2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

REVIEW OF MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE USE FOR COST ESTIMATION IN CONSTRUCTION PROJECTS

Emel Sadikoglu, Sevilay Demirkesen
Gebze Technical University, Kocaeli, Turkiye
esadikoglu@gtu.edu.tr, demirkesen@gtu.edu.tr

Abstract

Cost estimation is crucial in construction project management that influences project success. Machine learning and artificial intelligence methods have recently emerged as efficient tools in construction project cost estimation due to their ability to generate accurate predictions. This study presents a systematic literature review on the application of ML and AI methods in construction cost estimation based on PRISMA method using 75 publications. The review provides a comprehensive analysis of existing research, including data collection, project characteristics and predictive models. This study enhances understanding of ML and AI in construction cost estimation, aiding decision-makers and researchers in developing accurate models.

Introduction

The nature of the construction projects requires large capital for the initial phase and throughout the project (Wang and Ashuri, 2017). In construction project management, cost estimation is one of the most critical steps. Cost prediction is significant for project budget and resource allocation. Considering the time gap between initial estimation and actual construction, it is crucial to account for potential fluctuations in costs over time (Wang and Ashuri, 2017). Overestimation or underestimation causes deviations between expected and realized costs, which leads to critical problems. Obtaining input data for cost estimation is difficult when the scope of work is ambiguous. The accuracy of cost estimates varies throughout the project lifecycle with early-stage estimates being less precise due to limited design, documentation, and information. As the project progresses with clearer scope and defined specifications, estimates become more accurate and reliable (Hashemi et al., 2020; Kumar et al., 2023). The accuracy of predictions plays a significant role in determining the success or failure of the project. Hence, developing a simple and accurate cost estimate during the project's preliminary stage is one of the most challenging estimations in construction project management (Meharie et al., 2022).

Traditional methods remain limited in providing accurate predictions. To address inaccuracies and errors in cost estimation, studies have increasingly employed systematic approaches incorporating mathematical models, machine learning methods, and similar advanced methods for more reliable predictions (Hashemi et al., 2020). ML and AI methods are widely used in construction project cost estimation mainly due to their ability to generate accurate predictions without depending heavily on expert opinion or predefined rules (Meng et al., 2024). Several methods including Artificial Neural Networks (ANNs), hybrid models, Case-Based Reasoning (CBR), Regression Analysis (RA), Decision Tree (DT), and Support Vector Machine (SVM) have been proposed to achieve better and more accurate cost predictions (Hashemi et al., 2020). This study aims to provide an in-depth literature review of ML and AI methods used in construction cost prediction. In this context, 75 publications were investigated. Despite the array of publications, the area is yet to be developing fast, so this study mostly considers the recent publications in the field as they reflect the recent trends and applications.

Research method

This study conducts a systematic literature review based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method.

The PRISMA method which provides a structured guideline for systematic reviews (Page et al., 2020) is presented in Figure 1. Scopus database was used. The search query was structured as follows: TITLE-ABS-KEY((cost OR budget OR financ*) W/5 (predict* OR estimat* OR forecast* OR projection) AND ("machine learn*" OR "artificial intelligence" OR "deep learning") AND ("construction project" OR "civil engineering project" OR "building project")). Non-English studies were excluded. Publication years were selected 2000-2024. Document type was selected as journal articles and conference papers. In the screening phase, studies were scrutinized based on their title, abstract, and keywords. Further, full text papers were examined. The inclusion criteria: i) conducting a study specifically on cost estimation, ii) using at least one machine learning, artificial intelligence, or deep learning method, and iii) conducting a study in the context of construction projects.

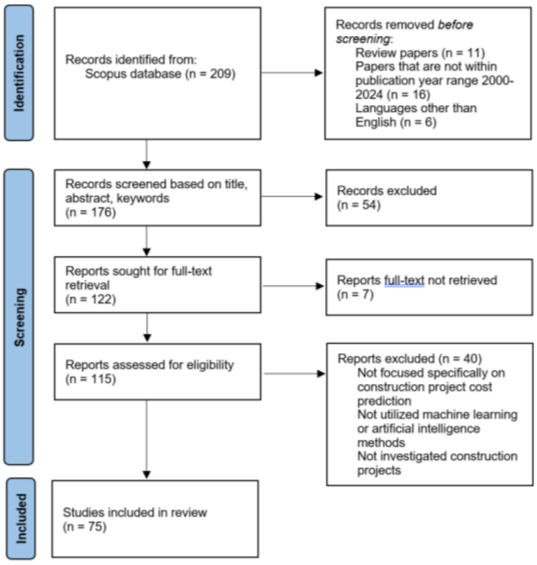

Figure 1: Flowchart of articles screened in this study

Reviewed Studies Description

Descriptive statistics related to reviewed studies are investigated to reveal the trends and patterns. Figure 2 shows the number of studies by year. There were relatively few studies published until 2019. The overall graph shows a slow but steady increase in the number of publications. In recent years, emerging and rapidly growing interest in ML and AI methods has been observed with a significant increase in 2024.

