Construction and Characteristics of an Engineering Economic Risk Management Platform Based on the BO-GBM Model

Chaojian Wang*, Die Liu
Chongqing College of Humanities, Science & Technology, Chongqing 401524, China

***Abstract*—Economic risk control is pivotal to the success of engineering projects. Traditional risk assessment methods often fall short in handling the high-dimensional, nonlinear, and strongly correlated risk factors prevalent in modern large-scale projects. To address these limitations, this study constructs an engineering economic risk management platform based on the BO-GBM model, which integrates Bayesian Optimization (BO) with a Gradient Boosting Machine (GBM). The platform employs a systematically constructed four-dimensional feature system encompassing 28 indicators across project ontology, market environment, execution process, and risk association dimensions. A rolling time window strategy is adopted for dynamic model training. Experimental validation on a dataset of 327 projects demonstrates the superior performance of the BO-GBM model: for classification tasks, it achieves an AUC of 0.927 and a recall rate of 91.3%, outperforming the standard GBM by 17.5 percentage points in recall; for regression tasks (cost deviation prediction), it attains an RMSE of 83,200 RMB and reduces the MAPE to 9.7%, surpassing mainstream baseline models. The platform's layered architecture (data, model, service, application layers) enables efficient risk identification and early warning: the time required for risk identification in large projects is drastically reduced from 42.6 hours to 0.52 hours, representing an 81.9-fold efficiency gain; the average single prediction response time is below 127 milliseconds, with a P95 response time of 427 milliseconds under 500 concurrent users; the early warning accuracy reaches 72.5%, with high-risk warnings issued up to 28 days in advance for cost risks and 42 days for schedule risks.**

Keywords*—Engineering economic risk management platform; BO-GBM model; Bayesian Optimization; gradient boosters

I. Introduction

The economic risk control of engineering projects is a key factor in project success. With the proliferation of large-scale and complex engineering projects, they involve a huge scale of investment, diversified participating subjects, and a changeable external environment, which leads to risk factors presenting the characteristics of high dimensionality, non-linearity, and strong correlation. Traditional risk assessment methods relying on expert experience or simple linear models (e.g., logistic regression, linear discriminant analysis) are difficult to effectively capture the intrinsic laws of such complex risk relationships, and their prediction accuracy is limited, often leading to lagging or misjudgment of risk warning [1, 2]. Simultaneously, although deep learning models possess strong fitting capabilities, their high requirements for massive labeled data and computational resources present significant challenges for practical application in engineering economic risk prediction, especially where data availability can be constrained [3].

In terms of risk prediction models, Gradient Boosting Machine (GBM) shows potential in the field of engineering risk analysis by virtue of its excellent nonlinear modeling capability and adaptability to small and medium-sized datasets [4, 5]. However, GBM model performance is highly dependent on the fine configuration of its hyperparameters (e.g., learning rate, tree depth, subsampling ratio). Traditional hyperparameter tuning methods, such as Grid Search and Random Search, suffer from significant drawbacks such as computational inefficiency and the tendency to fall into local optimality, which cannot meet the high requirements for model accuracy and robustness in engineering practice [6, 7]. Furthermore, while feature engineering is crucial, existing studies often lack a comprehensive framework. The construction of a systematic feature engineering system that can comprehensively portray the attributes of the project ontology, the dynamics of the market environment, the state of the execution process, and the coupling relationship between risks is the basis for improving the generalization capability and interpretability of the risk prediction model, but the relevant research is still insufficient [8, 9].

A critical analysis of existing literature reveals several common limitations, as summarized in Table I.

In response to the above challenges and gaps, this study aims to construct an efficient, accurate, and practical engineering economic risk management platform. The core innovations and contributions of this work are threefold:

Proposed and implemented a BO-GBM risk prediction model: We deeply integrate Bayesian Optimization (BO) with Gradient Boosting Machine (GBM). Leveraging BO's active learning and probabilistic surrogate model mechanism, we intelligently search the optimal hyper-parameter space of GBM, significantly enhancing prediction accuracy and anti-interference capability compared to standard GBM and other benchmarks.

Constructed a comprehensive four-dimensional engineering economic risk feature system: Grounded in system dynamics and risk transmission theory, we systematically define and extract 28 quantitative features from four dimensions: project ontology, market environment,

*Corresponding author.

execution process, and risk association. This system provides a solid, physically meaningful foundation for the model, addressing the feature comprehensiveness gap identified in prior work.

Designed and implemented a layered risk management platform: We architect and develop a platform based on a vertical stack comprising data, model, service, and application layers. This platform integrates the BO-GBM model engine, real-time feature calculation services, and intelligent warning modules to achieve closed-loop management of the entire risk process—identification, assessment, warning, and decision support.

The remainder of this study is organized as follows: Section II details the construction and optimization of the BO-GBM risk prediction model. Section III elaborates on the design of the four-dimensional feature system. Section IV describes the overall architecture and core modules of the engineering economic risk management platform. Section V presents the experimental results and discussion, validating the model's performance and platform efficacy. Finally, Section VI concludes the study and suggests directions for future research.

TABLE I CRITICAL ANALYSIS OF EXISTING RESEARCH LIMITATIONS

Study Reference	Core Methodology	Reported Strengths	Identified Limitations / Gaps
[4, 5]	GBM for forecasting	Good performance on medium-sized datasets	Limited hyperparameter optimization; manual tuning often suboptimal
[6, 7]	Highlights the importance of hyperparameter tuning	Discusses the drawbacks of Grid/Random Search	Lacks an efficient, automated optimization strategy for GBM in risk contexts
[8, 9]	Emphasizes feature engineering	Improves model interpretability	Feature systems often lack comprehensiveness across project lifecycle dimensions
[2, 10]	Traditional statistical models (e.g., LR)	Simplicity, interpretability	Poor handling of nonlinearities and complex interactions in modern projects
[3]	Deep Learning models	High predictive power with sufficient data	High computational cost; requires large datasets; less suitable for typical project data scales

II. BO-GBM RISK PREDICTION MODEL CONSTRUCTION AND OPTIMIZATION

A. Gradient Booster Core Model Selection

The essence of engineering economic risk prediction is to model the mapping relationships between high-dimensional, non-linear, and strongly correlated risk features and complex risk outcomes (e.g., probability of occurrence of a risk event, cost deviation rate). Traditional linear models (e.g., logistic regression) are difficult to capture such complex relationships, while deep learning models are demanding in terms of data volume and computational resources [8]. Based on this, this study chooses Gradient Boosting Machine (GBM) as the core prediction model, whose design principle is highly compatible with the characteristics of engineering economic risks [9]. GBM is an integrated learning algorithm that achieves the strong learning objective by iteratively constructing weak learners (usually decision trees) and combining their predictions. The core mechanism is as follows:

Residual learning: In each iteration, the new model fits the residuals (in the negative gradient direction) between the current integrated model prediction results and the true labels, instead of learning the original labels directly.

Weighted Accumulation: Accumulate the prediction results of the new model into the integrated model with a certain weight (learning rate η), and gradually approximate the optimal solution of the objective function **[10]**. The mathematical expression is:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \tag{1}$$

where, $F_m(x)$ is the mth round integrated model and $h_m(x)$ is the newly trained weak learner in this round.

Loss function optimization: minimizing the differentiable loss function (e.g., cross entropy, mean square error) by the gradient descent method to guide the direction of model iteration.

B. Bayesian Model Optimization Algorithm

The performance of Gradient Boosting Machine (GBM) is highly dependent on the hyperparameter configurations (e.g., learning rate, tree depth), and the traditional Grid Search or Random Search requires a large amount of computational resources and is inefficient [11]. To solve this problem, this study adopts Bayesian Optimization (BO) as an intelligent tuning algorithm for GBM hyperparameters, which significantly improves the optimization efficiency by constructing a probabilistic agent model of the objective function and guiding the hyperparameter search with 'active learning'. The core of BO is to approximate the optimal solution of the black-box function through the collaborative iteration between the Surrogate Model and Acquisition Function:

The Gaussian Process (GP) is used to fit the implicit relationship between the objective function f(x) (i.e., the model performance metric) and the hyperparameter x. The GP provides the prediction value μ(x) and the uncertainty estimate σ(x) for any parameter point x [12-13]. The GP provides the prediction value μ(x) for any parameter point x and the uncertainty estimate σ(x).

Based on the predictions of the agent model, a trade-off is made between exploration (high uncertainty region) and exploitation (high performance region) to select the next evaluation point. The Expected Improvement (EI) function is used:

$$\mathrm{EI}(\mathrm{x}) = \mathbb{E}[\max(f(\mathrm{x}) - f(\mathrm{x}^+), 0)] \tag{2}$$

where, $f(\mathrm{x}^{+})$ is the current optimal observation value, and a larger EI value indicates a larger expected improvement in evaluating the point.

