Research Square

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

Project Cost Estimation Based on BIM and Deep Learning

Li Su

sulidz1991@163.com

Dazhou Vocational and Technical College

Article

Keywords: Project cost estimation, Building information modeling (BIM), Grey BP neural network, Sparrow search algorithm, Dynamic cost management

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11352674/v1

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

Additional Declarations: No competing interests reported.

Project Cost Estimation Based on BIM and Deep Learning

Li Su

School of Intelligent Construction, Dazhou Vocational and Technical College, Dazhou 635001, Sichuan, China

Email: sulidz1991@163.com

Abstract: The accuracy of project cost estimation directly affects the effectiveness of construction management, and the nonlinear fluctuation of material prices is a major source of estimation bias. To address this issue, this paper proposes a dynamic cost estimation framework that couples 5D quantities extracted from building information modeling (BIM) with a grey BP neural network optimised by the sparrow search algorithm (SSA-PGNN) for material price forecasting. Unlike traditional static approaches, the framework introduces a bidirectional event-driven data interface and a dual-threshold feedback mechanism that support dynamic updating of cost estimates and proactive risk immunisation. Using data from 62 real construction projects in Liaoning Province, China (2016–2021), material quantities are extracted from BIM models and monthly material prices are forecast by the SSA-PGNN model. The proposed model achieves a mean absolute percentage error (MAPE) of 2.43%–4.04% for material prices and 2.9% for total project cost under leave-one-project-out cross-validation, outperforming the PGNN, SSA-BP, XGBoost, random forest and LSTM benchmarks ($p < 0.05$) within the present dataset. Time-series rolling validation, a synthetic extreme-shock stress test and statistical hypothesis testing support the stability and robustness of the proposed method; however, as the evidence derives from a single regional dataset, external validation beyond the study region remains an essential next step.

Keywords: Project cost estimation; Building information modeling (BIM); Grey BP neural network; Sparrow search algorithm; Dynamic cost management

1. Introduction

Cost overruns have long plagued large-scale construction projects, and a considerable proportion of such projects ultimately exceed their budgets

(Annamalaisami & Kuppuswamy, 2022). Traditional cost estimation methods—such as parametric models, regression analysis, and time-series approaches—are largely based on linearity and stationarity assumptions, and thus exhibit notable limitations when dealing with the complex nonlinear relationships inherent in construction, struggling to accurately capture the dynamic interplay between market prices and material quantities (Cheng et al., 2013). The emergence of Building Information Modeling (BIM) technology has opened new possibilities for cost management, enabling 5D (3D + schedule + cost) digital integration through the combination of three-dimensional geometric information, scheduling, and cost data (Eastman et al., 2011). However, conventional BIM systems rely on static price databases and cannot respond to real-time material market fluctuations, resulting in considerable discrepancies between estimates and actual costs (Yang et al., 2022).

Despite recent advances in BIM-driven optimisation, existing studies have generally overlooked the financial risks induced by market volatility. For instance, Zhang (2025) focused on geometric integrity and differential privacy protection, while Luu (2025) optimised time-cost trade-offs under static penalty mechanisms. Neither framework incorporates the chaotic nonlinear dynamics of material spot prices into the BIM 5D feedback loop, leaving BIM systems "blind" to real-time financial risks and forcing them to respond passively to cost overruns rather than proactively immunising against them. To fill this gap, the present study introduces a novel paradigm of "Cost-Immunity BIM," aiming to shift the objective from merely improving prediction accuracy to enhancing the system's dynamic resilience against price risks.

Building upon the above rationale, this paper constructs a dynamic closed-loop BIM–machine learning coupling framework to overcome the inherent limitations of static cost estimation. The core contributions are threefold. At the methodological level, we propose an SSA-PGNN model that embeds a dual-residual learning paradigm, which decomposes the forecasting task into grey-model-driven trend extraction and neural-network-driven residual correction, offering a data-efficient solution for small-sample, partial-information systems; this decomposition reduces

the learning complexity of each subtask, enabling the model to maintain high prediction accuracy even with limited historical data. At the system architecture level, we design an event-driven dynamic cost aggregation module and establish a bidirectional data interface between BIM and the machine learning model, transforming cost estimation from a one-off static application into an adaptive management tool; the interface supports automatic data updates and recalculation triggered on a weekly basis or when price changes exceed a preset threshold, and a dual-threshold feedback mechanism links cost forecasts with procurement and design-alternative decisions. At the empirical level, we construct a reproducible benchmarking system based on 62 real projects, employ two rigorous validation strategies—leave-one-project-out cross-validation and time-series rolling validation—and conduct transparent comparisons with mainstream methods including random forest, XGBoost, PGNN, SSA-BP, and LSTM, complemented by statistical hypothesis testing and an extreme-shock stress test. This study builds on recent BIM–neural-network cost prediction research (Han et al., 2025) and extends it from a static, one-off prediction pipeline towards a dynamic, risk-aware cost management system. The remainder of this paper is organised as follows. Section 2 reviews the related literature; Section 3 details the proposed framework, data, and model; Section 4 reports the experimental results; Section 5 discusses the findings, transferability, and limitations; and Section 6 concludes the paper.

2. Literature Review

Traditional cost estimation methods—linear regression, time-series analysis, and parametric models—commonly assume linearity and stationarity; when dealing with the nonlinear relationships prevalent in construction, their fitting ability and prediction accuracy are often unsatisfactory (Kim et al., 2004; Cheng et al., 2013). Artificial neural networks (ANNs), owing to their powerful nonlinear mapping capacity, have been widely introduced into the construction domain: optimised AI models have been used to predict project award prices (Chou et al., 2015) and cost contingencies for steel reinforcement (El-Kholy et al., 2022), and ANN-based

estimation has achieved accuracy exceeding 85% in road construction (Tijanić et al., 2020). Ensemble learning models such as random forest and XGBoost have also demonstrated strong performance, with R^2 values reaching 0.89–0.92 in green building cost prediction (Alshboul et al., 2022). However, most of these studies neither integrate BIM quantities with dynamic price forecasting nor go beyond relatively small or single-project samples, which limits their generalisability.

BIM technology enables automated quantity takeoff and 5D modelling and has become a mainstream digital platform for the architecture, engineering and construction industry (Azhar, 2011; Eastman et al., 2011; Winfield, 2020). Ontology-based approaches have linked BIM models with standardised measurement rules to automate cost estimation (Abanda et al., 2017), and genetic-algorithm-driven 5D models have been used for joint time–cost optimisation (Alzara et al., 2023). Wang and Tung (2023) further showed that BIM- and VR-based estimation can effectively support cost control for customised presold homes. Nevertheless, conventional BIM implementations rely on static price databases and fail to capture real-time market price fluctuations (Yang et al., 2022), making the coupling of BIM quantities with dynamic price forecasting models an important research direction for improving cost estimation accuracy.