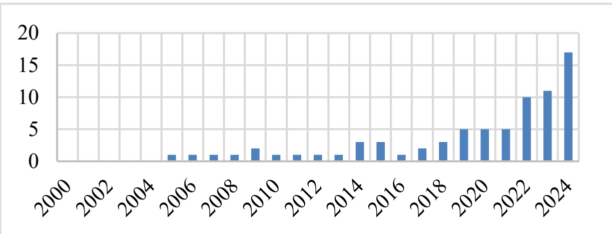

Figure 2: Number of studies by year

In Figure 3, the bar chart provides an overview of the number of studies based on publication titles. This figure highlights the major publication sources. Automation in Construction is the leading journal, followed by the Asian Journal of Civil Engineering and Applied Sciences.

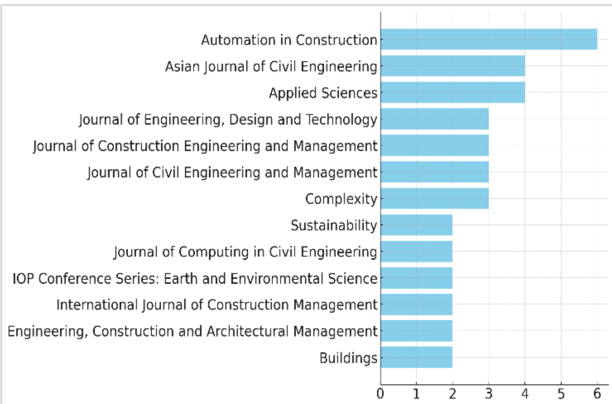

Figure 3: Number of studies by journal

Analysis of Reviewed Studies

Project type

Nearly all studies have developed a model focusing on a specific structure type. Hence, the models are expected to best represent and predict the structure types they modelled. Mostly, studies focused on one type of project because of the differing characteristics of project types. A large number of studies examined building projects. Many studies considered infrastructure construction projects such as bridge projects, road projects, metro construction projects, water pipeline projects, wastewater pipeline projects. Governmental construction projects such as governmental offices and educational buildings were also examined. Industrial projects were considered in a few studies such as industrial factory buildings and EPC projects in the energy sector. Differently, a few studies investigated specialized construction projects. Figure 4 presents the number of studies by project type, showing that most studies focus on building projects, while other project types are limited.

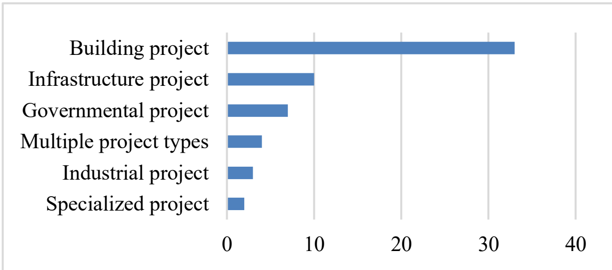

Figure 4: Number of studies by project type

Project phase

In construction projects, cost estimation methods depend on the project phase, its objectives, and the level of information available. In the inception phase, the level of information is limited, and the main objective is to identify the feasibility of the project. Preliminary cost estimates at this phase mostly rely on simple and quick calculation methods. In early phases, accuracy is a challenge (Miranda et al., 2022). A major portion of studies estimated conceptual cost and mostly focused on area, quantity, and cost to estimate total construction cost. Meng et al. (2024) estimated the cost of metro construction projects during the project initiation phase.

In reviewed studies, the investigated project phase is predominantly the design phase. Cheng et al. (2009) predicted cost in the conceptual and preliminary design stages. Some studies predicted cost in the detailed design phase which presents detailed cost prediction at the building component level (Banihashemi et al., 2022; Cheng et al., 2010). In the bidding phase, Chou et al. (2015) predicted the bid award price for bridge construction projects. Kusonkhum et al (2022) predicted the cost of different governmental projects such as building and road projects, in the bidding phase. Car-Pušić and Mlađen (2020) investigated the relation of real construction cost and contracted tender price. In the construction stage, project managers usually estimate costs relying on the Earned Value Management (EVM) method that provides an estimate at completion, project status tracking, and project performance measurement (Cheng and Hoang, 2014). Cheng and Hoang (2014) developed a comprehensive model to assist cost planning and monitoring at the construction phase. Cheng and Roy (2011) predicted the construction project cash flow using time-series data. Cheng and Khasani (2024) introduced a comprehensive model which provided a better prediction performance than the earned value management method. Few studies considered the operation and maintenance (O&M) phase. An (2023) developed an early warning system for cost deviations in the O&M phases of large-scale construction projects. Kim et al. (2024) developed a deep-learning model to estimate repair and maintenance costs of apartment buildings.