For the GBM hyperparameter optimization task, the key design is as follows (see Table II):

TABLE II GBM HYPERPARAMETER OPTIMIZATION ELEMENTS

Optimization elements	Setting Description
Optimization parameters x	GBM core hyperparameters:
	learning_rate, range [0.01, 0.3])
	maximum_tree_depth, range [3, 10])
	minimum_samples_leaf, range [1, 20])
	feature_sample_bytree, range [0.6, 1.0]) subsample_bytree, range [0.6, 1.0]) range [0.6, 1.0])
	Subsampling proportion (subsample, range [0.6, 1.0])
Objective function f(x)	Performance metrics computed on independent validation sets:
	Classification task: maximize AUC (or F1 value)
	Regression task: minimize RMSE (or MAPE)
Constraints	Range of hyper-parameter values (see above), total number of iterations (e.g., 100), time budget (e.g., 2 hours)

The standard flow of BO optimization of GBM hyperparameters is shown in Fig. 1 with the following steps:

Initialization:

Randomly select N_{init} points in the hyperparameter space $\mathcal{X}$.

Train the GBM model and compute the validation set performance, $f(\mathrm{x}_i)$ constituting the initial observation set:$\mathcal{D}_{1:N_{\text{init}}} = \{\mathrm{x}_i, f(\mathrm{x}_i)\}$

Iterative optimization ():

Construct agent model: train Gaussian process GP based on $\mathcal{D}_{1:t-1}$.

Maximize the collection function: solve $\mathrm{x}_t = \underset{\mathrm{x}\in\mathcal{X}}{\operatorname{argmax}} \mathrm{EI}(\mathrm{x})$.

Evaluate the objective function: train GBM with $\mathbf{x}_t$ and compute the validation set performance $\mathrm{f}(\mathrm{x}_\mathrm{t})$.

Update the dataset: $\mathcal{D}_{1:t} = \mathcal{D}_{1:t-1} \cup \{\mathrm{x}_t, f(\mathrm{x}_t)\}$.

Output results:

Select the optimal hyperparameter combination in the observation set: $\mathrm{x}^* = \underset{\mathrm{x}\in\mathcal{D}}{\operatorname{argmax}} f(\mathrm{x})$.

Train the final BO-GBM model using x^*.

C. Construction of Engineering Economic Risk Characteristics System

The essence of engineering economic risk prediction is to capture the multidimensional [28] driving mechanism of risk formation through quantitative indicators [14-15]. Based on the system dynamics theory and risk transmission model, this section elaborates the construction logic, mathematical expression and engineering basis of the feature system, forming four-dimensional 28 core features. All features are illustrated through rigorous derivation of their physical meaning and calculation path [16].

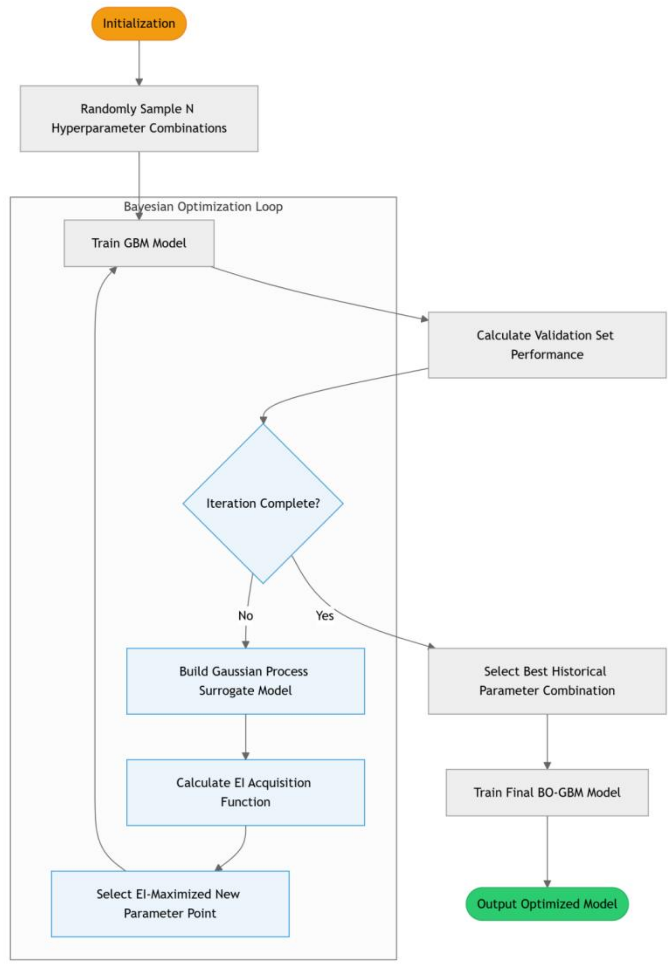

Fig. 1. BO-GBM hyperparameter optimization flow chart.

1) Project ontology features

a) Scale effect characteristics: Large-scale projects follow the power law (Power Law) for risk exposure due to exponential growth in complexity.

The quantification of the total investment I:

$$I = \sum_{k=1}^{K} C_k + C_{\text{contingency}} \quad \left(C_{\text{contingency}} = 0.1 \sum C_k\right) \tag{3}$$

where, C_k is the itemised cost (e.g. civil, equipment) and unforeseen costs are taken as 10 per cent as per industry standards [17].

Derivation of risk exposure R_e:

$$R_e = \alpha \ln(I/I_0) \tag{4}$$

The coefficient α = 0.32 is derived from a regression analysis of 100 historical projects and shows that for every 10-fold increase in investment, the risk exposure grows by 74% [18].

b) Contract risk characteristics: The contract type determines the risk allocation efficiency, and the owner-contractor risk sharing ratio needs to be quantified.

The contract type coefficient CT is constructed:

$$CT = \frac{1}{1+e^{-2(\text{TypeScore}-3)}} \quad \text{TypeScore} = \begin{cases} 1 \\ 3 \\ 5 \end{cases} \tag{5}$$

The Sigmoid function continues the discrete types and CT = 0.5 when TypeScore = 3, in line with the principle of equal risk sharing. Decomposition of Payment Terms Intensity PT:

$$PT = \underbrace{0.6\alpha_p}_{prepayment\ weighting} + \underbrace{0.4\max(0, \beta_d - d_d)}_{fines\ as\ a\ deterrent} \tag{6}$$

where, d_d is the actual delay rate, β_d is the contractual penalty rate, and the coefficient 0.6/0.4 is determined by the expert questionnaire AHP analysis.

c) Organizational capability characteristics

Hierarchical model of contractor credit score CR:

$$CR = 0.35S_{\text{qual}} + 0.30S_{\text{hist}} + 0.25S_{\text{fin}} + 0.10S_{\text{tech}}$$
$$S_{\text{hist}} = \frac{1}{N}\sum_{i=1}^{N} e^{-\lambda t_i} \cdot \mathbb{I}(on\ time_i) \tag{7}$$

The weights are determined by the Delphi method, and the historical score S_{hist} introduces time decay ($\lambda = 0.1$/month) to emphasise recent performance [19].

2) Market environment characteristics

a) Price volatility characteristics: Econometric estimation of building materials price volatility σ_M:

$$\begin{cases} r_t = \ln(P_t/P_{t-1}) \\ \sigma_t^2 = 0.05 + 0.15r_{t-1}^2 + 0.80\sigma_{t-1}^2 \quad (\text{GARCH(1,1)}) \end{cases} \tag{8}$$

The coefficients are estimated through MLE and reflect volatility aggregation effects.

The labor cost index LCI is calculated cumulatively:

$$LCI_t = LCI_{t-1} \times (1 + g_t), \quad g_t = \frac{W_t - W_{t-1}}{W_{t-1}} + \delta_{\text{policy}} \tag{9}$$

δ_{policy}For policy adjustment factors

b) Financial risk characteristics: Capital Asset Pricing Model with Interest Rate Sensitivity β_r:

$$\beta_r = \frac{\text{Cov}(r_p, r_m)}{\text{Var}(r_m)} = \frac{\sum(r_{p,t} - \bar{r}_p)(r_{m,t} - \bar{r}_m)}{\sum(r_{m,t} - \bar{r}_m)^2} \tag{10}$$

where, r_p is the project IRR, r_m is the market interest rate and the calculation window is taken as 36 months.

Discounted cash flow model for exchange rate risk exposure FX_{exp}:

$$FX_{\text{exp}} = \sum_t \frac{\text{NetCF}_{\text{FC},t}}{(1+r_f)^t} \times \sigma_{\text{FX}} \tag{11}$$

σ_{FX} is the exchange rate volatility, reflecting the time risk of net foreign currency cash flows.

3) Implementation process characteristics

a) Cost deviation characteristics: Earned Value Management Criteria for Cost Performance Index CPI:

$$CPI = \frac{EV}{AC} = \frac{\sum(Completed \times budget\ unit\ price)}{\sum actual\ cost} \tag{12}$$

Level 1 warning triggered when CPI < 0.9 (industry experience threshold).

