JURNAL TEKNIK SIPIL CENDEKIA
e-ISSN : 2723-3375
doi : **10.51988/jtsc.v7i3.563**
pp : 471-481

Application of Machine Learning in Predicting Project Cost Overruns: A Review

Johanis Anggawan, Agustinus Purna Irawan[2], Mega Waty[3]
[1,3] Fakultas Teknik Program Studi Doktor Ilmu Teknik Sipil, Universitas Tarumanagara, Kota Jakarta Barat, DKI Jakarta, Indonesia
[2]School of STEM, Universitas Prasetiya Mulya, Jakarta, Indonesia, Jakarta, Indonesia
*) johanis.328211004@stu.untar.ac.id

Received : 4 Februari 2026 Revised : 11 Juli 2026 Accepted : 20 Juli 2026

Abstract

Cost overrun is one of the main problems in project management that has a significant impact on the success of construction projects. High project complexity, uncertainty of field conditions, design changes, and external factors such as inflation and supply chain disruptions make predicting cost overruns a challenge that requires a more sophisticated analytical approach than conventional methods. The development of Machine Learning (ML) technology has opened new opportunities to improve the accuracy of project cost predictions through the use of historical data and complex pattern analysis. This study aims to review the development of research related to the application of Machine Learning in predicting project cost overruns, identify the most widely used algorithms, influential predictor variables, and evaluate the advantages and limitations of each approach. The research method is carried out through a systematic literature review of scientific publications indexed by Scopus in the period 2015–2025. The study focuses on various ML algorithms, including Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Gradient Boosting, and Extreme Gradient Boosting (XGBoost), which are used in project cost prediction. The study results show that Random Forest and XGBoost to produce higher levels of accuracy than traditional statistical methods or single ML models. The factors most frequently used as predictor variables include project characteristics, implementation duration, changes in work scope, contract conditions, labor productivity, technical risks, and macroeconomic factors. This study concludes that Machine Learning has great potential in supporting early warning systems and data-driven decision-making to control the risk of project cost overruns.

Keywords: *cost overrun, machine learning, project cost prediction, construction, systematic literature review, project management.*

INTRODUCTION

Construction projects play a strategic role in supporting economic growth, infrastructure development, and achieving sustainable development goals. Project success is generally measured by achieving three main objectives: cost, time, and quality, known as the iron triangle in project management [1-2]. However, cost overruns remain one of the most common problems in construction projects across various countries. This phenomenon has attracted significant attention from researchers and practitioners due to its high frequency and significant impact on project performance [3-4]. Cost overruns not only cause financial losses for project owners and contractors, but can also lead to delayed project completion, reduced work quality, disrupted project cash flow, and reduced investor and stakeholder confidence in the project's success [5-6] Various studies show that cost overruns are influenced by a combination of technical, managerial, contractual, economic, and environmental factors that interact in a complex manner, making them difficult to predict using conventional approaches[7-9].

Various studies have shown that most construction projects experience cost deviations from the initial budget. This phenomenon occurs in both small-scale projects and mega-infrastructure projects. Cost overruns have been identified as a major cause of project failure in various countries and industrial sectors, particularly in construction and transportation infrastructure projects [3],[5]. The factors causing cost overruns are diverse, ranging from scope changes, initial estimation errors, implementation delays, low labor productivity, material price increases, unpredictable weather conditions, and external factors such as inflation and economic instability. Various literature reviews have identified design changes, scope changes, project delays, poor project management, cost estimation errors, material price fluctuations, and economic factors as the dominant causes of construction project cost overruns [10-12]. The complexity of interactions between these factors makes predicting cost overruns increasingly difficult using conventional approaches, which generally assume linear relationships between variables. Traditionally, project cost predictions have been conducted using statistical methods and expert judgment-based estimation techniques, such as linear regression, historical trend analysis, and parametric approaches. While these methods are relatively easy to implement, their predictive capabilities are often limited when dealing with project data that is nonlinear, multidimensional, and subject to high levels of uncertainty. Conventional methods generally assume linear relationships between variables, making them less able to capture the complex patterns often found in modern project data.