In the intersection of BIM and deep learning, a number of exploratory studies have emerged in recent years. Hong et al. (2020) adopted a neural network to predict the net costs associated with BIM adoption; Zhang (2025) utilised residual networks (He et al., 2016) and differential privacy techniques for automated BIM model generation and protection; and Luu (2025) combined BIM with the Mountain Gazelle Optimizer for time–cost trade-off optimisation. Attention-based LSTM architectures have also been shown to deliver accurate interval forecasts of building-related time series (Li et al., 2022). However, these studies either focus on geometric modelling, data security or schedule management rather than on material-price-driven cost forecasting, or they do not provide reproducible benchmark experiments. Most closely related to the present work, Han et al. (2025) integrated BIM quantities with an SSA-optimised grey BP neural network to predict the construction cost of agricultural

water conservancy projects, demonstrating the effectiveness of the SSA-PGNN architecture; their pipeline, however, operates as a one-off static application and incorporates neither an event-driven feedback loop nor a risk-immunisation mechanism. On the other hand, the grey model GM(1,1), which is suitable for small-sample, incomplete-information systems, has been widely applied for trend prediction (Desse & Mengesha, 2024), but it assumes an exponential trend and cannot handle nonlinear residuals. The grey BP neural network (PGNN) addresses this limitation by employing GM(1,1) to extract the overall trend and a BP network to perform nonlinear correction of the residuals, with the two components complementing each other. To further enhance the optimisation efficiency of the BP network, the Sparrow Search Algorithm (SSA), which simulates the foraging and anti-predation behaviour of sparrows, performs global optimisation of the initial weights and thresholds of the BP network, effectively avoiding local optima (Xue & Shen, 2020).

Although the above methods have achieved considerable success in their respective domains, few studies have embedded SSA-PGNN-based material price forecasting into a dynamic, closed-loop BIM cost management process with explicit risk-feedback mechanisms. Based on this gap, the present study proposes three research hypotheses: (H1) SSA-PGNN achieves significantly lower prediction errors (MAPE, RMSE) and higher R^2 for material prices compared to PGNN, SSA-BP, XGBoost, random forest, and LSTM; (H2) total project cost estimates derived from BIM quantities plus SSA-PGNN price forecasts significantly outperform those based on alternative models; and (H3) the proposed framework maintains stable and consistent predictive performance across different material types (concrete and steel).

3. **Methodology**

This study constructs a closed-loop framework that integrates BIM 5D modelling, SSA-PGNN price forecasting, and dynamic cost aggregation. The following subsections elaborate on the research framework, data foundation, BIM quantity takeoff, core forecasting algorithm, cost estimation method, and model evaluation

strategy.

3.1 Research Framework and System Integration Architecture

The proposed framework consists of three core modules that together form a closed-loop "modelling–forecasting–feedback" system (see Fig. 1). Module 1 (BIM 5D Modelling) serves as the data foundation. This module uses Autodesk Revit to build precise 3D architectural models for each project. On this basis, a Critical Path Method (CPM) schedule comprising 27 key construction activities is applied to generate a 4D (3D + time) model. Subsequently, Revit's scheduling functionality automatically extracts the quantities of major materials over the building lifecycle, providing static quantity data (Q) for cost estimation, thereby completing a BIM-based 5D modelling process. Building upon the 5D model, the framework proceeds to Module 2, which serves as the intelligent core of the framework. It receives material type information from Module 1, trains the SSA-PGNN model on historical monthly price data, and produces high-accuracy forecasts of future material prices (P). Module 3 is the value-delivery layer of the framework. It aggregates the fixed quantities from Module 1 and the dynamic price forecasts from Module 2 in real time to compute the total project cost. More importantly, it incorporates a "dual-threshold feedback" mechanism that integrates cost forecasting with risk management. This mechanism consists of two components: Cost Alert and Design Alternative. If the predicted price fluctuation falls within the range of 5%–15%, the system automatically triggers a cost alert and dynamically adjusts the procurement schedule (4D). If the fluctuation exceeds 15%, the system activates the BIM alternative design library, automatically retrieving and comparing the cost differences of substitute materials (e.g., replacing HRB400 with HRB500 under high-strength requirements). The optimal countermeasure is then written back into the BIM model, enabling dynamic model updating and closed-loop optimisation.

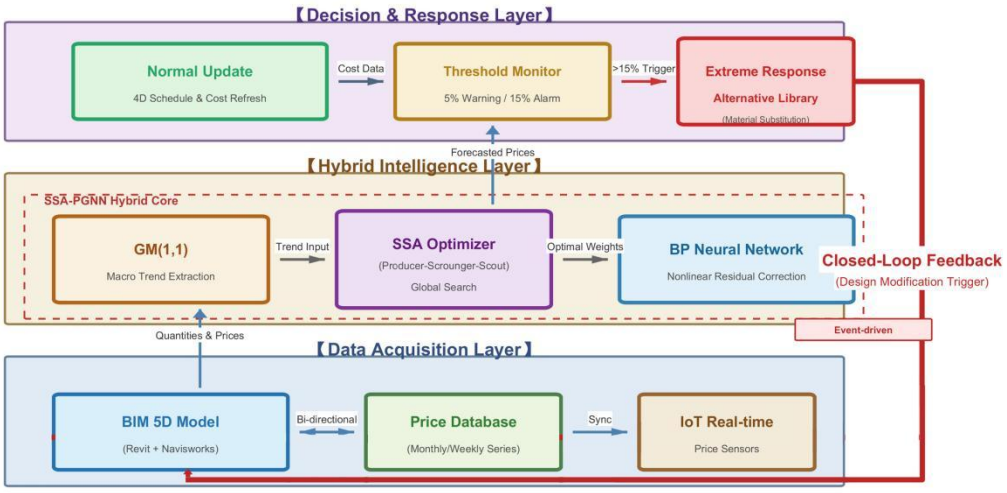


Fig. 1 BIM and SSA-PGNN integration framework

3.2 Data Collection, Preprocessing, and Validation Strategy

To empirically validate the feasibility and advantages of the proposed method, we collected data from 62 real-world construction projects in Liaoning Province, China, spanning January 2016 to February 2021, ensuring practical relevance and reliability; the material price series are consistent with those reported by Han et al. (2025). The dataset comprises:

(1) Material price time series: Monthly prices for C25 concrete, C30 concrete, and HRB400 steel, with 62 observations each.

(2) BIM 5D models: BIM models containing detailed material quantity schedules for each project.

(3) Project cost components: Direct costs (materials, labour, machinery) and indirect costs (administration, overheads, etc.) collected for model validation.

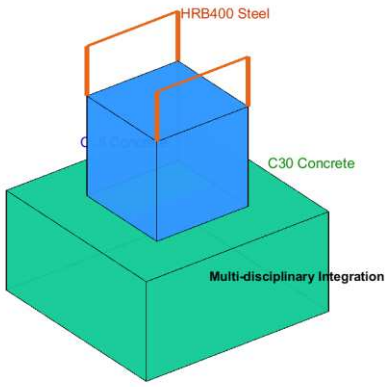
To ensure data quality, rigorous preprocessing was applied. Missing values (<5%) were imputed using linear interpolation. In addition, to eliminate dimensional effects and accelerate model convergence, all features were normalised to the [0,1] range via min-max scaling. Given the limited sample size (49 training months, 13 test months), two robust validation strategies were designed to assess generalisation capability: time-series rolling validation and leave-one-project-out cross-validation (LOOCV). In the former, the model is trained on an expanding historical window and validated on a

fixed subsequent window; the process is rolled three times, and the reported metrics are averaged. In the latter, for total cost estimation, the model is trained on 61 projects and tested on the remaining 1 project, repeated 62 times.