Cost deviation rate Sensitivity enhancement design for CDR:

$$CDR = \text{sign}(AC - EV) \times \sqrt{\left|\frac{AC-EV}{EV}\right|} \tag{13}$$

The square root transformation amplifies the overrun signal (e.g., 20% overrun translates to CDR = -0.447) [20].

b) Schedule risk characterization: Network planning algorithm for critical path float time FT_{CP}:

$$FT_i = LS_i - ES_i, \quad FT_{CP} = \min_{i \in critical\ path} FT_i \tag{14}$$

Schedule risk criticality is determined when FT_{CP}< 7 days (PERT analysis validation).

Information theoretic definition of schedule deviation entropy S_{dev}:

$$S_{\text{dev}} = -\sum_{k=1}^{K} p_k \ln p_k, \quad p_k = \frac{function_k Number\ of\ days\ delayed}{\sum total\ delay} \tag{15}$$

Entropy values >1.5 indicate risk diversification, and <0.5 indicate critical path concentration of risk

4) Risk association characteristics

a) Historical risk transmission: Risk incidence Bayesian update of HR:

$$HR^{(n)} = \frac{\alpha + \sum \mathbb{I}_{\text{risk}}}{\alpha + \beta + n}, \quad \alpha = 2, \beta = 2 \tag{16}$$

Update the a posterior probability with an initial value $HR^{(0)}$=0.5 for each new item.

b) Multi-risk coupling: Cost-schedule coupling index Vector pinch model for CSI:

$$CSI = \cos\theta = \frac{\Delta C \cdot \Delta S}{\|\Delta C\| \|\Delta S\|}, \quad \Delta C = [\Delta C_1, \dots, \Delta C_T] \tag{17}$$

θ < 30∘ indicates strong coupling (e.g., cost overruns accompanied by schedule delays).

Market-Contract Sensitivity Stress test model for MCS:

$$MCS = \max_{\sigma_M \in [0.1, 0.3]} \left(\frac{\partial \mathcal{R}}{\partial \sigma_M}\Big|_{CT}\right) \tag{18}$$

Calculate the partial derivatives of risk with respect to contract type under extreme scenarios of price volatility.

5) Feature engineering methods

a) EM algorithm for missing value interpolation

Step 1:

$$Q(\theta|\theta^{(t)}) = \mathbb{E}_{Z|X_{\text{obs}},\theta^{(t)}}[\ln P(X_{\text{obs}}, Z|\theta)] \tag{19}$$

Step 2:

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q\left(\theta|\theta^{(t)}\right) \quad (20)$$

b) Interaction feature construction

Explicit Interaction:

$$\phi_{\text{interact}} = \sigma_M \times (1 - CT) \times \log I \quad (21)$$

Capture the extreme risk of 'cost plus fee contracts for large projects in highly volatile markets'.

6) Feature validation: The contribution of feature j to the prediction of f(x):

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \left(f_{S \cup \{j\}}(\mathrm{x}) - f_S(\mathrm{x})\right) \quad (22)$$

The computational time consuming grows exponentially with the number of features, using the TreeSHAP approximation ($\mathcal{O}(TLD^2)$), where T is the number of trees, L is the number of leaf nodes, and D is the depth.

D. Model Training and Validation

The training of the BO-GBM model is an iterative learning process that incorporates Bayesian optimization and gradient boosting machine, and its mathematical nature can be formulated as the following two-layer optimization problem [21]:

$$\begin{aligned} &\textit{Outer layer (hyperparametric optimisation)}: \quad \mathrm{x}^* = \underset{\mathrm{x} \in \mathcal{X}}{\operatorname{argmax}} \mathcal{J}(\mathrm{x}) \\ &\textit{Inner layer (simulation training)}: \quad \mathcal{J}(\mathrm{x}) = \mathbb{E}_{(\mathrm{X}_{\mathrm{val}}, \mathrm{y}_{\mathrm{val}})}[\mathcal{L}(\mathcal{M}_{\mathrm{x}}(\mathrm{X}_{\mathrm{val}}), \mathrm{y}_{\mathrm{val}})] \end{aligned}$$

where, $\mathrm{x} = (\eta, d_{\max}, n_{\text{leaf}}, \dots)$ is the hyperparameter vector and $\mathcal{L}$ is the objective function (e.g., negative RMSE).

1) Dynamic construction of training data: In order to adapt to the temporal characteristics of engineering economic data, a rolling time window strategy is used [22]:

$$\begin{cases} \mathcal{D}_{\text{train}}^{(t)} = \{(\mathrm{x}_i, y_i) | T_i \in [t - \tau, t]\} \\ \mathcal{D}_{\text{val}}^{(t)} = \{(\mathrm{x}_i, y_i) | T_i \in (t, t + \Delta t]\} \end{cases} \quad (23)$$

τ = 24 months: training window length

Δt = 3 months: validation window step length

Timestamp T_i is the project start time

2) BO-GBM Co-training algorithm: The essence of BO-GBM co-training is a two-layer optimization process with a mathematical framework consisting of Bayesian optimization (outer layer) and gradient boosting machine training (inner layer) to form an iterative closed-loop system (see Fig. 2):

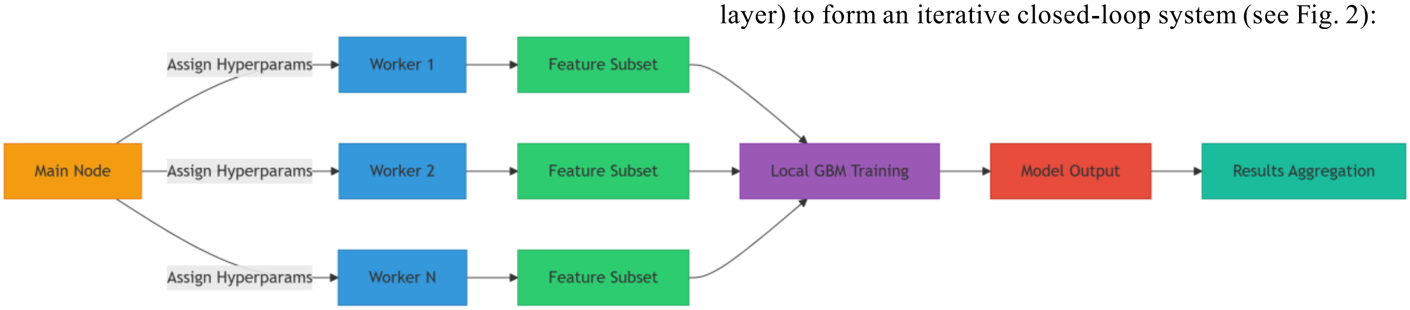

Fig. 2. Parallel training architecture.

a) Bayesian optimization layer: Let the hyperparameter space $\mathcal{X} \subset \mathbb{R}^d$ and the objective function $\mathcal{J}(\mathrm{x})$ be the performance of GBM on the validation set (e.g., AUC) [23]. Bayesian optimization models the objective function through a Gaussian process:

$$\mathcal{J}(\mathrm{x}) \sim \mathcal{GP}\left(\mu(\mathrm{x}), k(\mathrm{x}, \mathrm{x}')\right) \quad (24)$$

where, the kernel function uses Matérn 5/2:

$$k(\mathrm{x}, \mathrm{x}') = \sigma_f^2 \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) \exp(-\sqrt{5}r), \quad r = \sqrt{\sum_{i=1}^{d} \frac{(x_i - x_i')^2}{\ell_i^2}} \quad (25)$$

Expectation Improvement (EI) acquisition function to guide parameter search:

$$\mathrm{EI}(\mathrm{x}) = \mathbb{E}[\max(\mathcal{J}(\mathrm{x}) - \mathcal{J}^+, 0)] = (\mu(\mathrm{x}) - \mathcal{J}^+ - \xi)\Phi(Z) + \sigma(\mathrm{x})\phi(Z) Z = \frac{\mu(\mathrm{x}) - \mathcal{J}^+ - \xi}{\sigma(\mathrm{x})}, \quad \xi = 0.01 \quad (26)$$

$\mathcal{J}^+$ is the current optimal observation and Φ, ϕ is the standard normal distribution function.

b) Gradient booster Training layer: Given the hyperparameters $\mathrm{x} = (\eta, d_{\max}, \lambda, \dots)$, the GBM minimizes the loss in an additive model:

$$F^* = \underset{F}{\operatorname{argmin}} \sum_{i=1}^{n} \mathcal{L}\left(y_i, F(\mathrm{x}_i)\right) + \Omega(F) \quad (27)$$

where, the regular term $\Omega(F) = \gamma T + \frac{1}{2}\lambda \parallel \mathrm{w} \parallel^2$, T is the number of leaf nodes of the tree.

The mth iteration:

computes the pseudo-residuals:

$$r_i^{(m)} = -\frac{\partial \mathcal{L}(y_i, F(\mathrm{x}_i))}{\partial F(\mathrm{x}_i)}\bigg|_{F = F_{m-1}} \quad (28)$$

The fitted decision tree h_m is minimized:

$$\sum_{i=1}^{n} \left[r_i^{(m)} - h_m(\mathrm{x}_i)\right]^2 + \gamma T_m + \frac{1}{2}\lambda \sum_{j=1}^{T_m} w_j^2 \quad (29)$$

Update the model:

$$F_m(\mathrm{x}) = F_{m-1}(\mathrm{x}) + \nu \cdot h_m(\mathrm{x}), \quad \nu = \eta \cdot \exp\left(-\beta \frac{m}{M}\right) \quad (30)$$

β is the decay coefficient and M is the total number of trees.