Research steps

The research steps for developing prediction models can be seen in Figure 5. In the studies, first, the problem is defined, and research objectives are set. Then, input and output parameters (or features variables, or items) are specified. In data collection, relevant data from historical projects, databases, or case studies are gathered. Collected data is preprocessed, cleaned, and prepared for data analysis. Exploratory data analysis is used in some studies to better understand data distribution, patterns and relationships. Besides, some studies utilize feature selection to reveal the most important features to improve the model. In the method selection stage, appropriate methods and model parameters are selected. In data analysis, the outputs of the developed model are assessed. Then, the model is trained, and model performance is evaluated based on performance metrics. Validation methods are used to ensure the generalization of the model. Finally, the model can be tested and used.

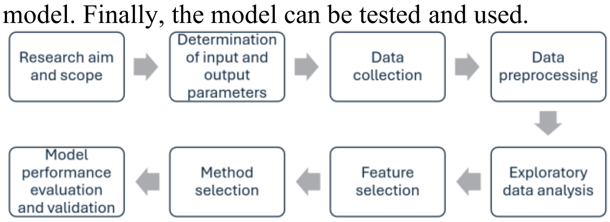

Figure 5: Prediction model development steps (Adapted from Shmueli and Koppius, 2010)

Input and output variables

The selection of input variables varied across studies, especially based on the project type and project phase. Input variables used in building construction cost estimation studies across different project phases are summarized in Table 1.

Input variables specific to infrastructure projects were also identified in previous studies. Wastewater pipeline construction project cost estimation input parameters were determined as year, duration, governorate, type of finance, depth, pipe length, type, diameter, number of manholes, soil type, worksite, groundwater table level, works on paved and unpaved surface (Abbas and Aswed, 2024; Juszczyk et al., 2023). In the pre-design and design phases of bridge construction project cost estimation, input variables include structure type, bridge type, project type, bridge length and width, number of spans, span length, pier height, supports, structural solution, material solution, foundation method, load class, and construction technology (Juszczyk, 2020; Beljkaš and Knežević, 2021). In the bidding stage of bridge construction projects, Chou et al. (2015) identified input variables as budget amount, compliance period, unit-price analysis table, number of detailed estimates, road constructed area, upper constructed area, number of bridge abutments and piers and bid authority. In the pre-design and design phases of highway construction, Abd et al. (2024) considered cost-driving variables, including the project year, road length, land preparation, earthworks, sub-base works, paving works, and shoulder works. Meharie et al. (2022) focused on the number of bridges, inflation rate, terrain type, and project type. In the study by Meng et al. (2024), the variables for metro construction cost estimation included project-specific variables such as length of line, number of stations, station distance, design speed, economic variables. The study by Mahdavian et al. (2020) focused on the construction phase of highway projects, utilized monthly historical data. In the study by Kusonkhum et al. (2022), used input variables including project location, project duration, project scale, project type, project department and bidding method for government construction projects' procurement phase.

Table 1: Input variables for building projects in different project phases

Project phase	Input variables
Pre-design and design	duration, location, year, building type, building height, total floor area, number of floors, story height, foundation type, soil condition, seismic zone, design-related variables, material-based variables, other building components related variables, economic variables
Construction	project characteristic (total area, duration), project progress and performance indicators (construction progress, Earned Value, Actual Cost, Schedule Performance Index, Cost Performance Index, cost ratio of component), contract-related variables (contract payment, subcontractor billed index, owner billed index, change order index), economic variables, environmental variables
O&M	building age, usage types, weather-related factors (temperature, precipitation, wind speed, solar radiation)

A large number of studies predicted output variable as total project cost (Jiang, 2020; Juszczyk et al., 2023; Liu et al., 2024). Output variable was most frequently total construction cost (in \$, € or local currency) or construction cost per m^2 (Kong et al., 2008; Cheng et al., 2010; Beljkaš and Knežević, 2021; Liu and Lin, 2024; Das et al., 2024). There were some studies that predict only one cost component. The majority of studies approached construction project cost prediction as a regression problem, however, some studies considered it as a classification problem.

Data collection

In a large number of studies, historical data of previously completed projects such as project specification, project documentation, surveys were gathered mostly from general contractors, governmental bodies, consultancy companies and construction owners. Many studies used official and governmental databases, data presented by other research articles, questionnaire survey and BIM data. Data size significantly impacts the model accuracy. As the data size increases, reduced sample variance and bias are obtained. In the reviewed studies, the smallest dataset contains 10 samples. Even though most of the studies utilized relatively small datasets, there are also exceptionally large datasets. The histogram represents the categorization of the sample sizes in Figure 6. The majority of studies use datasets with fewer than 200 samples. This shows that studies mostly rely on relatively small datasets.