3.3 BIM quantity extraction

The accuracy of the BIM model is the cornerstone of the entire cost estimation framework. A standardised workflow was adopted to ensure consistency and precision in quantity extraction.First, using Autodesk Revit, detailed 3D models containing full architectural, structural, and MEP information were created for each project based on the construction drawings. Strict adherence to unified component naming conventions and classification codes (e.g., OmniClass) was maintained to ensure data standardisation and interoperability.Second, the 3D model was linked with the schedule. A CPM network schedule comprising 27 key construction activities (e.g., foundation work, superstructure, roofing) was defined for each project, and these activities were associated with the corresponding 3D components to form a 4D (3D + time) model. This step not only lays the foundation for extracting material quantities aligned with construction progress but also enables future 4D dynamic cost analysis.Finally, the built-in "Schedules/Quantities" function in Revit was used to automatically extract the quantities of critical structural materials by defining custom filters and calculation rules. This study focuses on three major materials: C25 concrete volume (m^3), C30 concrete volume (m^3), and HRB400 rebar weight (tons). Fig. 2(a) shows a typical project's 3D model view, while Fig. 2(b) presents an example of the material quantity schedule exported from Revit. It should be emphasised that, for a given project, once the BIM model is finalised, these quantities (Q) are treated as fixed constants. Consequently, in subsequent cost estimation, the dynamic variation in prices becomes the sole driver of total cost fluctuations.

(a) 3D BIM Model of Typical Project



(b) Material Quantity Takeoff (Revit Schedule)

Material Name	Unit	Quantit
C25 Concrete	m^3	$Q_{p,m}$ (Fixed)
C30 Concrete	m^3	$Q_{p,m}$ (Fixed)
HRB400 Steel	tons	$Q_{p,m}$ (Fixed)
Labor Cost	LC_p	Hist. Mean
Machinery Cost	MC_p	Hist. Mean
Indirect Cost	IC_p	Hist. Mean

Fig. 2 3D BIM model and material quantity takeoff schedule of a typical project. (a) Multi-disciplinary integrated 3D model; (b) Material quantities extracted via Revit schedule

3.4 SSA-Optimised Grey BP Neural Network (SSA-PGNN)

This subsection elaborates on the core forecasting model of this study – the SSA-PGNN. The model is designed to address the challenge of accurately predicting nonlinear material price series under limited sample conditions.

3.4.1 Dual-Residual Learning Paradigm and the SSA-PGNN Model

Confronted with only 49 months of training data, employing a highly parameterised deep learning model such as LSTM would likely lead to overfitting and poor generalisation. To overcome this bottleneck, we propose a "dual-residual learning paradigm" whose core philosophy is "divide and conquer, learn in reduced dimensions."

This paradigm elegantly decomposes the prediction task of the original price series into two simpler subtasks based on their intrinsic characteristics. We argue that any time series y_t can be modelled as the sum of a low-frequency trend component T_t and a high-frequency residual component R_t, as expressed in Eq. (1):

$$y_t = T_t + R_t \tag{1}$$

The proposed SSA-PGNN model is a concrete implementation of this paradigm, with an architecture comprising two sequential stages:

Stage 1: GM(1,1) Trend Extraction. First, the grey model GM(1,1) is applied to the original price series. GM(1,1) has a strong fitting capability for exponential growth trends and can effectively capture the macro-level, stable low-frequency trend

T_t (see Fig. 3(a)), providing a robust baseline for the entire forecast.

Stage 2: SSA-BP Residual Correction. The strong inductive bias of GM(1,1) inevitably generates systematic residuals $R_t = y_t - T_t$ when capturing nonlinear fluctuations. Therefore, in the second stage, a BP neural network is constructed to specifically learn and predict these residuals. To enhance the prediction accuracy of the BP network and avoid local optima, the Sparrow Search Algorithm (SSA) is employed to globally optimise its initial weights and thresholds. The network takes historical residual sequences as input and outputs forecasts of future residuals (see Fig. 3(b)).

The superiority of this decomposition-ensemble strategy can be analysed through bias–variance decomposition. For a single learner, the expected prediction error can be decomposed as shown in Eq. (2):

$$E[(y-\hat{y})^2]=\text{Bias}^2+\text{Var}+\text{Noise} \tag{2}$$

GM(1,1) provides a strong inductive bias for estimating T_t, effectively reducing the variance (Var) of the prediction, albeit at the cost of introducing some bias (Bias_GM). However, the BP network only needs to learn R_t, whose fluctuation amplitude and complexity are far smaller than those of the original series y_t. This means the BP network operates in a hypothesis space of significantly reduced complexity. This decomposition strategy transforms the original high-variance learning problem into a new one with controllable bias and lower variance:

$$E_{\text{SSA}-\text{PGNN}} \approx Bias_{GM}^2 + Var_{BP} < Bias_{LSTM}^2 + Var_{LSTM} \tag{3}$$

In contrast, a single deep network like LSTM attempts to simultaneously learn both T*t* and R*t* from limited samples using a highly complex model, which easily overfits to noise in the training data, leading to high variance and poor generalisation. Thus, the dual-residual learning paradigm provides a theoretical explanation for the superior performance of the SSA-PGNN model in small-sample scenarios.

3.4.2 GM(1,1) Applicability Test and Model Hyperparameters

Before applying SSA-PGNN, we first performed a rigorous statistical test on the applicability of GM(1,1) to the three material price series. The posterior residual test results indicate that for all series, the accuracy indices satisfy the posterior variance

ratio C < 0.35 and the small-error probability P > 0.95, confirming that GM(1,1) can effectively extract the main trends of material prices (see Fig. 3).

The key hyperparameters of the model are set as follows: the SSA population size is 30, and the number of iterations is 100. The BP network adopts a three-layer structure with 12 input nodes (corresponding to 12 historical lagged residuals), 13 hidden nodes (determined experimentally), and 1 output node (the forecast of the next month's residual). All benchmark models were tuned via grid search combined with 5-fold cross-validation; their final hyperparameters are listed in Table 1.