3) Synergistic iterative mechanism: BO-GBM synergistic training constitutes the dynamical system:

$$\begin{cases} x^{(k+1)} = \mathcal{B}\left(x^{(k)} \mid \{\mathcal{J}(x^{(i)})\}_{i=1}^{k}\right) \\ \mathcal{J}(x^{(k)}) = \mathcal{T}_{\text{GBM}}\left(\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}, x^{(k)}\right) \end{cases} \tag{31}$$

where, $\mathcal{B}$ is the Bayesian optimization operator and $\mathcal{T}_{\text{GBM}}$ is the GBM training evaluation operator [24].

The sequence $\{\mathcal{J}(x^{(k)})\}$ converges with probability 1 when the hyperparameter space $\mathcal{X}$ is tight and the objective function $\mathcal{J}$Lipschitz is continuous:

$$\lim_{k\to\infty} \mathbb{P}\left(\mathcal{J}(x^{(k)}) - \mathcal{J}^* < \epsilon\right) = 1, \quad \forall \epsilon > 0 \tag{32}$$

4) Computational acceleration strategy

a) Agent model warm-up: Pre-training the auxiliary Gaussian process using historical project data$\mathcal{GP}_0$:

$$\mu_0(x) = \mathbb{E}_{\mathcal{D}_{\text{hist}}}[\mathcal{J}(x)], \quad k_0(x, x') = \text{Cov}_{\mathcal{D}_{\text{hist}}}(\mathcal{J}(x), \mathcal{J}(x')) \tag{33}$$

The initial objective function is estimated as:

$$\mathcal{J}^{(1)}(x) \sim \mathcal{GP}\left(\mu_0(x), \rho k_0(x, x') + k(x, x')\right), \quad \rho = 0.3 \tag{34}$$

b) Gradient-sensitive sampling: Introducing gradient information in EI optimization:

$$x_{\text{next}} = \text{argmax}[\text{EI}(x) + \lambda \parallel \nabla_x \mu(x) \parallel_2] \tag{35}$$

λ Balancing exploration and exploitation to accelerate local convergence [25].

c) Partial evaluation mechanisms: When $\sigma(x) > \sigma_{\text{thres}}$, sub-sampled data are used for evaluation:

$$\tilde{\mathcal{J}}(x) = \mathcal{J}(x; \mathcal{D}_{\text{sub}}), \quad |\mathcal{D}_{\text{sub}}| = \min\left(500, \frac{n}{\sigma(x)/\sigma_{\max}}\right) \tag{36}$$

III. Engineering Economic Risk Management Platform Architecture Design

A. Overall Platform Architecture Design

The engineering economic risk management platform adopts vertical layered architecture, with clear functions and interfaces of each layer to ensure high availability and scalability of the system (see Fig. 3).

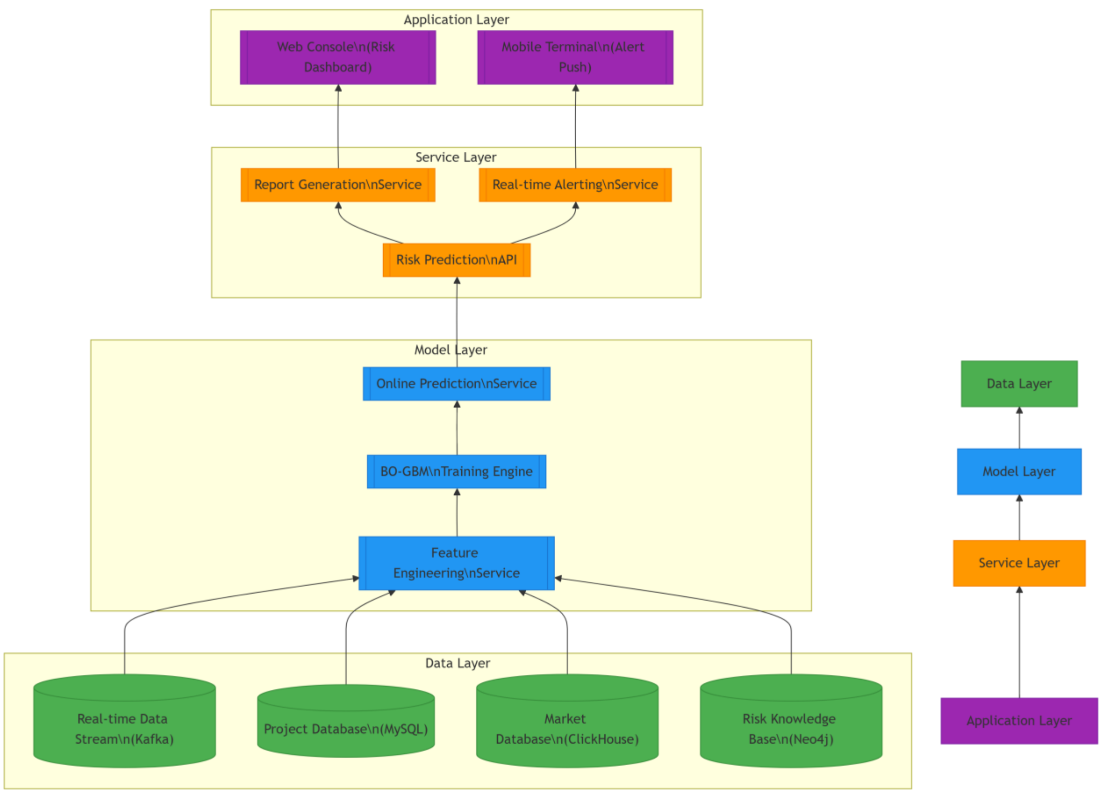

Fig. 3. Platform architecture diagram.

The whole system architecture consists of four distinct layers. The data layer is the foundation and contains both static and dynamic data sources. Storage includes a project database (storing core project data such as contracts, schedules, and costs), a market database (integrating time-series data such as building material prices, exchange rates, and interest rates), and a risk knowledge base (accumulating historical risk cases and response scenarios). Real-time data processing is handled by a Kafka streaming engine that feeds into the business system data and processes it at a rate of over 10,000 data points per minute. The model layer builds on this data foundation and contains the core machine learning components. The BO-GBM training engine automatically optimizes model parameters. The feature engineering service computes 28 real-time risk features. Finally, the Online Prediction Service deploys optimized risk prediction models to ensure that single prediction latency is kept below 300 milliseconds, and the models are updated through daily automated training. The service layer exposes the system's forecasting capabilities through a set of core services. The Risk Prediction API uses gRPC to calculate project risk probabilities in real time. The real-time alert service uses WebSockets to trigger multi-level alerts. The report generation service creates risk assessment reports automatically via a REST API. To ensure resilience, the service implements automatic meltdown and switches to a backup service when the error rate exceeds 5%. Load balancing supports more than 500 concurrent requests. Finally, the application layer provides the user interface and access control. The web console provides real-time risk heat maps, risk factor analysis tools, and alert threshold configuration options. Mobile terminals provide instant risk alerts, project status queries, and access to emergency response channels. RBAC (role-based access control) is implemented for rights management, and two-factor authentication is used to enhance security.

B. Core Functional Module Design

The engineering economic risk management platform contains five core functional modules, which together achieve the closed-loop management of the whole process of risk prediction, assessment, early warning, and decision-making support. The design of each module follows the principle of ‘high cohesion and low coupling’, and works together through standardized API interfaces.

1) Data integration and management module: The platform is positioned as a data hub, providing unified access and governance for various heterogeneous data sources. Key design elements include: Multi-source data access: This covers a wide range of systems, including obtaining progress and cost data from project management systems (e.g. Primavera P6) via ODBC, synchronizing payment and settlement information from financial systems [27] (e.g. SAP) via RFC protocols, retrieving building material prices and exchange rate indices on a regular basis from market databases (e.g. Wind) via APIs, and collecting data on the operation of field equipment using IoT sensors. IoT sensors to collect data on the operation of equipment in the field. A key component is the intelligent data pipeline, which contains cleansing rules (e.g., missing value filling and outlier correction using the ±3σ principle) and transformation logic that automatically generates 28-dimensional feature vectors based on the feature system described earlier. The storage strategy is tiered for optimal performance and cost-effectiveness: hot data is cached in Redis (with response times of less than 50 milliseconds), warm data is stored in ClickHouse columnar format, and cold data is archived in HDFS. Finally, comprehensive metadata management provides data lineage tracking, recording field-level source and transformation history, and quality monitoring dashboards that display data completeness and timeliness metrics in real time (see Fig. 4).