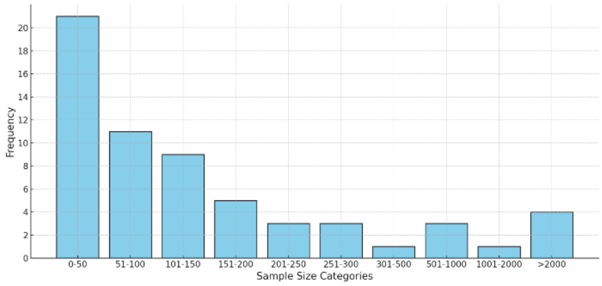

Figure 6: Histogram of sample sizes in reviewed studies

Studies mostly collected data and verified their results based on a specific country. In different studies, data was collected from many countries all around the world. Figure 7 presents the countries of reviewed studies' datasets. Asia leads the research output, with China, Taiwan, and Korea contributing the highest number of studies, followed by European countries such as Poland and the United Kingdom.

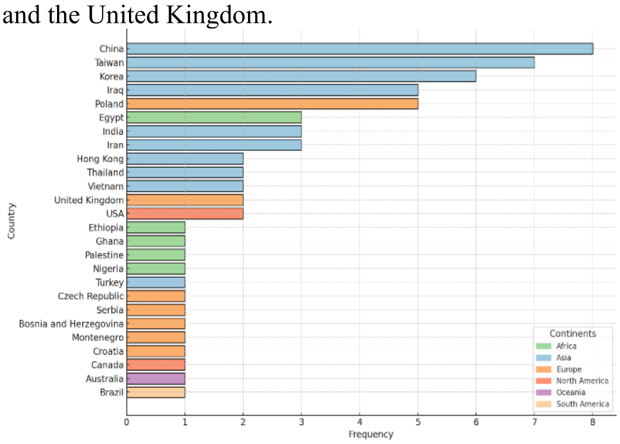

Figure 7: Geographic distribution of datasets

Figure 8 shows the time range of datasets in reviewed studies. The longest data collection period was 22 years, observed in studies covering 1993-2015 and 2000-2022. The shortest period was 1 year, in studies collected data in 2008 and 2019.

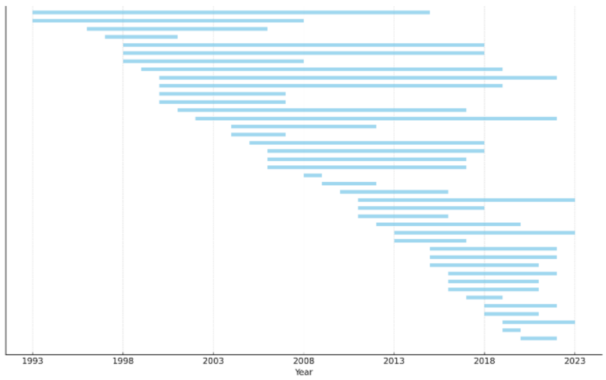

Figure 8: Time range of datasets

Feature selection

Feature selection is the method of choosing the most appropriate variables and removing unnecessary variables from a set of potential variables. Feature selection is significant since the selected features affect the model. Many studies determined features based on qualitative approach such as literature review, expert opinion and questionnaire survey. Quantitative approaches were also utilized. Within exploratory data analysis, dimension reduction is used to reduce the number of parameters (Miranda et al., 2022). PCA was utilized in many studies for feature selection (Jiang, 2020; Liu and Lin, 2024; Abd et al., 2024; Son et al., 2012; Wang et al., 2024a). Some regression-based methods were also employed such as stepwise regression (Chou et al., 2015) and LASSO (Tajziyehchi et al., 2021).

Method selection

ML and AI methods in construction cost estimation can be categorized as single models, ensemble models, and hybrid models. A single model indicates that a single ML or AI model is used for the problem. Single models can be base models, or they can be derived from the base models (Wang et al., 2024b). Regression models and ANN are broadly employed methods for predicting construction costs (Mahdavian et al., 2020). Linear and logistic regressions are commonly used methods, but they are mostly limited to detecting simple and linear relations. DT, ANN, and SVM have the capacity to capture complex patterns and nonlinear relationships; however, they have risk of overfitting (Miranda et al, 2022).

One of the widely employed methods in construction cost prediction is ANN which can capture complex patterns in the data; however, its black-box nature makes interpretability difficult (Meng et al., 2024). There are advanced variations of ANN such as Deep Neural Network (DNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Back-Propagation Neural Network (BPNN), Multi-Layer Perceptron (MLP), Extreme Learning Machine (ELM) and Radial Basis Function Network (RBFN). Some studies utilized ANN variations as single models such as DNN (Park and Yun, 2023), BPNN, RBFN (Jiang, 2020), LSTM, RNN, and GRU (Kim et al., 2024).