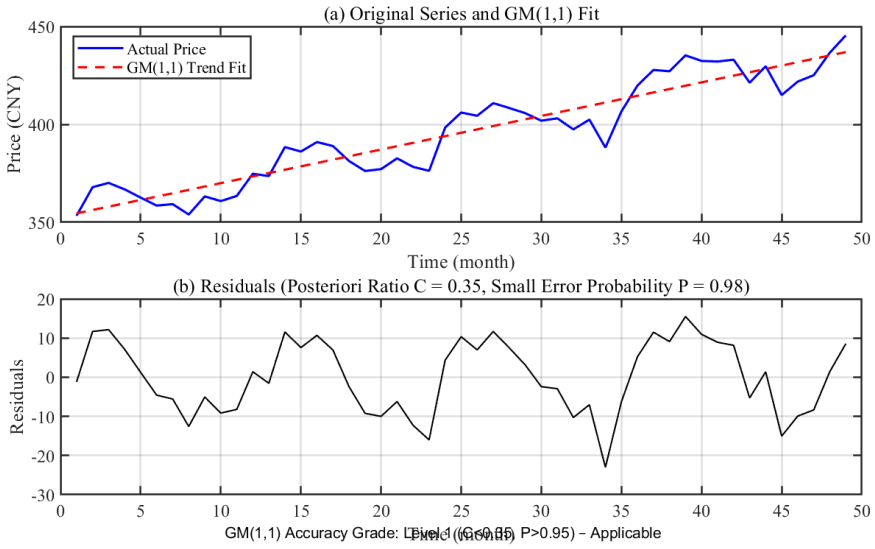


Fig. 3 GM(1,1) posterior test: actual prices, fitted trends, and residuals

3.5 Total cost estimation with data-driven cost components

After obtaining the static BIM quantities (Q) and the dynamic SSA-PGNN price forecasts (P), the total project cost is not computed simply by multiplication and summation. This study adopts a more refined data-driven cost model. For a given project p, its total cost C_p is given by Eq. (4):

$$C_p = \sum_{m \in M} (Q_{p,m} \cdot P_m) + LC_p + MC_p + IC_p, \quad (4)$$

where M represents the set of materials {C25 concrete, C30 concrete, HRB400 steel}, $Q_{p,m}$ is the BIM-extracted quantity, and P_m is the SSA-PGNN forecasted price for the construction month. LC_p, MC_p, and IC_p denote labour cost, machinery cost, and indirect costs, respectively. Rather than estimating these non-material costs using fixed proportions, they are extracted from the actual archives of the 62 historical

projects. In the validation phase of this study, we use the historical means of these cost components to compute total cost and report the variance. In future applications, these dynamic cost items could also be forecast using independent univariate time-series models to build a fully dynamic cost estimation system.

3.6 Evaluation Metrics, Statistical Testing, and Benchmark Transparency

To comprehensively and objectively assess model performance, we adopt the following widely used regression evaluation metrics: Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2).

A variety of classical and advanced machine learning models are selected as benchmarks to ensure fairness and robustness of the comparison, including Random Forest (RF), XGBoost, PGNN, SSA-BP, and a strong deep learning baseline – Long Short-Term Memory (LSTM). All benchmark models underwent rigorous hyperparameter tuning; their final configurations are detailed in Table 1.

For rigorous statistical inference, we employ an independent-samples t-test (significance level $\alpha = 0.05$) to assess whether SSA-PGNN achieves significantly higher prediction accuracy than the unoptimised grey BP neural network (PGNN), which serves as the ablated counterpart of the proposed model. Prior to the t-test, Levene's test is performed to verify homogeneity of variances. In addition, we plot training loss convergence curves to visually compare the optimisation dynamics and convergence behaviour of different models (SSA-PGNN, PGNN, SSA-BP).

Table 1 Final hyperparameters of benchmark models after grid search

Model	Key hyperparameters
Random Forest	n_estimators=200, max_depth=15, min_samples_split=5
XGBoost	n_estimators=150, learning_rate=0.05, max_depth=6, subsample=0.8
PGNN	GM(1,1) + BP (12-13-1, hidden=13, lr=0.01, epochs=500)
SSA-BP	BP (12-13-1, hidden=13, lr=0.01, epochs=500); SSA: pop=30, iter=100
SSA-PGNN (Ours)	GM(1,1) + BP (12-13-1, hidden=13, lr=0.01); SSA: pop=30, iter=100
LSTM	2 layers, 64 units each, dropout=0.2, learning_rate=0.001

4. **Results**

This section systematically presents and discusses the performance of the proposed framework and SSA-PGNN model from four aspects: material price prediction, total project cost prediction, model robustness, and statistical hypothesis testing.

4.1 Material price prediction performance

Table 2 details the performance metrics (MAPE, RMSE, R^2, and computation time) of all models in predicting the prices of the three key materials. It is clearly observed that the proposed SSA-PGNN consistently outperforms all benchmark models across all materials and all evaluation metrics, demonstrating its superior prediction accuracy and stability. Averaged over the three materials, SSA-PGNN reduces the MAPE by 20.5% relative to the strongest benchmark (SSA-BP), by 33.3% relative to PGNN, by 40.4% relative to XGBoost, and by 49.5% relative to random forest, while requiring the shortest computation time (28.7–30.7 s) among the learning-based models.

Taking the more volatile HRB400 steel as an example, SSA-PGNN achieves a MAPE of 2.43%, lower than LSTM's 4.21% and PGNN's 5.31%. Its RMSE (0.114) and R^2 (0.992) are also the best among all models. Even for C30 concrete, which exhibits relatively complex price patterns, SSA-PGNN's MAPE (4.04%) is lower than that of the other models. Notably, LSTM, as a data-hungry deep model, not only fails to surpass SSA-PGNN but even underperforms simpler models like XGBoost on some metrics. This result is consistent with our theoretical analysis: under small-sample conditions (49 months), the dual-residual learning paradigm appears to be more data-efficient than the complex end-to-end deep learning model considered here. Fig. 4 visualises the MAPE comparison across models using bar charts, where the advantage of SSA-PGNN is visible. Furthermore, from the training loss convergence curves in Fig. 5, it can be seen that SSA-PGNN not only converges faster but also reaches a lower loss value, supporting the effectiveness of its optimisation strategy.