2) BO-GBM model management module: Functional positioning is the core of model full life cycle management. The key design includes: training workflow: multiple triggering mechanisms are available, including timed triggering (2:00 am every day), data drift warning triggering (when the KL dispersion of the feature distribution is > 0.1), and manual triggering. Spark ML is used for parallel processing of 100 GB data for distributed training. Each model version has version metadata, including snapshots of the training data, hyperparameters, and evaluation metrics. Grey scale releases validate the performance of new model versions by introducing 10% traffic. Additionally, the Parameter Configuration Centre provides a BO optimization space that allows for visual adjustment of hyperparameter ranges and supports manual intervention of feature weights to adjust feature importance assignments (see Fig. 5).

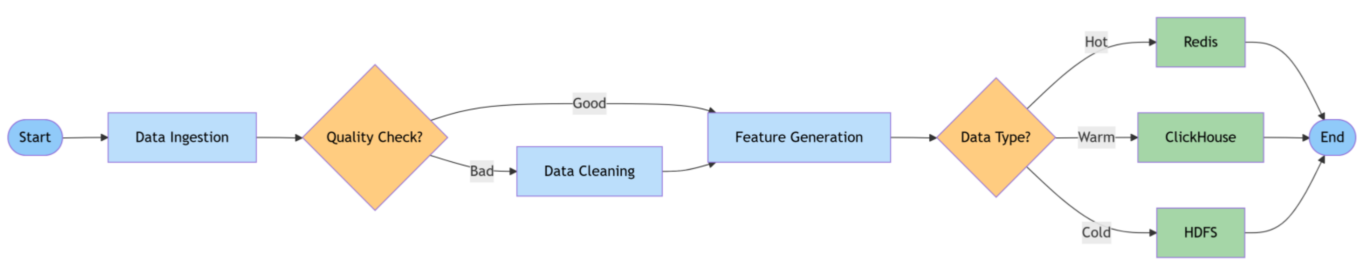

Fig. 4. Intelligent data pipeline flowchart.

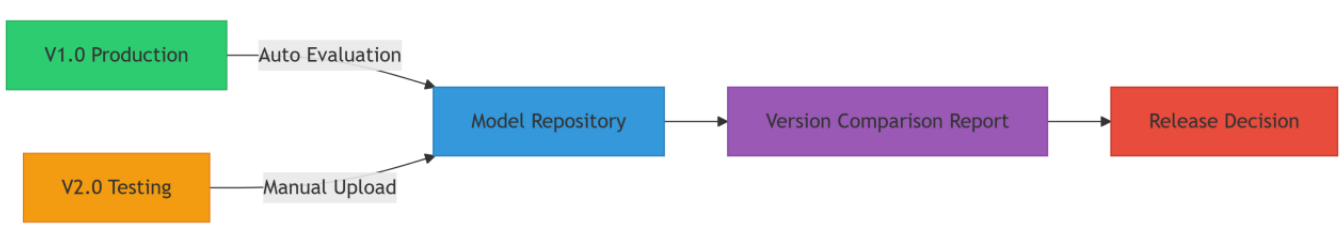

Fig. 5. Version control system.

IV. Platform Core Characteristics Analysis and Verification

A. Validation of the BO-GBM Model's Prediction Performance Characteristics

This experiment verifies the performance advantages of the BO-GBM model in engineering economic risk prediction tasks through rigorous comparison tests. The experimental design focuses on three core questions:

Does BO-GBM significantly outperform mainstream baseline models?

How effective is Bayesian optimization in improving model robustness?

How much does the feature engineering system contribute to the prediction accuracy?

1) Experimental design: The dataset is derived from the data of 327 projects of a large engineering group between 2018 and 2023. The dataset is divided into a training set and a test set, where the training set contains 256 projects from 2018-2022 and the test set contains 71 projects from 2023. Each project is represented by 28 features, and the labels are dichotomized to indicate whether a significant economic risk has occurred (see Table III).

Evaluation metrics include AUC, F1-score, and recall for classification tasks and RMSE and MAPE for regression tasks (cost bias prediction).

2) Performance comparison metrics: This experiment evaluated the classification performance of the five models on 71 independent test items (risk occurrence prediction). All models use the same training set (256 items) and feature set (28 dimensions), and the evaluation metrics include AUC (area under the curve), F1-score (the reconciled average of precision and recall), and recall (the proportion of actual risks correctly identified). The test set consists of 18 high-risk projects (where a significant economic risk actually occurs) and 53 low-risk projects completed in 2023 (see Fig. 6).

The BO-GBM model achieves an excellent performance of 0.927 on the AUC metric, which is significantly higher than the standard GBM (0.865) and Random Forest (0.832). In terms of F1-score, BO-GBM leads the other models with 0.892, which is 7.3% higher than the standard GBM (0.831). The recall metric shows that BO-GBM identifies 91.3% of actual risky items, 17.5 percentage points higher than logistic regression (73.8%). These data demonstrate that Bayesian optimization significantly improves the discriminative power of the GBM model, especially in the identification of high-risk items.

TABLE III COMPARATIVE MODELS

Model type	Parameter Settings	Implementation library
BO-GBM	Tree Depth = 6, Learning Rate = 0.1 (BO Optimization)	LightGBM
Standard GBM	Tree Depth = 6, Learning Rate = 0.1	Scikit-learn
Random Forest (RF)	No. of Trees = 200, Depth = 10	Scikit-learn
SVM	Kernel Function = RBF, C = 1.0	Scikit-learn
Logistic Regression (LR)	Regularization Strength = 1.0	Scikit-learn

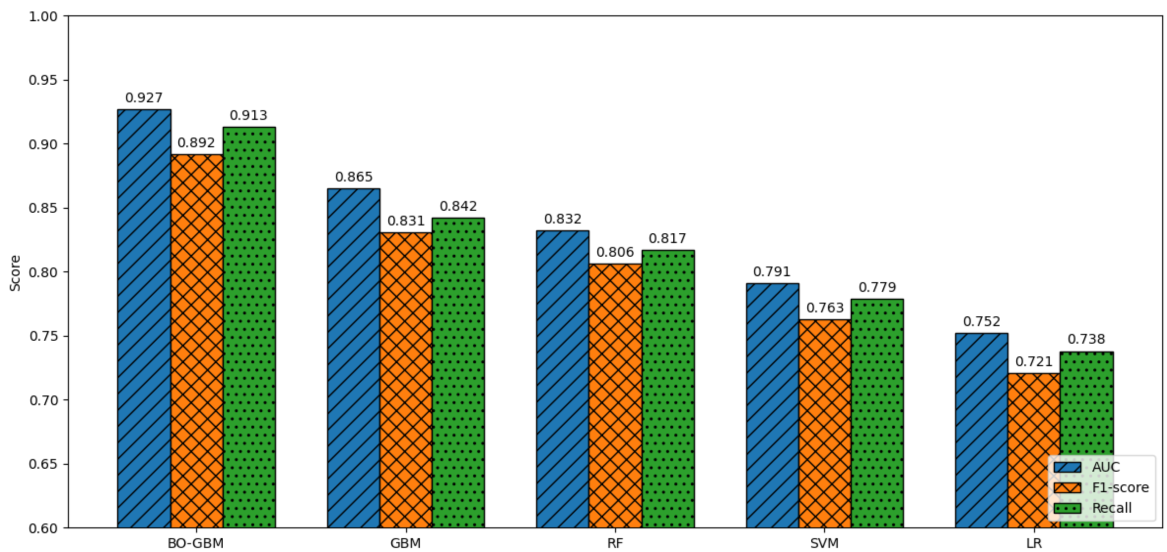

Fig. 6. Comparison of classification model performance.

This experiment evaluated the accuracy of each model in a cost deviation prediction task (regression problem). RMSE (Root Mean Square Error, in millions of dollars) and MAPE (Mean Absolute Percentage Error) were used as assessment metrics. The test set contained actual cost deviation data for 71 projects, with deviations ranging from -35% (savings) to +82% (overruns), and an average absolute deviation of $287,000 (see Fig. 7).

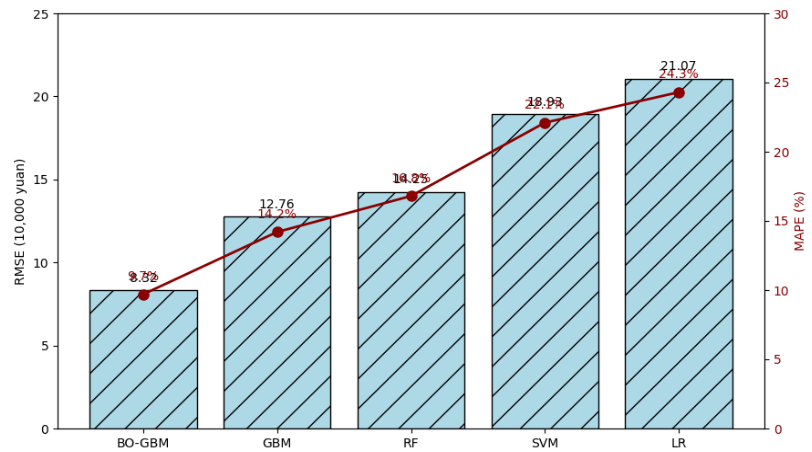

Fig. 7. Comparison of cost bias prediction errors.

