society Real Estate Law Construction Law Research Society, expert of Chongqing Social Science expert Database