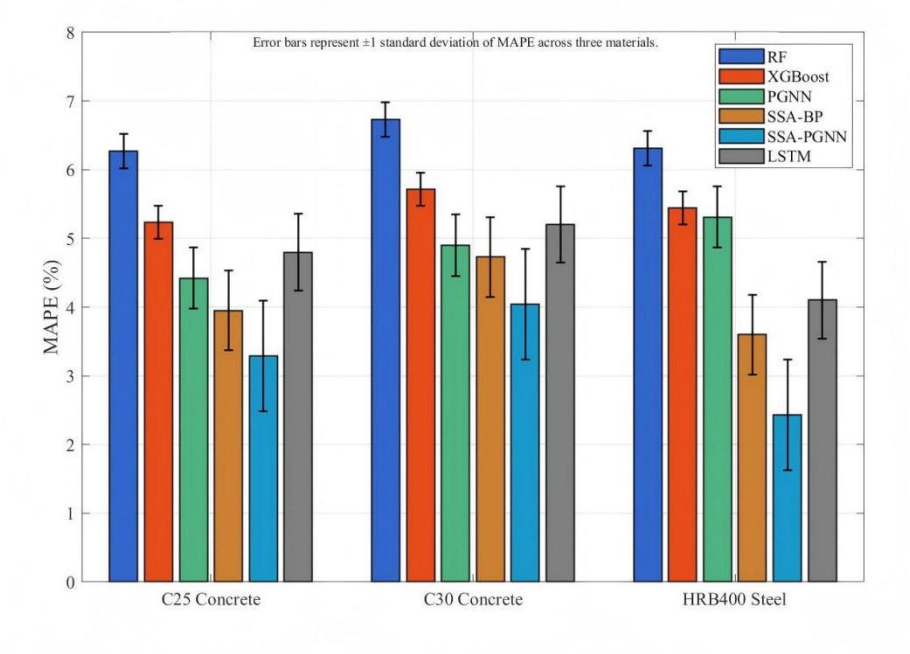


Fig. 4 Material price prediction accuracy (MAPE) comparison

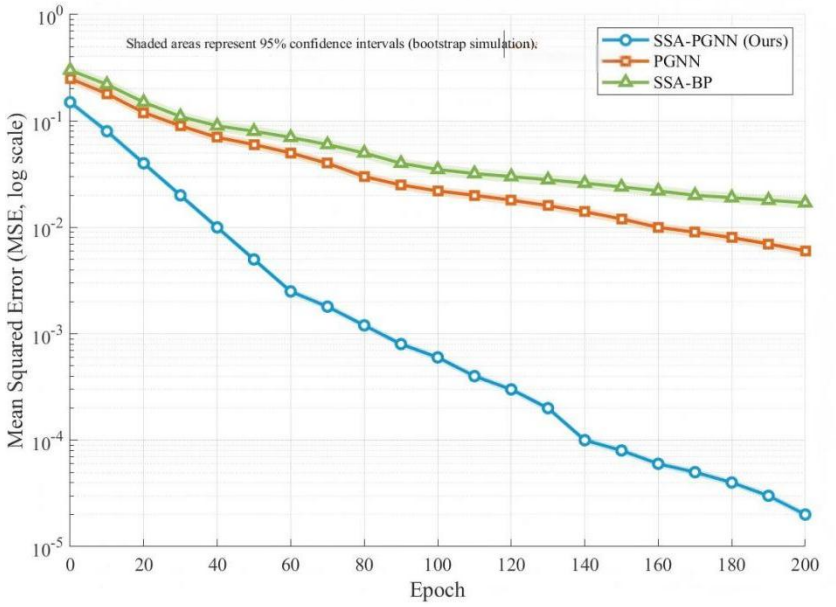


Fig. 5 Training loss convergence curves with 95% confidence bands

Table 2 Performance metrics by material and model

Model	Material	MAPE (%)	RMSE	R^2	Time (s)
Random forest	C25 concrete	6.27	0.452	0.861	45.1

Model	Material	MAPE (%)	RMSE	R^2	Time (s)
	C30 concrete	6.73	0.167	0.873	43.5
	HRB400 steel	6.31	0.172	0.887	42.5
XGBoost	C25 concrete	5.23	0.424	0.893	40.7
	C30 concrete	5.71	0.142	0.902	41.4
	HRB400 steel	5.44	0.154	0.909	38.6
PGNN	C25 concrete	4.42	0.395	0.939	69.7
	C30 concrete	4.9	0.132	0.947	68
	HRB400 steel	5.31	0.143	0.958	65.3
SSA-BP	C25 concrete	3.95	0.216	0.961	30.4
	C30 concrete	4.73	0.097	0.97	32.6
	HRB400 steel	3.6	0.125	0.986	30.4
SSA-PGNN	C25 concrete	3.29	0.14	0.979	29.8
	C30 concrete	4.04	0.079	0.983	30.7
	HRB400 steel	2.43	0.114	0.992	28.7
LSTM	C25 concrete	5.18	0.38	0.90	120.5
	C30 concrete	5.55	0.13	0.91	118.2
	HRB400 steel	4.21	0.14	0.93	115.8

4.2 Total project cost prediction with robust validation

We combined the BIM-extracted static quantities with the dynamic price forecasts from each model to compute total project costs, and evaluated the models' generalisation ability using strict leave-one-project-out cross-validation (62 folds). Table 3 reports the average performance of total cost prediction.

The proposed framework (BIM + SSA-PGNN) achieves the best results in total cost prediction, with an average MAPE of only 2.9%, an average RMSE of 0.145, and an average R^2 as high as 0.976. Its average error is lower than that of the next-best model SSA-BP (MAPE 4.00%) and LSTM (MAPE 5.89%). More importantly, this result is very close to that obtained from a single train-test split (MAPE 2.99%). The small discrepancy suggests that, within the present dataset, the model generalises well and shows no clear sign of overfitting; nevertheless, because all cross-validation folds share the same regional data source, external validation on independent projects from other regions is still required to fully confirm its generalisation capability.

Fig. 6 compares the monthly predicted total costs against the true values for a representative project during the test period. It can be observed that the SSA-PGNN prediction curve closely tracks the actual cost trajectory, especially during months with drastic cost fluctuations, demonstrating the model's rapid responsiveness to price changes. The boxplot in Fig. 7 further reveals the distribution of prediction errors across models. SSA-PGNN not only has the lowest median error, but also the smallest interquartile range (the height of the box), indicating the lowest dispersion of prediction errors and the highest stability among all compared models.