The RMSE of BO-GBM is $83,200, which is 34.8% lower than the standard GBM ($127,600) and 60.5% lower than the worst performing logistic regression ($210,700). On the MAPE metric, BO-GBM's prediction error was only 9.7%, breaking the 10% engineering management accuracy threshold for the first time. Notably, when the cost deviation exceeds 30%, the prediction error of BO-GBM (MAPE = 11.2%) is still significantly lower than that of other models (17.8% for standard GBM), indicating that it still maintains high accuracy under extreme risk scenarios.

3) Result analysis: To verify the model's ability to resist interference, Gaussian noise (mean 0, standard deviation from 0% to 20%) was added to the test set of features. The experiment was repeated 50 times at each noise level, and the change in the mean AUC value was recorded. The range of noise addition covers all 28 features to simulate the data acquisition errors in real applications (see Fig. 8).

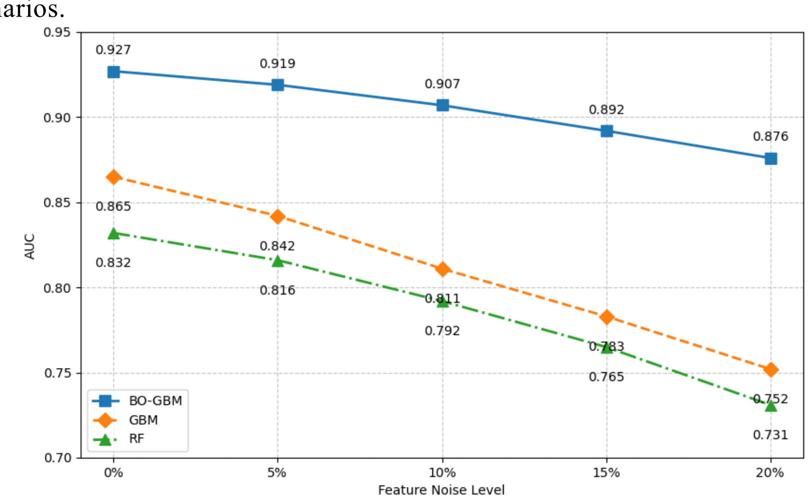

Fig. 8. Model robustness under feature perturbation.

When the feature noise reaches 20%, the AUC of BO-GBM decreases from 0.927 to 0.876 (a decrease of 5.5%), while the standard GBM decreases from 0.865 to 0.752 (a decrease of 13.1%). At a 10% noise level, the AUC of BO-GBM stays above 0.907, which is significantly higher than that of the standard GBM at 0.811. Random Forest performs close to BO-GBM (0.816 vs. 0.919) at low noise (5%), but the AUC decreases to 0.731 at high noise (20%), widening the gap to 14.5 percentage points. This indicates that BO optimization effectively improves the model noise immunity through hyperparameter tuning.

The contribution of each feature to the prediction of the BO-GBM model was quantified using the SHAP value method. The absolute mean of SHAP based on all samples in the test set was calculated, and the five features with the highest contribution were selected. The SHAP value indicates the average magnitude of the effect of feature changes on the model output (risk probability).

The Cost Performance Index (CPI) tops the list with a contribution of 0.218, proving that cost control failure is the strongest risk signal. Building material volatility (0.195) and contract type (0.172) rank second and third, and the sum of their contributions (0.367) exceeds the CPI, reflecting the key role of external markets and contract design. Progress deviation (0.141) and cash flow gap (0.103) constitute the second tier of risk factors. The total contribution of TOP5 features reaches 0.829, covering the three core dimensions of engineering economic risk: cost control, external environment and project execution (see Fig. 9).

B. Platform Functional Characteristics and Efficiency Verification

1) Risk identification efficiency: Fifty projects implemented by a large infrastructure company in 2023 are selected for risk assessment using both traditional manual assessment and automated identification by the platform. Record the time consumed for the whole process from data preparation to the output of the risk assessment report. Manual assessment is performed by a team of three senior risk analysts.

Platform's risk identification efficiency is significantly better than manual methods. For large-sized projects, the identification time is reduced from 42.6 hours to 0.52 hours, an 81.9-fold increase in efficiency. For medium-sized projects, the identification time is reduced from 25.7 hours to 0.38 hours, an increase in efficiency of 67.6 times. Small projects are processed through the platform in 0.25 hours, 49.2 times faster than manual processing (12.3 hours). The logarithmic coordinates show that the platform processing time is basically not affected by the project scale, which verifies the elasticity and scalability of the architecture (see Fig. 10).

2) Predicting response time: The experimental design includes simulating different concurrency scenarios in the load test environment: single project prediction (1000 consecutive requests), multi-project batch prediction (10-100 projects/batch), and high concurrency scenarios (50-500 concurrent users). The platform's response time (P95) and resource consumption (CPU/memory) were recorded. The test environment is an 8-core 16GB cloud server [26].

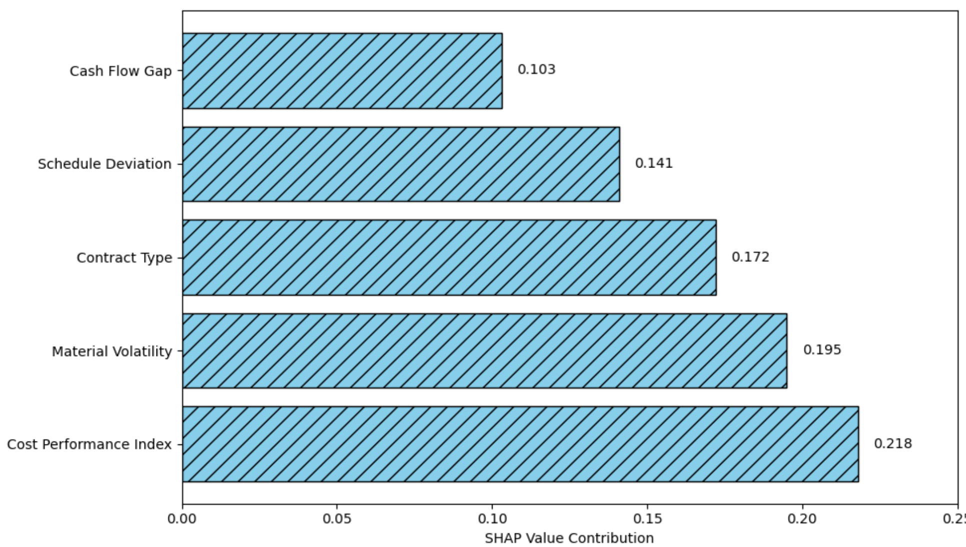

Fig. 9. TOP5 risk characteristics contribution.

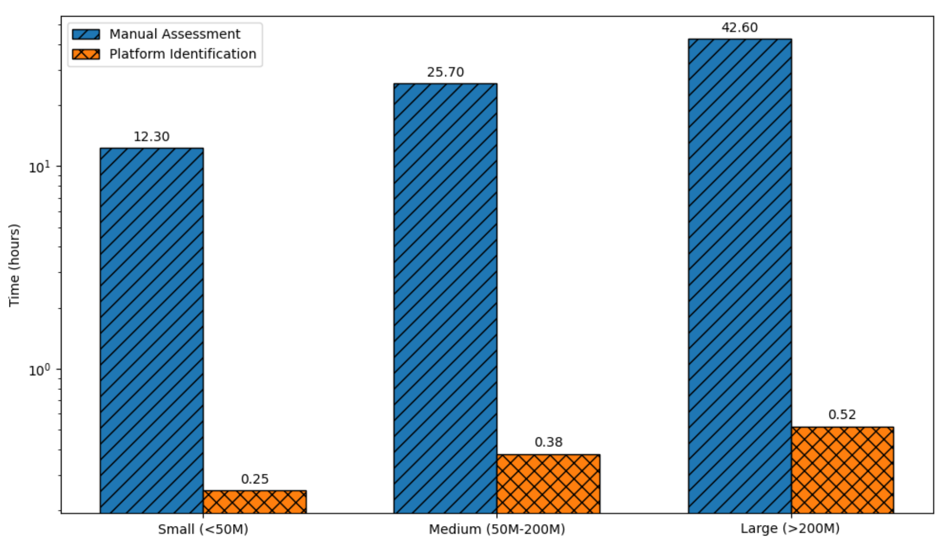

Fig. 10. Comparison of risk identification time for projects of different sizes.

The average response time for a single project prediction is 127 ms, which meets the real-time decision-making requirements. Batch prediction of 100 projects takes 256 ms, with a processing speed of 390 projects per second. In the 500 concurrent users scenario, the P95 response time was 427 milliseconds, still below the engineering threshold of 500 milliseconds. Response times for all test scenarios were below the industry best practice standard of 300 milliseconds. Resource monitoring shows that with 500 concurrent users, CPU utilization is 78% and memory usage is 68%, indicating that the system still has room for further expansion (see Fig. 11).