Regression-based models were also used in reviewed studies such as Multiple Regression (ML), Generalized Linear Regression (GENLIN), and Multivariate Adaptive Regression Spline (MARS). Sharqi and Bhattarai (2021) utilized ELM, MARS, and Partial Least Square Regression (PLS) methods. Based on the results, MARS was the best-performing model. Several studies utilized SVM (Cheng and Hoang, 2014) and SVM-Regression (Juszczyk, 2019; Juszczyk, 2020). Kong et al. (2008) compared SVM-Regression and RBFN methods and SVM-Regression showed better prediction accuracy. Car-Pusić and Mladen (2020) employed single models including Linear Regression (LR), General Regression Neural Network (GRNN), Radial Basis Function (RBF), and SVM, where the RBF model achieved the best performance.

Ensemble models combine multiple weak or strong models and aggregate their predictions to enhance accuracy and robustness. Ensemble methods involve bagging, boosting, stacking-based, averaging and voting methods. Juszczyk et al. (2019) and Juszczyk et al. (2023) created ensemble models of ANN by combining MLP networks with simple and generalized averaging methods. Liu and Lin (2024) integrated DT, BPNN, and SVM using a weighted average method. Meharie et al. (2022) proposed a stacking ensemble model by combining predictions of LR, SVM, and ANN methods. The stacking ensemble model outperformed LR, SVM, and ANN. Tree-based ensemble models are Random Forest (RF), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), LightGBM (LGBoost), Natural Gradient Boosting (NGBoost) and Stacked Decision Trees. Ali et al. (2022) compared RF, ANN, and SVM methods and RF outperformed others. Chakraborty et al. (2020) compared LR, ANN, RF, XGBoost, LGBoost, and NGBoost and found that hybrid LGBoost and NGBoost model provides the most desirable results. Ali and Burhan (2023) compared XGBoost, ELM, and MARS models and XGBoost provided the best results.

A hybrid model combines multiple separate models aiming to leverage their strengths and compensate for their limitations (Wang et al., 2024b; Liu and Lin, 2024). Researchers developed hybrid models to combine the advantageous properties of different methods and benefit from their synergy. Wang et al. (2024a) proposed a hybrid RA+ANN model which outperformed standalone RA, ANN, and CBR models. De Soto and Adey (2016) proposed a hybrid CBR+RA+NN model and the results showed better performance compared to single models. Some studies proposed hybrid models that combine optimization algorithms with base models. Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Differential Evolution (DE) are metaheuristic optimization algorithms. Evolutionary Fuzzy Neural Inference Model (EFNIM) was developed by integrating Fuzzy Logic (FL), NN, and GA by Cheng et al. (2009). Cheng et al. (2010) developed an Evolutionary Fuzzy Hybrid Neural Network (EFHNN) combining neural network (NN), high order neural network (HONN), FL and GA. The comparison of results revealed that EFHNN outperformed EFNIM. Cheng et al. (2015) presented the Adaptive Time-Dependent Least Squares Support Vector Machine (LS-SVMAT) model by incorporating LS-SVM, Adaptive Time Function (ATF), and DE. Du and Li (2017) proposed an improved BPNN model with GA optimization. Mhady et al. (2024) integrated Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and ANN with AOA. They reported the hybrid model achieved improved predictive capabilities with fewer input parameters.

Model performance evaluation and validation

Predictive accuracy is the model's ability to make accurate predictions for new observations. Predictive performance can be evaluated by using predictive measures to evaluate the model's accuracy and by selecting an appropriate validation method (Shmueli and Koppius, 2011; Miranda et al., 2022). Generic predictive measures for model performance evaluation are Mean Absolute Error (MAE), Mean Square Error (RMSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) in regression problems, on the other hand, accuracy, precision, and recall are mostly reported for classification problems. MAE, MSE, and RMSE are scale-dependent metrics, while MAPE is scale-independent which can be utilized for comparison across datasets. MAPE formula (1) is as follows (Dang-Trinh et al., 2023), where n is number of samples, y_a is actual value and y_p is predicted value.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_p - y_a|}{y_a} \quad (1)$$