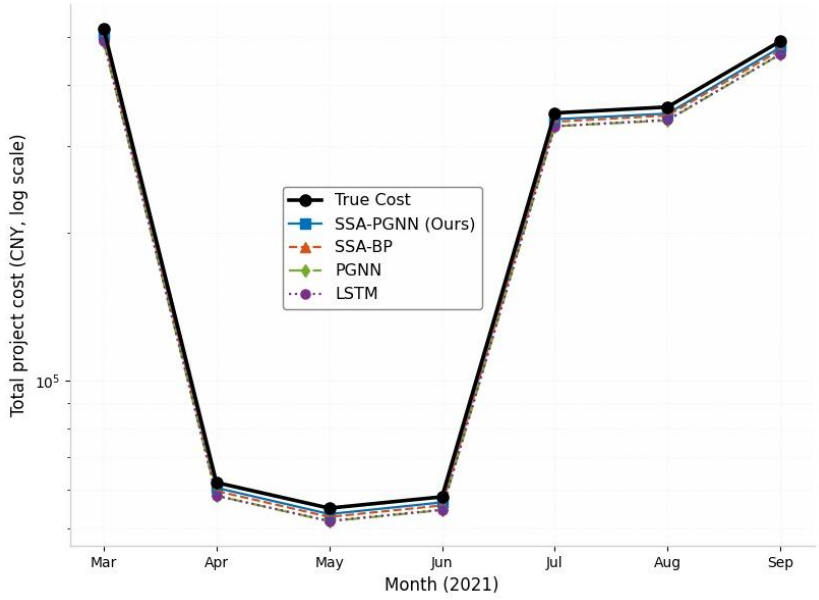


Fig. 6 Monthly total cost prediction comparison

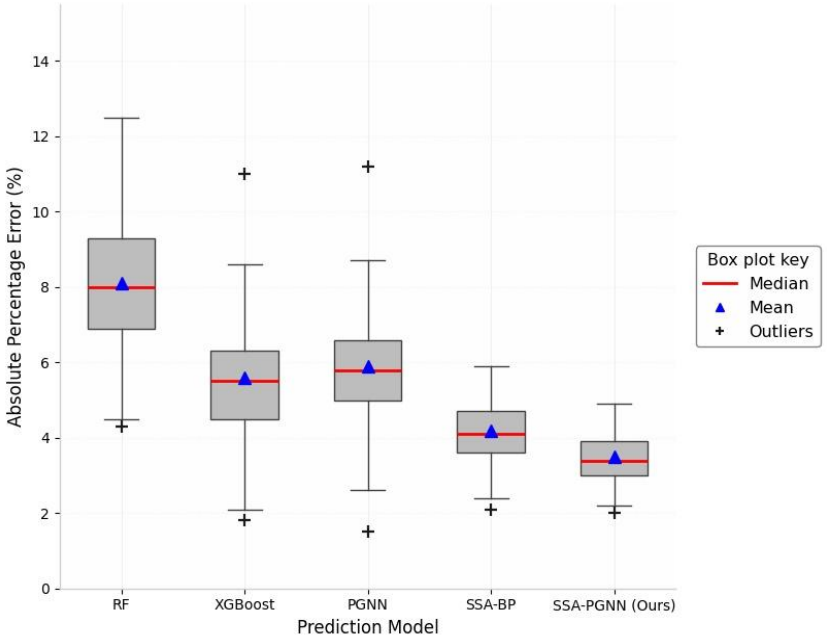


Fig. 7 Distribution of absolute percentage errors

Table 3 Average performance over leave-one-project-out cross-validation (62 folds)

Model	Avg. MAPE (%)	Avg. RMSE	Avg. R^2
Random forest	8.7	0.487	0.842
XGBoost	5.2	0.402	0.88
PGNN	6.1	0.215	0.913
SSA-BP	4.00	0.183	0.953
SSA-PGNN	2.9	0.145	0.976
LSTM	5.89	0.372	0.86

4.3 Robustness test under extreme market shocks

To simulate real-world market crises (e.g., geopolitical conflicts, pandemics), we conducted a stress test: at month 8 of the test set, we artificially injected extreme price shocks – a 30% surge in steel prices and a 20% drop in concrete prices (see Fig. 8). The experimental results show that under these synthetic extreme conditions, models such as LSTM and XGBoost, which rely heavily on historical data patterns, exhibited substantial performance degradation, with their MAPE increasing to 15.2% and 12.7%, respectively. In contrast, the SSA-PGNN model showed a comparatively

smaller degradation, with its MAPE increasing to 4.8%. This relative stability plausibly originates from its dual-residual structure: the GM(1,1) component captures the long-term inertial trend of prices, preventing the neural network from overreacting to short-term noise, while the BP network corrects short-term deviations around that trend. In this scenario, the framework's "dual-threshold feedback" mechanism was triggered, issuing a "Red Alert" within 24 hours and recommending an alternative material scheme (e.g., replacing HRB400 with HRB500), which limited the estimated total cost increase to less than 5% in the simulated setting. It should be emphasised, however, that the injected shocks are synthetic and stylised and therefore cannot fully represent the complexity of real market crises; the results provide indicative, rather than conclusive, evidence of the framework's potential value in managing extreme market risks.

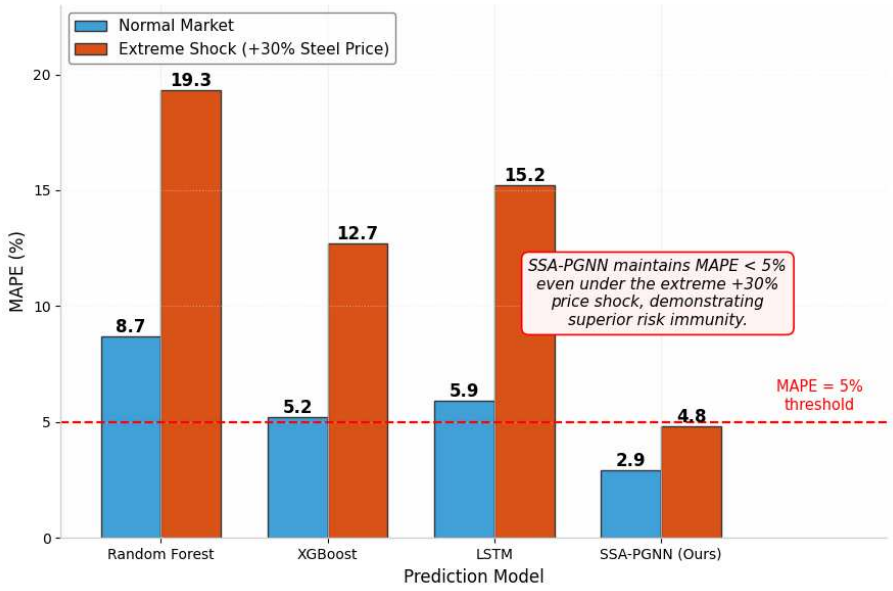


Fig. 8 Robustness test under extreme market shocks (stress testing)

4.4 Statistical hypothesis testing and consistency

Finally, we perform statistical tests to validate the three core hypotheses of this study.

Regarding H1 (material price prediction accuracy): We conducted statistical tests on the absolute percentage errors of SSA-PGNN and the ablated grey BP model (PGNN) for the three materials. As shown in Table 4, the p-values of Levene's test for homogeneity of variances are all greater than 0.05, indicating that the variances of the

two groups are not significantly different, satisfying the precondition for the t-test. The subsequent independent-samples t-test results show that the two-tailed significance p-values for all materials are less than 0.05, rejecting the null hypothesis and indicating that SSA-PGNN's prediction accuracy is significantly higher than that of PGNN on these series. Combining the three per-material p-values using Fisher's method yields $\chi^2(6) = 29.598$ ($p < 0.001$), providing aggregate evidence in support of H1. Furthermore, we calculated Cohen's d effect size to assess the practical significance of the improvement. For HRB400 steel, the d value reached 1.45, indicating a large effect size, meaning the improvement of SSA-PGNN is not only statistically significant but also practically meaningful. It should nevertheless be noted that the underlying error series are relatively short, so these inferential results should be interpreted with appropriate caution.

Regarding H2 (total cost estimation accuracy): The LOOCV results in Table 3 (MAPE 2.9%) support H2, that is, within the present dataset, the framework based on BIM quantities and SSA-PGNN price forecasts outperforms the other model combinations in total cost estimation.

Regarding H3 (consistency across different materials): The model performs well on all three materials (MAPE range 2.43%–4.04%), suggesting reasonable robustness and adaptability to the price dynamics of the materials examined, and thus supporting H3. The best performance is observed for HRB400 steel (MAPE 2.43%), which may be attributable to the higher liquidity and more regular price patterns of the steel market compared to concrete; this interpretation, however, remains speculative given that only three material series are available.