The development of digital technology and the transformation of the industry towards the industry 4.0 era have resulted in the availability of large amounts of project data (big data) from various sources, such as project management information systems, Building Information Modeling (BIM), field sensors, work progress reports, and project financial data. The digitalization of construction processes and the integration of Information technology enables the real-time collection, storage, and analysis of project data on an increasingly large and complex scale [13-14]. This condition encourages the development of the application of artificial intelligence (AI), and specifically machine learning (ML), in project management. Machine learning is a branch of artificial intelligence that enables computer systems to learn patterns from historical data and generate predictions without having to be explicitly programmed. In the context of project management, this technology has been used to support decision-making, cost prediction, duration estimation, risk identification, and project performance evaluation with a higher level of accuracy than conventional methods [15-17]. In the past decade, the application of machine learning to predict project cost overruns has experienced significant development. Various algorithms have been used, such as artificial neural networks (ANN), support vector machines (SVM), decision trees (DT), random forests (RF), gradient boosting machines (GBM), extreme gradient boosting (XGBoost), and deep learning techniques. Unlike traditional statistical methods, machine learning algorithms are able to identify complex relationships, hidden patterns, and nonlinear interactions between variables that influence project costs. Various studies have shown that machine learning models are capable of producing higher levels of predictive accuracy than conventional approaches, especially for projects with complex and dynamic characteristics.

Research developments in this area still face several challenges. First, there is no consensus on the most effective machine learning algorithm for predicting cost overruns across various project types because model performance is heavily influenced by data characteristics, sample size, and the project context being analyzed [15],[18]. Second, variations in data characteristics, predictor indicators, and validation methods used in various studies often make the results difficult to compare directly and limit the generalizability of the models to different project conditions [14],[17]. Third, most machine learning models, particularly those based on ensemble learning and deep learning, still face interpretability issues (black box problems), making the relationship between input variables and predicted results difficult to explain transparently to decision-makers [19-20]. In addition, data quality issues, incomplete data, dataset imbalance, and limited historical data availability are also major challenges in developing reliable, accurate, and generalizable predictive models for various types of construction and infrastructure projects [13-14],[16].

Previous research has discussed the application of Machine Learning in project management, particularly for cost estimation, duration prediction, risk identification, and project performance evaluation. However, most studies still focus on the development and testing of specific algorithms, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), on specific project case studies and limited datasets [14-15],[18]. On the other hand, various reviews on artificial intelligence in the construction industry generally discuss AI applications broadly without providing an in-depth analysis of project cost overrun prediction as a specific research topic [16-17]. Comprehensive studies evaluating algorithm developments, dominant predictor variables, model performance, implementation challenges, and future research directions in the context of project cost overrun prediction are still relatively limited. A systematic literature review method is necessary to integrate existing research findings, compare the effectiveness of various Machine Learning approaches, and identify research gaps that can serve as a basis for further research.
This research aims to analyze the development of Machine Learning applications in predicting project cost overruns based on Scopus-indexed scientific publications. Specifically, this research aims to: (1) identify research trends related to the use of Machine Learning in project cost overrun prediction; (2) examine the most widely used algorithms and their performance levels; (3) identify the factors most influential in cost overruns; and (4) formulate challenges, opportunities, and directions for future research development. These objectives are crucial given the increasing use of artificial intelligence in project management and the need for more accurate, adaptive prediction systems capable of supporting data-driven decision-making [14],[16-17]. The results of this study are expected to provide theoretical contributions to the development of project management science and serve as a practical reference for project owners, contractors, consultants, and policymakers in implementing artificial intelligence-based cost prediction systems to improve project success and reduce the risk of cost overruns [13-14].

Unlike previous review studies, which generally discuss artificial intelligence applications in construction management from a broad perspective or evaluate only specific machine learning techniques, this study specifically focuses on the application of Machine Learning for predicting construction project cost overruns using a Systematic Literature Review (SLR) approach. The novelty of this study lies in its comprehensive synthesis of recent Scopus-indexed publications by simultaneously examining research trends, comparing the predictive performance of widely used machine learning algorithms (ANN, SVM, Random Forest, XGBoost, and Deep Learning), identifying the most influential predictor variables, and highlighting current implementation challenges and future research directions. Furthermore, this review emphasizes emerging opportunities for integrating Explainable Artificial Intelligence (XAI), Building Information Modeling (BIM), Digital Twins, and Internet of Things (IoT) technologies into intelligent cost prediction systems. Therefore, this study provides a more comprehensive conceptual framework that supports both future academic research and practical decision-making in construction project cost management.