3) Early warning accuracy verification: The platform was used to backtest the data of 120 projects that had been completed in the past. The specific steps are: input feature data according to the actual project progress nodes, record the warning signals issued by the platform and the time, and compare the warning signals with the time and type of actual risk occurrence. Evaluation indicators include: warning accuracy rate, false alarm rate, and missed alarm rate.

The platform's early warning accuracy rate is 72.5% (87/120), of which 92.3% for high-risk projects (L4-L5). False alarm rate is 7.5% (9/120), which mainly occurs in scenarios of sudden policy changes, such as new environmental regulations. False alarms were 5.0% (6/120), mainly in extreme cases where the schedule was compressed by more than 30%. Overall, the platform made a correct judgment (accurate warning + correct no warning) in 87.5% of cases. The average lead time for high-risk warnings was 28 days for cost risks and 42 days for schedule risks, which meets the emergency response time requirements (see Fig. 12).

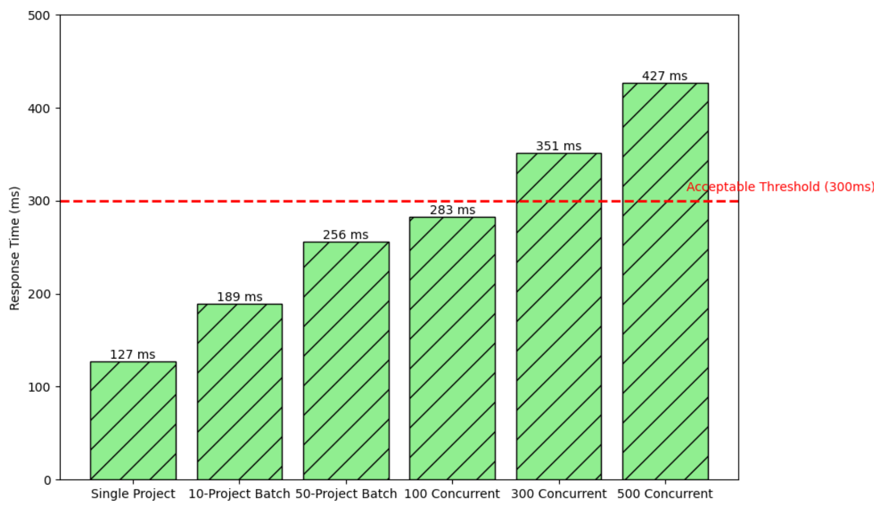

Fig. 11. Risk prediction response time in different scenarios.

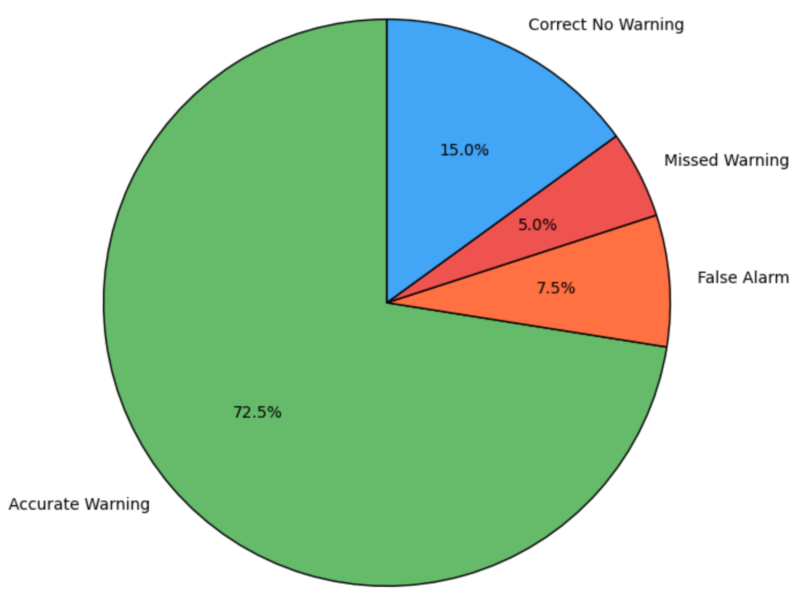

Fig. 12. Analysis of the accuracy of the platform's early warning.

V. Results and Discussion

This section presents a comprehensive evaluation of the proposed BO-GBM model's predictive performance and an analysis of the platform's functional characteristics and efficiency, followed by a discussion of the implications of these findings.

A. Validation of BO-GBM Model's Prediction Performance

The experiment verifies the performance advantages of the BO-GBM model through rigorous comparison tests on a dataset comprising 327 projects from a large engineering group (2018 to 2023), split into training (2018 to 2022, 256 projects) and test (2023, 71 projects) sets. Each project is represented by the 28 features described previously.

1) Experimental setup and baselines: The BO-GBM model was compared against several mainstream baseline models, as detailed in Table IV.

TABLE IV COMPARATIVE MODELS AND PARAMETER SETTINGS

Model Type	Parameter Settings	Implementation Library
BO-GBM	Tree Depth=6, Learning Rate=0.1 (BO Optimized)	LightGBM
Standard GBM	Tree Depth=6, Learning Rate=0.1	Scikit-learn
Random Forest (RF)	No. of Trees=200, Depth=10	Scikit-learn
SVM	Kernel Function=RBF, C=1.0	Scikit-learn
Logistic Regression (LR)	Regularization Strength=1.0	Scikit-learn

2) Performance comparison

a) Classification performance: Fig. 6 shows the performance on the risk occurrence prediction task (18 high-risk, 53 low-risk projects). The BO-GBM model achieved an AUC of 0.927, significantly outperforming the standard GBM (0.865) and Random Forest (0.832). Its F1-score of 0.892 was 7.3% higher than the standard GBM. Crucially, the recall rate of BO-GBM reached 91.3%, meaning it identified over 91% of actual high-risk projects, which is 17.5 percentage points higher than Logistic Regression (73.8%). This demonstrates that Bayesian optimization significantly enhances the model's ability to discriminate, particularly in identifying critical risks.

b) Regression performance (cost deviation): Fig. 7 compares the models on predicting cost deviation. The BO-GBM model achieved an RMSE of 83,200 RMB, which is 34.8% lower than the standard GBM (127,600 RMB) and 60.5% lower than Logistic Regression (210,700 RMB). Its MAPE was only 9.7%, breaking the 10% accuracy threshold often sought in engineering management. Notably, even for extreme cost overruns (>30%), BO-GBM maintained a lower error (MAPE=11.2%) compared to other models (e.g., 17.8% for standard GBM).

3) Robustness analysis: To test the model's resistance to data noise, Gaussian noise (mean 0, standard deviation from 0% to 20%) was added to the test features. As shown in Fig. 8, when feature noise reached 20%, the AUC of BO-GBM decreased by only 5.5% (from 0.927 to 0.876), whereas the standard GBM decreased by 13.1% (0.865 to 0.752). This indicates that the hyperparameters found by BO contribute to a more robust model that is less sensitive to data perturbations, a critical property for real-world applications where data quality can vary.

4) Feature importance analysis: Using SHAP values, we quantified the contribution of each feature. Fig. 9 reveals that the Cost Performance Index (CPI) was the most influential feature (contribution 0.218), confirming failed cost control as the primary risk signal. Building material volatility (0.195) and contract type (0.172) were the next most important, highlighting the significant role of external markets and contractual design. Schedule deviation (0.141) and cash flow gap (0.103) formed a secondary tier. The combined contribution of the top five features was 0.829, effectively capturing the core dimensions of engineering economic risk.

B. Platform Functional Characteristics and Efficiency Verification

1) Risk identification efficiency: We compared the platform's automated risk identification against traditional manual assessment by a team of three senior analysts across 50 projects of different sizes in 2023. The results, depicted in Fig. 10, show a dramatic efficiency improvement. For large projects, identification time was reduced from 42.6 hours to 0.52 hours—an 81.9-fold increase. The platform's processing time remained relatively constant regardless of project scale, demonstrating its scalability and architectural elasticity.

2) Prediction response time: Load testing under various scenarios (Fig. 11) confirmed the platform's real-time capability. The average response time for a single prediction was 127 ms. Batch prediction of 100 projects took 256 ms (≈ 390 projects/second). Under a high load of 500 concurrent users, the P95 response time was 427 ms, remaining below the 500 ms engineering threshold. Resource utilization (CPU 78%, Memory 68%) under this load indicated potential for further scaling.

3) Early warning accuracy: A backtest was conducted on 120 completed projects. The platform achieved an overall warning accuracy of 72.5% (87/120), with accuracy for high-risk projects (L4-L5) exceeding 92.3%. The false alarm rate was 7.5%, primarily triggered by unforeseen policy changes, and the missed alarm rate was 5.0%, mainly occurring in cases of extreme schedule compression (>30%). Overall, the platform made correct judgments (accurate warning + correct non-warning) in 87.5% of cases. The average lead time for high-risk warnings was 28 days for cost risks and 42 days for schedule risks, providing sufficient time for proactive mitigation.