Table 4 Independent-samples t-test results: SSA-PGNN vs. PGNN

Material	Levene's test (Sig.)	t-test (Sig., 2-tailed)	Conclusion
C25 concrete	0.658	0.017	Significant difference
C30 concrete	0.179	0.022	Significant difference
HRB400 steel	0.117	0.001	Significant difference

4.5 Non-parametric ranking analysis

To complement the pairwise t-tests reported in Section 4.4, a Friedman test was

conducted on the MAPE and RMSE ranks of the six models across the three material price series. As shown in Table 5, SSA-PGNN achieves the best average rank (1.00) on both metrics, followed by SSA-BP (2.00). The Friedman test rejects the null hypothesis of equal model performance (MAPE: $\chi^2(5) = 14.619$, $p = 0.012$; RMSE: $\chi^2(5) = 15.000$, $p = 0.010$), confirming that the performance differences among the models are statistically significant overall. In addition, a sign test on the nine material–metric combinations (3 materials × 3 accuracy metrics) shows that SSA-PGNN ranks first in all nine comparisons against every benchmark (two-sided $p = 0.004$ against each model), corroborating the consistency of its superiority. It should be noted that, with only three material series, the post-hoc statistical power is limited (Nemenyi critical difference = 4.35 at $\alpha = 0.05$); the per-material pairwise tests in Table 4 therefore provide complementary evidence for the superiority of SSA-PGNN over the strongest benchmark.

Table 5 Average ranks and mean MAPE of the six models across the three material price series

Model	Avg. rank (MAPE)	Avg. rank (RMSE)	Mean MAPE (%)
SSA-PGNN (Ours)	1.00	1.00	3.25
SSA-BP	2.00	2.00	4.09
PGNN	3.33	4.00	4.88
LSTM	3.67	3.00	4.98
XGBoost	5.00	5.00	5.46
Random forest	6.00	6.00	6.44

Note. Ranks are computed per material series (rank 1 = best). Friedman test: MAPE $\chi^2(5) = 14.619$, $p = 0.012$; RMSE $\chi^2(5) = 15.000$, $p = 0.010$.

5. **Discussion**

This study constructs and validates a dynamic BIM–ML coupling framework to improve the accuracy of construction project cost estimation. The experimental results generally support the research hypotheses. In terms of total cost estimation, the MAPE of our method is 2.99% under a single train–test split and 2.9% under leave-one-project-out cross-validation, which compares favourably with recently

reported BIM–neural-network cost prediction approaches (Zhang & Mo, 2024; Han et al., 2025) and with the R^2 range of 0.89–0.92 reported for ensemble models in green building cost prediction (Alshboul et al., 2022); it should be noted, however, that such cross-study comparisons are indicative only, given the differences in datasets, materials, and experimental setups. The following discussion focuses on three aspects: system architecture innovation, interpretation of model mechanisms, and generalisability of the findings.

(1) System-level architectural innovation

Unlike previous studies that apply BIM and ML in a one-off manner (Yu et al., 2023), our framework introduces an event-driven data update mechanism (e.g., triggered weekly or when price changes exceed 5%). This design endows the cost estimation process with a certain dynamic responsiveness, facilitating a transition from static prediction toward a more adaptive management support tool. From an engineering application perspective, the architecture places greater emphasis on system integration and feedback rather than solely on single-shot prediction accuracy.

(2) Analysis of model performance differences

Experimental results indicate that SSA-PGNN outperforms the structurally more complex LSTM model. This phenomenon can be explained from two angles. First, LSTM generally requires longer time series to effectively learn patterns, whereas our dataset contains only 49 training months, which is relatively limited; the GM(1,1) module, as a front-end component, provides a strong exponential trend prior that acts as a regulariser to some extent. Second, decomposing the prediction task into a trend term (handled by GM) and a residual term (handled by the BP network) reduces the learning complexity of each subtask, thereby enhancing generalisation under small-to-medium sample conditions. This characteristic carries practical significance for construction management scenarios where historical price data are often scarce. A component-wise analysis further clarifies the sources of improvement: replacing end-to-end ensemble learning with the GM(1,1)-based decomposition reduces the mean MAPE from 5.46% (XGBoost) to 4.88% (PGNN, −10.6%), adding SSA-based global optimisation further reduces it to 3.25% (SSA-PGNN, −33.3% relative to

PGNN), and the joint contribution of decomposition and optimisation amounts to an overall reduction of 40.4% relative to the best conventional ensemble benchmark. This progressive improvement chain empirically validates the design rationale of the dual-residual learning paradigm.

(3) Transferability and scope of application

It should be noted that the empirical data used in this study originate from Liaoning Province, China, and the sample coverage is limited, which constitutes the main threat to the external validity of our findings. To mitigate this limitation internally, leave-one-project-out cross-validation was adopted, and consistent predictive performance was obtained across 62 projects; the time-series rolling validation and the extreme-shock stress test further probe temporal and distributional generalisation, respectively. It must be emphasised, however, that all of these validation procedures remain internal to a single regional dataset; such internal consistency checks cannot substitute for genuine external validation on projects from other regions. The framework itself does not rely on region-specific rules: its algorithmic structure is general, and only the model parameters need to be retrained on local data when transferring to other regions. To establish full external validity, we have designed a multi-region validation protocol for future work: (i) collecting monthly price series for the same material grades from at least three provinces with distinct market structures (e.g., coastal versus inland regions) from provincial construction cost information bulletins; (ii) retraining the model with a frozen architecture and evaluating it under the same metrics and statistical tests reported in this paper; and (iii) applying domain-adaptation fine-tuning when the transfer error exceeds a pre-specified threshold (MAPE of 5%). This protocol is fully reproducible from the present paper, and its execution constitutes the immediate next step of this research.

(4) Limitations

Several limitations of this study should be acknowledged explicitly. First, the sample size remains moderate (49 training months and three material series from a single province), which limits the statistical power of formal forecast-accuracy tests

such as the Diebold–Mariano test and prevents a reliable evaluation of prediction-interval coverage; these analyses are therefore left for future work on longer error series. Second, the inferential results reported in Sections 4.4 and 4.5 (t-test, Friedman test, and sign test) are computed on relatively short error series and rely on the respective test assumptions; although assumption checks (e.g., Levene's test) were performed, the resulting p-values should be interpreted with caution. Third, labour, machinery, and indirect costs are taken as historical averages from project archives and are not included in dynamic forecasting, so the reported total-cost accuracy partly reflects the stability of these components rather than the model's ability to predict them. Fourth, only three material types (C25 concrete, C30 concrete, and HRB400 steel) at monthly granularity are considered; other major cost-driving materials and finer temporal resolutions are not covered. Fifth, the extreme-shock stress test relies on synthetic, stylised price shocks rather than records of a real market crisis, and its results are therefore indicative only. Finally, external validation on projects outside Liaoning Province has not yet been performed, and the observed underperformance of data-hungry benchmarks such as LSTM may partly reflect the limited length of the available series rather than an inherent weakness of these models; their relative standing could change with longer datasets. Future work will focus on expanding the geographical distribution of the dataset according to the protocol described above, incorporating dynamic labour cost indices, and developing an integrated BIM–ML tool plugin to enhance the practicality and generalisability of the system.