METHOD

This study used the Systematic Literature Review (SLR) method to identify, evaluate, and synthesize research findings related to the application of Machine Learning to predict project cost overruns. The SLR approach was chosen because it provides a systematic, transparent, replicable review process, and minimizes bias in literature selection and analysis. This method is widely used to map research developments, identify scientific trends, and identify research gaps in a field (Kitchenham & Charters, 2007; Tranfield et al., 2003).

The study was conducted through several main stages: (1) formulating research questions, (2) determining a literature search strategy, (3) selecting articles based on inclusion and exclusion criteria, (4) data extraction, (5) bibliometric analysis and thematic synthesis, and (6) identifying research gaps and future research agendas.
This study was designed to answer the following research questions (RQs):

- RQ1: What are the trends in research developments related to the application of Machine Learning to predict project cost overruns?
- RQ2: What machine learning algorithms are most commonly used to predict project cost overruns?
- RQ3: What are the challenges, opportunities, and future research directions in developing project cost overrun prediction models?

The primary data source used in this study was the Scopus database, as it is one of the largest scientific publication indexes, covering highly reputable international journals with broad multidisciplinary coverage.
The article search was conducted using a combination of keywords structured around the main research concepts: machine learning, cost overrun, and project management. The search string used was as follows:

TITLE-ABS-KEY (
("machine learning" OR "artificial intelligence"
OR
"deep learning" OR "neural network" OR
"random forest" OR "support vector machine")
AND
("cost overrun" OR "cost prediction" OR
"cost estimation" OR "project cost")
AND
("construction project" OR "project management"
OR "infrastructure project")

A total of 256 articles were initially identified through a Scopus database search using the predefined search string. After removing duplicate records, 232 articles remained for title and abstract screening. Subsequently, 78 articles were assessed through full-text review to determine their relevance to the research objectives. Following the application of the inclusion and exclusion criteria, 43 articles were finally selected and included in the systematic literature review. The detailed article selection process is illustrated in Figure 1 following the PRISMA framework.

Inclusion and Exclusion Criteria:
The literature search was limited to publications indexed in the Scopus database between 2015 and 2025, consistent with the study scope presented in the abstract. Only peer-reviewed journal articles written in English and focusing on the application of Machine Learning for predicting construction project cost overruns or construction cost estimation were included in the review. Conference papers, book chapters, review articles unrelated to construction cost prediction, studies without sufficient methodological information, and publications outside the specified period were excluded. These inclusion and exclusion criteria ensured that only relevant and high-quality studies were synthesized in this review.

Data Extraction:
For each selected article, the extracted information included publication year, machine learning algorithm, predictor variables, project type, dataset characteristics, model performance indicators, and the main research findings. These data were subsequently synthesized to identify research

trends, compare algorithm performance, and determine future research opportunities.

The article selection process follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

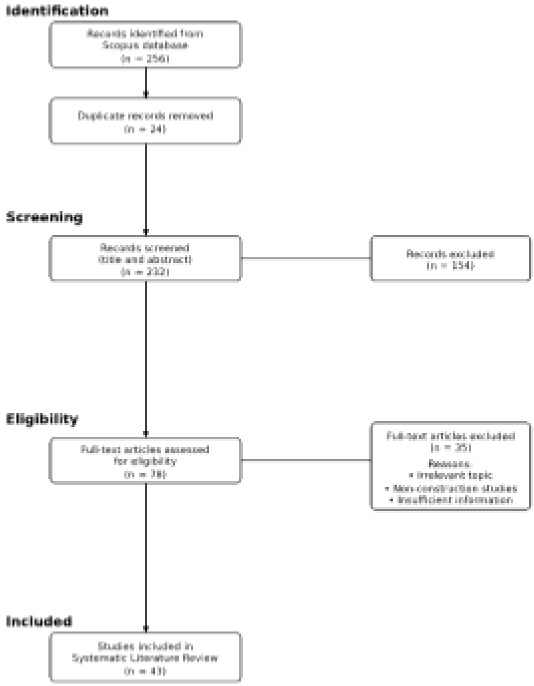

Figure 1. PRISMA 2020 Flow Diagram of Article Selection Process

Figure 1 presents the PRISMA-based article selection process adopted in this study. A total of 256 records were initially identified through the Scopus database. After removing 24 duplicate records, 232 articles remained for title and abstract screening. Subsequently, 78 full-text articles were assessed for eligibility, of which 35 articles were excluded based on the predefined exclusion criteria. Finally, 43 articles met all inclusion criteria and were included in the qualitative synthesis.