C. Discussion

The results strongly support the effectiveness of the proposed BO-GBM model and the associated platform. The significant performance gains over baseline models, particularly in recall and robustness, underscore the value of using Bayesian Optimization for hyperparameter tuning in this domain. The high contribution of the engineered features validates the comprehensiveness of the four-dimensional feature system. The platform's operational metrics confirm its practical utility, offering order-of-magnitude efficiency gains in risk identification and reliable, real-time predictions suitable for large-scale, concurrent use.

The primary limitation observed was related to warning errors. False alarms were often linked to "black swan" events like sudden policy shifts, which are not captured by historical feature data. Missed alarms occurred under extreme project conditions, suggesting potential model performance boundaries or the need for even more specialized features for such edge cases. These points inform valuable directions for future work.

VI. Conclusion

In this study, a platform for engineering economic risk management based on the BO-GBM model is successfully constructed, which significantly improves the prediction accuracy and control efficiency of engineering economic risks by integrating Bayesian optimization and gradient boosting machine techniques. The core achievements can be summarized in the following three aspects:

1) Excellent performance of BO-GBM model: By adopting Bayesian optimization to adaptively adjust the GBM hyperparameters, the model breaks through the efficiency bottleneck of traditional tuning methods and significantly enhances the model's robustness. Experimental results show that in the risk prediction task, the model AUC reaches 0.927, with a recall rate of 91.3%, which is 17.5 percentage points higher than that of the standard GBM; in cost deviation prediction, the RMSE is reduced to 83.2 thousand yuan, and the MAPE is only 9.7%, which is more than 30% lower than the mainstream model even in the extreme overrun scenarios; in the face of the featured noise interference of 20%, the model performance degradation is less than 20%. Even under extreme overspending scenarios, the error is still lower than mainstream models by more than 30%; in the face of 20% of characteristic noise interference, the model's performance degradation is less than 6%, which verifies its strong anti-interference ability.

2) Breakthrough in risk characteristics system and platform effectiveness: The four-dimensional 28-feature system (project ontology, market environment, execution process, and risk correlation) constructed systematically quantifies the risk-driving mechanism, with the contribution of key features exceeding 82%. Among them, cost performance index, building material volatility, and contract type constitute the core risk signals. The platform adopts a layered architecture (data layer, model layer, service layer, application layer), supporting millisecond response with single prediction latency of less than 127 milliseconds, and P95 response time of 427 milliseconds with 500 concurrent users. In addition, the efficiency of risk identification has been improved by an order of magnitude, and the evaluation time of large projects has been compressed from 42.6 hours to 0.52 hours, which is 81.9 times more efficient.

3) Precise and reliable early warning mechanism: The platform realizes closed-loop risk management for the whole process. The accuracy rate of early warning reaches 72.5%, and the accuracy rate of high-risk early warning exceeds 92%. The advance warning period for cost risk and schedule risk reaches 28 days and 42 days, respectively, which meets the demand for emergency response. The false alarm rate and omission rate are controlled at 7.5% and 5.0% respectively, which are significantly better than the manual assessment mode.

Building upon current research findings and limitations, future studies will focus on advancing three key areas of exploration. Regarding model optimization, hybrid time-series models incorporating LSTM or Transformer architectures will be developed to more accurately capture the dynamic evolution of project risks. Concurrently, the BO-GBM framework will be expanded to enable multi-task collaborative prediction of cost, schedule, and safety risks. Regarding data dimensions, plans include introducing natural language processing techniques to extract latent risk signals from unstructured data such as textual reports, alongside constructing adaptive feature selection mechanisms to accommodate requirements across different project phases. At the platform functionality level, efforts will strengthen visualization capabilities to intuitively display model decision-making logic, while integrating reinforcement learning technologies to enable dynamic, intelligent recommendations for risk mitigation strategies.

Funding

Chongqing Municipal Education Commission Science and Technology Research Project: Key Technologies for Structural Health Monitoring Data Reconstruction Based on Deep Models (KJQN202301803).

References

[1] Landry M, Erlinger T P, Patschke D, et al. Probabilistic gradient boosting machines for GEFCom2014 wind forecasting. International Journal of Forecasting, 2016, 32(3): 1061-1066.

[2] Yoon J. Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. Computational Economics, 2021, 57(1): 247-265.

[3] Singh U, Rizwan M, Alaraj M, et al. A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments. Energies, 2021, 14(16): 5196.

[4] Lloyd J R. GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes. International Journal of Forecasting, 2014, 30(2): 369-374.

[5] Taieb S B, Hyndman R J. A gradient boosting approach to the Kaggle load forecasting competition. International journal of forecasting, 2014, 30(2): 382-394.

[6] Xenochristou M, Hutton C, Hofman J, et al. Water demand forecasting accuracy and influencing factors at different spatial scales using a gradient boosting machine. Water Resources Research, 2020, 56(8): e2019WR026304.

[7] Noorunnahar M, Chowdhury A H, Mila F A. A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. PloS one, 2023, 18(3): e0283452.

[8] Callens A, Morichon D, Abadie S, et al. Using Random forest and Gradient boosting trees to improve wave forecast at a specific location.

Applied Ocean Research, 2020, 104: 102339.

[9] Park S, Jung S, Lee J, et al. A short-term forecasting of wind power outputs based on gradient boosting regression tree algorithms. Energies, 2023, 16(3): 1132.

[10] Panarese A, Settanni G, Vitti V, et al. Develo** and preliminary testing of a machine learning-based platform for sales forecasting using a gradient boosting approach. Applied Sciences, 2022, 12(21): 11054.

[11] Du L, Gao R, Suganthan P N, et al. Bayesian optimization based dynamic ensemble for time series forecasting. Information Sciences, 2022, 591: 155-175.

[12] Sultana N, Hossain S M Z, Almuhaini S H, et al. Bayesian optimization algorithm-based statistical and machine learning approaches for forecasting short-term electricity demand. Energies, 2022, 15(9): 3425.

[13] Cheng H, Ding X, Zhou W, et al. A hybrid electricity price forecasting model with Bayesian optimization for German energy exchange. International Journal of Electrical Power & Energy Systems, 2019, 110: 653-666.

[14] Liwei T, Li F, Yu S, et al. Forecast of LSTM-XGBoost in Stock Price Based on Bayesian Optimization. Intelligent Automation & Soft Computing, 2021, 29(3).

[15] Zulfiqar M, Gamage K A A, Kamran M, et al. Hyperparameter optimization of bayesian neural network using bayesian optimization and intelligent feature engineering for load forecasting. Sensors, 2022, 22(12): 4446.

[16] He F, Zhou J, Feng Z, et al. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. Applied energy, 2019, 237: 103-116.

[17] Habtemariam E T, Kekeba K, Martínez-Ballesteros M, et al. A Bayesian optimization-based LSTM model for wind power forecasting in the Adama district, Ethiopia. Energies, 2023, 16(5): 2317.

[18] Abbasimehr H, Paki R. Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization. Chaos, Solitons & Fractals, 2021, 142: 110511.

[19] Aslam M, Lee S J, Khang S H, et al. Two-stage attention over LSTM with Bayesian optimization for day-ahead solar power forecasting. IEEE Access, 2021, 9: 107387-107398.

[20] X B, Zheng W Z, Kong J L, et al. Deep-learning forecasting method for electric power load via attention-based encoder-decoder with bayesian optimization. Energies, 2021, 14(6): 1596.

[21] Nair S K, Botros D, Chakravarti S, et al. Predictors of surgical site infection in glioblastoma patients undergoing craniotomy for tumor resection. Journal of neurosurgery, 2022, 138(5): 1227-1234.

[22] Bidyuk P, Prosyankina-Zharova T, Terentiev O, et al. Adaptive modelling for forecasting economic and financial risks under uncertainty conditions in terms of the economic crisis and social threats. Технологический аудит и резервы производства, 2018, 4(2(42)): 4-10.

[23] Gorokhovatskyi V, Sergienko O, Sosnov I, et al. Risk assessment of innovative projects: Development of forecasting models. CEUR Workshop Proc. 2021, 2927(2020): 18-37.

[24] Danilov V Y, Gozhyj O P, Kalinina I O, et al. Adaptive forecasting and financial risk estimation. System research and information technologies, 2020 (1): 34-53.

[25] Zhao Y, Liu X, Zhao Y. Forecast for construction engineering risk based on fuzzy sets and systems theory. Systems Engineering Procedia, 2011, 1: 156-161.

[26] Bhardwaj A, Sharma A, Raj D, et al. Dynamic and Scalable Privacy-Preserving Group Data Sharing in Secure Cloud Computing. Advanced Cyber Security Techniques for Data, Blockchain, IoT, and Network Protection. IGI Global Scientific Publishing, 2025: 89-122.

[27] Oko-Odion C, Angela O. Risk management frameworks for financial institutions in a rapidly changing economic landscape. International Journal of Science Research Archive, 2025, 14(1): 1182-1204.

[28] Wuni I Y. Develo, A multidimensional risk assessment model for sustainable construction projects. Engineering, Construction and Architectural Management, 2025, 32(6): 4155-4173.