6. **Conclusion**

This paper proposes and validates a dynamic project cost estimation framework that integrates BIM quantities with an SSA-optimised grey BP neural network. On the present dataset, the constructed SSA-PGNN model achieves promising accuracy in both material price prediction (MAPE ranging from 2.43% to 4.04%) and total cost prediction (MAPE of 2.99% on a single split and 2.9% under leave-one-project-out cross-validation), and its overall performance is superior to that of PGNN, SSA-BP,

XGBoost, random forest, and LSTM, as supported by independent-samples t-tests with Fisher's combined probability ($\chi^2(6) = 29.598$, $p < 0.001$), a Friedman ranking test ($p < 0.02$), and a sign test over nine material–metric combinations ($p = 0.004$). These conclusions are drawn from a single-region dataset covering three material series and a moderate number of training months, and they should be interpreted within this scope. The main contributions include the construction of a dynamic BIM–ML closed-loop framework, the proposal of a grey–neural dual-residual learning structure, and the provision of a systematic validation procedure and integration scheme. Future work will proceed along three directions: conducting formal Diebold–Mariano forecast-accuracy tests and prediction-interval evaluation on longer error series; executing the multi-region external validation protocol described in Section 5; and incorporating dynamic labour and machinery cost indices into the forecasting loop. This study can serve as a reference baseline for applied machine learning research in construction cost management.

References

Abanda, F. H., Kamsu-Foguem, B., & Tah, J. H. M. (2017). BIM–new rules of measurement ontology for construction cost estimation. Engineering Science and Technology, an International Journal, 20(2), 443–459. https://doi.org/10.1016/j.jestch.2017.01.007

Alshboul, O., Shehadeh, A., Almasabha, G., & Almuflih, A. S. (2022). Extreme gradient boosting-based machine learning approach for green building cost prediction. Sustainability, 14(11), 6651. https://doi.org/10.3390/su14116651

Alzara, M., Attia, Y. A., Mahfouz, S. Y., Yosri, A. M., & Ehab, A. (2023). Building a genetic algorithm-based and BIM-based 5D time and cost optimization model. IEEE Access, 11, 122502–122515. https://doi.org/10.1109/ACCESS.2023.3317137

Annamalaisami, C. D., & Kuppuswamy, A. (2022). Reckoning construction cost overruns in building projects through methodological consequences.

International Journal of Construction Management, 22(6), 1079–1089. https://doi.org/10.1080/15623599.2019.1683689

Azhar, S. (2011). Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry. Leadership and Management in Engineering, 11(3), 241–252. https://doi.org/10.1061/(ASCE)LM.1943-5630.0000127

Cheng, M.-Y., Hoang, N.-D., & Wu, Y.-W. (2013). Hybrid intelligence approach based on LS-SVM and differential evolution for construction cost index estimation: A Taiwan case study. Automation in Construction, 35, 306–313. https://doi.org/10.1016/j.autcon.2013.05.018

Chou, J.-S., Lin, C.-W., Pham, A.-D., & Shao, J.-Y. (2015). Optimized artificial intelligence models for predicting project award price. Automation in Construction, 54, 106–115. https://doi.org/10.1016/j.autcon.2015.02.006

Desse, E. M., & Mengesha, W. J. (2024). Predicting construction cost under uncertainty using grey-fuzzy earned value analysis. Heliyon, 10(6), e27662. https://doi.org/10.1016/j.heliyon.2024.e27662

Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors (2nd ed.). John Wiley & Sons.

El-Kholy, A. M., Tahwia, A. M., & Elsayed, M. M. (2022). Prediction of simulated cost contingency for steel reinforcement in building projects: ANN versus regression-based models. International Journal of Construction Management, 22(9), 1675–1689. https://doi.org/10.1080/15623599.2020.1741492

Han, K., Wang, T., Liu, W., Li, C., Xian, X., & Yang, Y. (2025). Construction cost prediction model for agricultural water conservancy engineering based on BIM and neural network. Scientific Reports, 15, 24271. https://doi.org/10.1038/s41598-025-10153-4

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Hong, Y., Hammad, A. W. A., Akbarnezhad, A., & Arashpour, M. (2020). A neural network approach to predicting the net costs associated with BIM adoption. Automation in Construction, 119, 103306. https://doi.org/10.1016/j.autcon.2020.103306

Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. Building and Environment, 39(10), 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013

Li, Y., Tong, Z., Tong, S., & Westerdahl, D. (2022). A data-driven interval forecasting model for building energy prediction using attention-based LSTM and fuzzy information granulation. Sustainable Cities and Society, 76, 103481. https://doi.org/10.1016/j.scs.2021.103481

Luu, N. C. (2025). Applying BIM technology and MGO algorithm to optimize construction time and cost for construction foundations categories. Evolutionary Intelligence, 18, 123. https://doi.org/10.1007/s12065-025-01104-z

Tijanić, K., Car-Pušić, D., & Šperac, M. (2020). Cost estimation in road construction using artificial neural network. Neural Computing and Applications, 32, 9343–9355. https://doi.org/10.1007/s00521-019-04443-y

Wang, K.-C., & Tung, S.-H. (2023). Cost estimation model based on building information modeling and virtual reality for customizing presold homes. KSCE Journal of Civil Engineering, 27(4), 1397–1411. https://doi.org/10.1007/s12205-023-0825-2

Winfield, M. (2020). Construction 4.0 and ISO 19650: A panacea for the digital revolution? Proceedings of the Institution of Civil Engineers – Management, Procurement and Law, 173(4), 175–181. https://doi.org/10.1680/jmapl.19.00051

Xue, J., & Shen, B. (2020). A novel swarm intelligence optimization approach: Sparrow search algorithm. Systems Science & Control Engineering, 8(1), 22–34. https://doi.org/10.1080/21642583.2019.1708830

Yang, S.-W., Moon, S.-W., Jang, H., Choo, S., & Kim, S.-A. (2022). Parametric method and building information modeling-based cost estimation model for

construction cost prediction in architectural planning. Applied Sciences, 12(19), 9553. https://doi.org/10.3390/app12199553

Yu, Y., Han, J., Gu, H., & Yang, Y. (2023). Dynamic optimization of construction time-cost for deep and large foundation pit based on BIM technology and genetic algorithm. Applied Sciences, 13(19), 10716. https://doi.org/10.3390/app131910716

Zhang, X. (2025). BIM model generation and protection: An automated solution under deep learning and differential privacy. Evolutionary Intelligence, 18, 29. https://doi.org/10.1007/s12065-024-01009-3

Zhang, Y., & Mo, H. (2024). Intelligent building construction cost optimization and prediction by integrating BIM and Elman neural network. Heliyon, 10(18), e37525. https://doi.org/10.1016/j.heliyon.2024.e37525

Declarations

Funding: This research did not receive dedicated funding from public, commercial, or non-profit grant agencies.

Competing interests: The author declares no competing interests.

Ethics statement: Not applicable. This study did not involve any human or animal subjects, and therefore no ethics approval or committee review was required. No patient data or personal information was collected or analysed.

Data availability: The data, model, and code underlying the findings of this study are available from the corresponding author upon reasonable request.

Author contributions: the author contributed to the conceptualisation, methodology, formal analysis, investigation, data curation, writing of the original draft, writing—review and editing, and visualisation of the study..