The selection stages consist of:

1) Identification: identifying articles from the Scopus database based on predetermined keywords.
2) Screening: removing duplicate articles, Selecting titles and abstracts.
3) Eligibility: reviewing the full text of articles that meet the initial criteria.
4) Inclusion: determining the articles to be analyzed in the study.

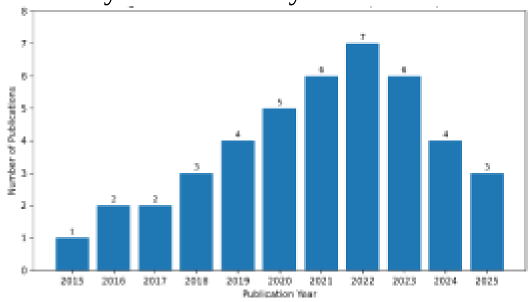

Figure 2. Annual Distribution of Publications (2015–2025)

RESULTS AND DISCUSSION

1. Artificial Neural Network (ANN)

   Publication analysis shows that research on project cost prediction using Machine Learning has consistently increased over the past decade. In the 2015–2018 period, research was still dominated by Artificial Neural Network (ANN) and Support Vector Machine (SVM) methods. Entering the 2019–2022 period, ensemble learning-based algorithms such as Random Forest (RF), Gradient Boosting Machine (GBM), and Extreme Gradient Boosting (XGBoost) began to be widely used due to their superior ability to handle nonlinear relationships and high-dimensional data.

   In the 2015–2025 period, research trends began to shift toward the use of Deep Learning, Explainable Artificial Intelligence (XAI), and the integration of Machine Learning with Building Information Modeling (BIM), the Internet of Things (IoT), and real-time project monitoring systems. These developments indicate that the research focus is no longer solely on improving prediction accuracy, but also on model transparency and the integration of intelligent systems in project management.

   The study results show that several Machine Learning algorithms dominate research on project cost overrun prediction.

   ANN is the earliest and most widely used algorithm in project cost prediction. Its ability to model nonlinear relationships makes it superior to traditional regression. However, ANN has the drawback of requiring large amounts of data and poor model interpretability. The extensive use of ANN in earlier studies demonstrates its capability to model complex nonlinear relationships between project variables and construction costs. However, its dependence on large training datasets and limited interpretability have encouraged researchers to explore alternative algorithms that provide greater transparency while maintaining high predictive accuracy.

2. Support Vector Machine (SVM)
   SVM performs well on small to medium-sized datasets. This algorithm is relatively resistant to overfitting and is capable of producing accurate predictions for various types of construction projects. SVM remains an effective solution for relatively small datasets because of its strong generalization capability and resistance to overfitting. Nevertheless, its performance may decrease when handling highly heterogeneous project data involving numerous predictor variables, making ensemble-based algorithms increasingly preferable in recent studies.

3. Random Forest (RF)
   RF has become one of the most popular algorithms in the last five years because it is able to handle complex data, reduce model variance, and provide information about the level of variable importance (feature importance). The increasing popularity of Random Forest reflects the growing demand for prediction models capable of handling nonlinear interactions among technical, managerial, and economic variables. In addition to providing high predictive accuracy, Random Forest offers feature importance analysis, allowing practitioners to identify the most influential factors contributing to project cost overruns. This practical advantage distinguishes Random Forest from many earlier machine learning techniques.

4. Extreme Gradient Boosting (XGBoost)
   XGBoost has demonstrated superior performance in numerous studies due to its ability to improve prediction accuracy and handle imbalanced data. This algorithm is one of the most promising methods for predicting project cost overruns. XGBoost has consistently demonstrated superior predictive performance because of its ability to optimize learning through gradient boosting while effectively controlling overfitting. Several reviewed studies reported lower prediction errors than ANN and SVM, indicating that XGBoost is particularly suitable for complex construction datasets. However, the algorithm requires careful hyperparameter tuning and remains relatively difficult to interpret, highlighting the need for Explainable Artificial Intelligence (XAI) in future implementations.