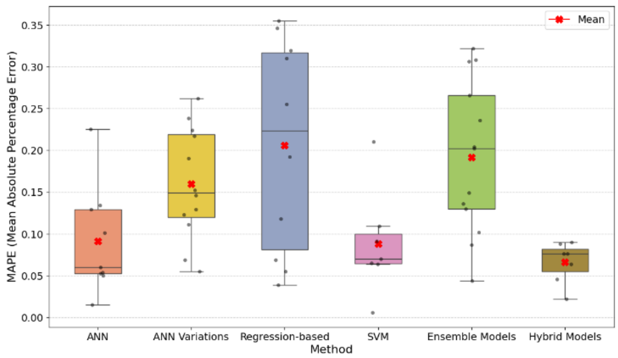

Figure 9: Comparison of MAPE values for different methods

Figure 9 presents the MAPE values for different methods reported in the reviewed studies. Regression-based models exhibit the highest MAPE values with a broad range, indicating the lowest overall performance among other methods. ANN and variations have moderate values with a wide spread, where the performance mostly depends on model parameters. SVM shows low median and mean MAPE values, indicating that it is a strong alternative, especially with small datasets. Ensemble models can have significant potential; however, parameter tuning is significant. Hybrid models have the lowest median and mean MAPE which offers high accuracy by combining strengths of different methods.

Cross-validation is a common method for evaluating the performance of machine learning models by assessing the model's generalizability to new data. It is used for the evaluation of the final model predictive performance. Holdout cross-validation splits the dataset into a training set and a testing set (Miranda et al., 2022). The k-fold cross-validation uses the entire dataset to be utilized for training by cycling through every subset for validation (Park and Yun, 2023). Most frequently, 5-fold cross-validation and 10-fold cross-validation were utilized. Leave One Out Cross Validation (LOOCV) is the extreme case where the number of subsets is equal to a number of data points (Miranda et al., 2022) (e.g. Car-Pusić and Mladen, 2020).

Predictive accuracy is influenced by the model type; however, models with higher accuracy may reduce the interpretability and objectivity of models (Miranda et al., 2022). Most studies only evaluated and compared predictive accuracy without explanation of variables affecting cost (Liu and Lin, 2024). Interpretation is important in understanding the model. Some studies addressed this limitation by using SHapley Additive exPlanations (SHAP) analysis for interpretability (Meng et al., 2024; Chakraborty et al., 2020; Das et al., 2024).

Discussion and Conclusions

This study conducted a comprehensive systematic literature review of employed ML and AI methods for cost estimation in construction projects. The number of publications indicates a growing interest and adoption of ML and AI methods. The majority of studies examined building projects. The higher number of studies may be attributed to data availability, whereas other project types such as infrastructure and industrial projects often have data restriction and confidentiality. The limited number of studies in other project types suggests that future studies can focus on infrastructure projects, governmental projects, industrial projects and specialized projects. More specific variables can be defined for each project type. Moreover, comprehensive models including several project types can be developed for broader applicability of developed models.

The project phase significantly influences cost estimation accuracy. The studies were mostly focused on planning and conceptual phases. There were also many studies investigating design phase. Early-stage predictions exhibit higher variability because of the limited availability of information and details. Studies focusing on planning and conceptual phases mostly use fewer variables which leads to increased uncertainty compared to cost estimation models used in later phases, such as bidding or construction. The scarcity of resources investigating the cost of construction phase can also be observed in the number of studies investigating operation and maintenance phase. Most studies focused on planning and conceptual design phase. Hence, it is recommended to develop a novel prediction model for the cost estimation specific to construction phase. The stakeholder perspective remains an unexplored area in cost estimation. Most studies focus on general contractor or owner perspectives, while other stakeholders such as subcontractors and material suppliers were overlooked. However, the construction industry involves many stakeholders which should be considered with a more comprehensive framework. In addition, the influence of project funding type and contract types is not considered. Future research should include these factors in the prediction models.

Input variables encompass a wide range of parameters in previous cost prediction studies. However, the incorporation of external variables into models is limited. Hence, external parameters including economic variables, the impact of force majeure events such as disasters, pandemic, war (Wang et al., 2024b) should be considered. For example, inflation was found to be the most significant parameter in some studies using data collected from developing countries (Ali et al., 2022). On the other hand, inflation may not affect the prediction extremely in developed countries that have stable economies. Hence, country specific parameters should also be incorporated into the prediction models.

Asia leads the research output, contributing the highest number of studies, followed by significant representation from European countries, while some developing regions are underrepresented. This geographical imbalance may be attributed to limited research opportunities, lack of open-access, structured datasets and weak collaboration between construction industry and academic institutions in developing countries. In contrast, developed countries often have greater access to datasets, stronger industry-academia collaborations, and advanced technological infrastructure. Further, language barriers and regional publication preferences may result in underreporting. This highlights the need for more studies from countries located in other continents for better generalization of the results. The average time range for the collected construction project data was approximately 9 years. In order to better predict construction costs and generalize the models, longer time range can be utilized in studies.