Table 1. Comparison of ML Algorithms [14-19]

Algoritma	Advantages	Disadvantages
ANN	Handling complex nonlinear relationships	Black-box, requires big data
SVM	Accurate on small datasets	Difficult to interpret
RF	Stable and easy to use	Requires parameter tuning
XGBoost	High accuracy	High model complexity
Deep Learning	Handling big and complex data	Requires high computing

Table 1 illustrates the comparison between ANN, SVM, RF, XGboost and Deep Learning where XGBoost has high accuracy with high model complexity.

Although Deep Learning provides excellent predictive capability for large-scale datasets, its practical implementation in construction project management remains limited. Many organizations still lack sufficient historical project data and computational resources to fully exploit deep learning models. Consequently, future studies should investigate hybrid approaches that combine predictive accuracy with model transparency and practical applicability.

5. Research Contribution

This study contributes to the existing literature in three important aspects. First, it provides a comprehensive synthesis of Machine Learning applications specifically for construction project cost overrun prediction based on recent Scopus-indexed publications. Second, it compares the strengths, limitations, and practical applicability of the major machine learning algorithms, enabling researchers and practitioners to better understand algorithm selection under different project conditions. Third, this study identifies emerging research opportunities involving Explainable Artificial Intelligence (XAI), Building Information Modeling (BIM), Digital Twins, and Internet of Things (IoT), which are expected to support the development of more transparent, intelligent, and reliable project cost prediction systems.

Table 2. SLR of various algorithms

Algorithm	**Reference**	**Main Predictor Variables**	**Performance Summary**
Artificial Neural Network (ANN)	Kim et al. (2004) [21]	Historical project cost, work volume, labor productivity	Demonstrated higher prediction accuracy than conventional regression models in estimating construction costs.
Support Vector Machine (SVM)	Son et al. (2019) [18]	Initial project cost, project risks, project schedule	Achieved lower RMSE than ANN when applied to relatively small datasets.
Support Vector Machine (SVM)	Cheng & Roy (2011) [22]	Cost-related variables, project duration	Produced reliable prediction performance for limited project samples.
Random Forest (RF)	Akinosho et al. (2020) [23]	Technical factors, economic factors, managerial factors	Achieved high predictive performance ($R^2 > 0.90$) while effectively handling nonlinear relationships.
Random Forest (RF)	Pan & Zhang (2021) [14]	Cost factors, risk factors, productivity indicators	Outperformed ANN and SVM in terms of prediction accuracy and model robustness.
Extreme Gradient Boosting (XGBoost)	Chen & Guestrin (2016) [24]	Multidimensional project data	Achieved the lowest MAE and RMSE among the compared machine learning algorithms.
Extreme Gradie	Li et al.	Cost risks, design changes,	Consistently outperfor

nt Boosting (XGBoost)	(2022) [25]	labor productivity	med ANN and RF in construction cost prediction tasks.
Deep Learning	Elbaz et al. (2022) [26]	Large-scale project datasets	Demonstrated superior predictive capability for complex and high-dimensional datasets.
Deep Learning	Hosseini et al. (2023) [17]	BIM data, IoT data, project progress	Showed strong potential for real-time project cost prediction through digital technology integration.

Table 2 summarizes the principal machine learning algorithms identified in the reviewed studies, together with their predictor variables and reported performance. Earlier studies mainly applied ANN and SVM using historical project cost, schedule, and productivity data. More recent studies increasingly adopted ensemble learning techniques such as Random Forest and XGBoost because of their superior predictive performance, lower prediction errors, and ability to handle complex nonlinear relationships. Recent developments also demonstrate growing interest in Deep Learning models integrated with BIM and IoT data to support real-time project cost prediction.

The identified research gaps include:

1. No studies have yet concluded the best algorithm for all project types.
2. Direct comparisons of ANN–SVM–RF–XGBoost–Deep Learning on the same dataset are still limited.
3. Most studies focus solely on accuracy without considering model interpretability.
4. The use of explainable artificial intelligence (XAI) in cost overrun prediction is still very limited.
5. The integration of predictive models with BIM, digital twins, and IoT remains a promising research opportunity.

The challenges faced in applying ML to predict project cost overruns are:

1. Data Quality: many organizations lack a standardized project data collection system, resulting in incomplete and inconsistent historical data.
2. Data Imbalance: project cases with extreme cost overruns are relatively few, leading to an imbalanced data distribution.
3. Black Box Problem: most highly accurate models are difficult to explain to users, reducing the level of confidence among decision-makers.
4. Model Generalization: models developed for a specific project or region often cannot be directly applied to other projects.
5. System Integration Limitations: the integration of machine learning models with BIM, ERP, and project monitoring systems is still relatively limited.

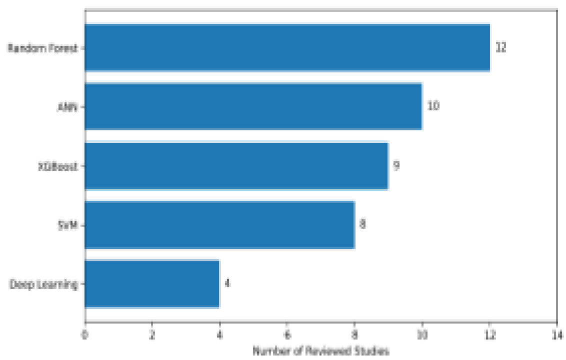

Figure 3. Frequency of Machine Learning Algorithms Identified in the Reviewed Studies

Among the 43 reviewed studies, Random Forest was the most frequently applied algorithm, appearing in 12 studies (27.9%), followed by Artificial Neural Network (10 studies; 23.3%), XGBoost (9 studies; 20.9%), Support Vector Machine (8 studies; 18.6%), and Deep Learning (4 studies; 9.3%). This distribution indicates a gradual transition from conventional machine learning approaches toward ensemble learning techniques because of their superior predictive performance and robustness when handling complex construction project datasets.

CONCLUSION

The following conclusions can be drawn from the research conducted:

1. The among the 43 selected studies, only one publication was identified in 2015. The number of publications gradually increased, reaching seven studies in 2022 before remaining relatively stable during the following years. Overall, approximately 65% of the reviewed publications were published between 2020 and 2025, indicating a substantial increase in research activity on machine learning applications for construction project cost prediction.
2. Based on the synthesis of 43 selected studies, Random Forest was identified as the most frequently applied machine learning algorithm (27.9%), followed by Artificial Neural Network (23.3%), XGBoost (20.9%), Support Vector Machine (18.6%), and Deep Learning (9.3%). The increasing adoption of ensemble learning algorithms reflects their superior capability to produce accurate and robust predictions for construction project cost overrun analysis.
3. The study results indicate that Machine Learning has significant potential to improve the effectiveness of project cost prediction systems and support more proactive decision-making integrated with digital technologies such as BIM, Digital Twin, and the Internet of Things. (IoT) in order to increase project success in a sustainable manner.

ACKNOWLEDGE

Special thanks are extended to Tarumanagara University for the research opportunity provided.

BIBLIOGRAPHY

[1] Atkinson, R. (1999). Project management: Cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, *17*(6), 337–342. https://doi.org/10.1016/S0263-7863(98)00069-6

[2] Kerzner, H. (2022). *Project management: A systems approach to planning, scheduling, and controlling* (13th ed.). Hoboken, NJ: John Wiley & Sons.

[3] Flyvbjerg, B., Holm, M. S., & Buhl, S. (2002). Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association*, *68*(3), 279–295. https://doi.org/10.1080/01944360208976273

[4] Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2004). What causes cost overrun in transport infrastructure projects? *Transport Reviews*, *24*(1), 3–18. https://doi.org/10.1080/0144164032000080494a

[5] Cantarelli, C. C., Flyvbjerg, B., Molin, E. J. E., & van Wee, B. (2012). Cost overruns in large-scale transportation infrastructure projects: Explanations and their theoretical embeddedness. *European Journal of Transport and Infrastructure Research*, *10*(1), 5–18.