Despite these advancements, several research gaps remain. Developing ML and AI models for construction cost prediction is challenging since it requires high-quality data, and large datasets for better predictive accuracy of the models. Data size significantly impacts the model accuracy. The construction industry, being a traditional sector, lacks data storage, sharing, integration, and effective use. Furthermore, resistance to change is a significant barrier to data collection and research. To access high quality and large datasets, issues related to data acquisition, lack of standardization, inconsistent data formats, security concerns, lack of transparency and trust should be addressed. Collaborative international data-sharing platforms can be created, and regional research funding can be provided. Moreover, novel models should be developed, incorporating country-specific input parameters and creating hybrid models by leveraging best performing models for better prediction accuracy.

This study is expected to guide both policymakers and industry practitioners in terms of conducting innovative studies in the construction industry utilizing the latest methods in ML and AI. The results of this study aim to provide what the main gaps are in the previous studies and encourage studies to produce models that can accurately predict the costs and durations in a construction project.

Acknowledgments

The authors acknowledge the support of the Scientific and Technological Research Council of Türkiye (TÜBİTAK) through the 2224-A International Scientific Events Participation Grant.

References

Abbas, A., & Aswed, G.K. (2024, August). Enhancing Sewage Pipeline Project Cost Estimations in Iraq through Artificial Neural Network Models. In IOP Conference Series: Earth and Environmental Science, 1374(1), p. 012086). IOP Publishing.

Abd, A.M., Kareem, Y.A., & Zehawi, R.N. (2024). Prediction and Estimation of Highway Construction Cost using Machine Learning. Engineering, Technology & Applied Science Research, 14(5), 17222-17231.

Ali, Z.H., Burhan, A.M., Kassim, M., & Al-Khafaji, Z. (2022). Developing an integrative data intelligence model for construction cost estimation. Complexity, 2022(1), 4285328.

Banihashemi, S., Khalili, S., Sheikhkhoshkar, M., & Fazeli, A. (2022). Machine learning-integrated 5D BIM informatics: Building materials costs data classification and prototype development. Innovative infrastructure solutions, 7(3), 215.

Beljkaš, Ž., & Knežević, M. (2021). Use of artificial intelligence for estimating cost of integral bridges. Građevinar, 73(03), 265-273.

Car-Pušić, D., & Mlađen, M. (2020). Early stage construction cost prediction in function of project sustainability. In 15th International Conference on Durability of Building Materials and Components.

Chakraborty, D., Elhegazy, H., Elzarka, H., & Gutierrez, L. (2020). A novel construction cost prediction model using hybrid natural and light gradient boosting. Advanced Engineering Informatics, 46, 101201.

Cheng, M.Y., & Hoang, N.D. (2014). Interval estimation of construction cost at completion using least squares support vector machine. Journal of Civil Engineering and Management, 20(2), 223-236.

Cheng, M.Y., & Khasani, R. R. (2024). Least Square Moment Balanced Machine: A New Approach to Estimating Cost to Completion for Construction Projects. Journal of Information Technology in Construction (ITcon), 29(23), 503-524.

Cheng, M.Y., & Roy, A.F. (2011). Evolutionary fuzzy decision model for cash flow prediction using time-dependent support vector machines. International journal of project management, 29(1), 56-65.

Cheng, M.Y., Tsai, H.C., & Hsieh, W.S. (2009). Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. Automation in construction, 18(2), 164-172.

Cheng, M.Y., Tsai, H.C., & Sudjono, E. (2010). Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. Expert Systems with Applications, 37(6), 4224-4231.

Cheng, M.Y., Hoang, N.D., & Wu, Y.W. (2015). Cash flow prediction for construction project using a novel adaptive time-dependent least squares support vector machine inference model. Journal of Civil Engineering and Management, 21(6), 679-688.

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.

Dang-Trinh, N., Duc-Thang, P., Nguyen-Ngoc Cuong, T., & Duc-Hoc, T. (2023). Machine learning models for estimating preliminary factory construction cost: case study in Southern Vietnam. International Journal of Construction Management, 23(16), 2879-2887.

Das, P., Kashem, A., Hasan, I., & Islam, M. (2024). A comparative study of machine learning models for construction costs prediction with natural gradient boosting algorithm and SHAP analysis. Asian Journal of Civil Engineering, 1-16.

Du, Z., & Li, B. (2017). Construction project cost estimation based on improved BP Neural Network. In 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA) (pp. 223-226). IEEE.

De Soto, B.G., & Adey, B.T. (2016). Preliminary resource-based estimates combining artificial intelligence approaches and traditional techniques. Procedia engineering, 164, 261-268.

Hashemi, S., Asadi, S., & Sheikholeslami, M. (2020). “A systematic review of machine learning techniques for cost estimation and prediction in construction projects”, Journal of Project Management, 5(3), 85-105.

Jiang, S. (2020). “Prediction of building cost based on BP neural network”, Journal of Physics: Conference Series, 1437(1), 012071.