[6] Love, P. E. D., Edwards, D. J., & Irani, Z. (2013). Moving beyond optimism bias and strategic misrepresentation: An explanation for social infrastructure

project cost overruns. *IEEE Transactions on Engineering Management*, *59*(4), 560–571. https://doi.org/10.1109/TEM.2011.2163628

[7] Odeck, J. (2004). Cost overruns in road construction—What are their sizes and determinants? *Transport Policy*, *11*(1), 43–53. https://doi.org/10.1016/S0967-070X(03)00017-9

[8] Doloi, H. (2013). Cost overruns and failure in project management: Understanding the roles of key stakeholders in construction projects. *Journal of Construction Engineering and Management*, *139*(3), 267–279. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000621

[9] Shah, R. K., Bhattarai, S., & Adhikari, K. (2022). Factors influencing cost overruns in construction projects: A systematic review. *International Journal of Construction Management*, *22*(14), 2715–2728. https://doi.org/10.1080/15623599.2020.1837714

[10] A. Afana, M. Enshassi, and S. Mohamed, (2024) "Factors affecting cost overruns in construction projects: A systematic review," International Journal of Construction Management, vol. 24, no. 2, pp. 215–230.

[11] S. Puteri, A. Nugroho, and D. Wibowo, (2022) "Analysis of factors influencing construction cost overruns in infrastructure projects," Journal of Engineering and Technological Sciences, vol. 54, no. 4, pp. 567–582.

[12] A. Suryawinata (2024) "Critical factors causing cost overruns in Indonesian construction projects: A review," Civil Engineering Dimension, vol. 26, no. 1, pp. 45–56.

[13] M. Bilal, L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, H. A. Owolabi, H. A. Alaka, and M. Pasha (2016) "Big Data in the construction industry: A review of present status, opportunities, and future trends," Advanced Engineering Informatics, vol. 30, no. 3, pp. 500–52.

[14] Y. Pan and L. Zhang,(2021), "Roles of artificial intelligence in construction engineering and management: A critical review and future trends," Automation in Construction, vol. 122, Art. no. 103517.

[15] M. Y. Cheng, M. T. Cao, and H. C. Hoang, (2010),"Predicting construction project success using machine learning techniques," Expert Systems with Applications, vol. 37, no. 7, pp. 4706–4713.

[16] A. Darko, A. P. C. Chan, Y. Yang, M. O. Tetteh, D. G. A. Osei-Kyei, and E. E. Amoah, (2020), "Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities," Automation in Construction, vol. 112, Art. no. 103081.

[17] M. R. Hosseini, E. Roelvink, T. Papadonikolaki, and N. Edwards, (2023),"Machine learning applications in construction management: A systematic review and future directions," Automation in Construction, vol. 148, Art. no. 104769, 2023.

[18] J. Son, C. Kim, and Y. K. Kim, (2019) "Machine learning-based construction cost prediction using Random Forest and Gradient Boosting algorithms," Journal of Construction Engineering and Management, vol. 145, no. 8, Art. no. 04019045.

[19] C. Molnar, (2022), Machine Learning Explainability. Birmingham, U.K.: Leanpub, 2022.

[20] A. Rai (2020) "Explainable AI: From black box to glass box," Journal of the Academy of Marketing Science, vol. 48, no. 1, pp. 137–141, 2020.

[21] G. H. Kim, S. H. An, and K. I. Kang, (2004), "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning," *Building and Environment*, vol. 39, no. 10, pp. 1235–1242, 2004.

[22] M. Y. Cheng and M. H. Roy, (2011), "Evolutionary fuzzy decision model for construction management applications," *Automation in Construction*, vol. 20, no. 4, pp. 438–447.

[23] A. Akinosho, L. O. Oyedele, M. Bilal, A. O. Ajayi, M. D. Delgado, and H. A. Akinade, (2020), "Deep learning in the construction industry: A review of present status and future innovations," *Journal of Building Engineering*, vol. 32, Art. no. 101827.

[24] C. Chen and A. Gastrin, (2016) "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, pp. 785–794.

[25] H. Li, Y. Wang, and X. Zhang (2022) "Machine learning approaches for construction cost overrun prediction using XGBoost and ensemble learning techniques," *Automation in Construction*, vol. 138, Art. no. 104233.

[26] M. Elbaz, A. AbouRizk, and Y. Mohamed (2022), "Deep learning applications in construction project performance prediction," *Automation in Construction*, vol. 137, Art. no. 104211.