Juszczyk, M. (2019). Cost estimates of buildings' floor structural frames with the use of support vector regression. In IOP Conference Series: Earth and Environmental Science, 222, 012007, IOP Publishing.

Juszczyk, M. (2020). On the search of models for early cost estimates of bridges: An SVM-based approach. Buildings, 10(1), 2.

Juszczyk, M., Hanák, T., Výskala, M., Pacyno, H., & Siejda, M. (2023). Early Fast Cost Estimates of Sewerage Projects Construction Costs Based on Ensembles of Neural Networks. Applied Sciences, 13(23), 12744.

Juszczyk, M., Zima, K., & Lelek, W. (2019). Forecasting of sports fields construction costs aided by ensembles of neural networks. Journal of Civil Engineering and Management, 25(7), 715-729.

Kim, J.M., Yum, S. G., Adhikari, M.D., & Bae, J. (2024). A LSTM algorithm-driven deep learning approach to estimating repair and maintenance costs of apartment buildings. Engineering, Construction and Architectural Management, 31(13), 369-389.

Kong, F., Wu, X.J., & Cai, L.Y. (2008). A novel approach based on support vector machine to forecasting the construction project cost. In 2008 International Symposium on Computational Intelligence and Design, 21-24, IEEE.

Kumar, S., Kumar, V., & Kumar, A. (2023). "Machine learning algorithms for cost estimation in construction projects: A review", Journal of Building Engineering, 62, 105551.

Kusonkhum, W., Srinavin, K., Leungbootnak, N., Aksorn, P., & Chaitongrat, T. (2022). Government construction project budget prediction using machine learning. Journal of Advances in Information Technology Vol, 13(1).

Liu, Z., & Lin, J. (2024). Combinatorial machine learning approaches for high-rise building cost prediction and their interpretability analysis. Journal of Asian Architecture and Building Engineering, 1-12.

Liu, H., Cheng, J.C., & Anumba, C.J. (2024). A Graph Neural Network Approach to Conceptual Cost Estimation. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 41, pp. 815-821). IAARC Publications.

Mahdavian, A., Shojaei, A., Salem, M., Yuan, J.S., & Oloufa, A. A. (2021). Data-driven predictive modeling of highway construction cost items. Journal of Construction Engineering and Management, 147(3), 04020180.

Meharie, M.G., Mengesha, W.J., Gariy, Z.A., & Mutuku, R.N. (2022). Application of stacking ensemble machine learning algorithm in predicting the cost of highway construction projects. Engineering, Construction and Architectural Management, 29(7), 2836-2853.

Meng, C., Qu, D., & Duan, X. (2024). Cost Estimation of Metro Construction Projects Using Interpretable Machine Learning. Journal of Computing in Civil Engineering, 38(6), 04024038.

Mhady, A.A., Budayan, C., & Gurgun, A.P. (2024). Estimate-at-completion (EAC) prediction using Archimedes optimization with adaptive fuzzy and neural networks. Automation in Construction, 166, 105653.

Miranda, S.L., Del Rey Castillo, E., Gonzalez, V., & Adafin, J. (2022). Predictive analytics for early-stage construction costs estimation. Buildings, 12(7), 1043.

Page, M.J., McKenzie, J.E., Bossuyt, P. M.,... & Brennan, S.E. (2020). The PRISMA, et al. statement: an updated guideline for reporting systematic reviews. Bmj, 2021(372), 71.

Park, D., & Yun, S. (2023). Construction Cost Prediction Using Deep Learning with BIM Properties in the Schematic Design Phase. Applied Sciences, 13(12), 7207.

Rafiei, M.H., & Adeli, H. (2018). Novel machine-learning model for estimating construction costs considering economic variables and indexes. Journal of construction engineering and management, 144(12), 04018106.

Son, H., Kim, C., & Kim, C. (2012). Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables. Automation in Construction, 27, 60-66.

Tajziyehchi, N., Moshirpour, M., Jergeas, G., & Sadeghpour, F. (2021, August). A Methodology and Tool for the Predictive Analysis of Cost Growth in Construction Projects. In 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 240-247). IEEE.

Wang, X., & Ashuri, B. (2017). "Forecasting construction cost index in the short-, mid-, and long-terms using k nearest neighbor and perfect random tree ensembles", Construction Management and Economics, 35(6), 349-364.

Wang, Y., Zuo, J., Pan, M., Tu, B., ... & Dong, N. (2024a). Cost prediction of building projects using the novel hybrid RA-ANN model. Engineering, Construction and Architectural Management, 31(6), 2563-2582.

Wang, R., Salleh, H., Abdul-Samad, Z., Radzuan, N.F.M., Wen, K. C. (2024b). Fundamentals of Developing Conceptual Cost estimation Models Using Machine Learning Techniques. Planning Malaysia: Journal of the Malaysian Institute of Planners, 3, 242-256.