

Review of Supervised Machine Learning Cost Estimation Techniques for Building Projects

Faith Dowelani

University of Pretoria, Hatfield, South Africa

ABSTRACT

Cost is considered a vital parameter in determining the success of a construction project. Project costs control and monitoring prevent budget overruns and safeguard expected profits, regardless of the project's size, scope, or complexity. Traditional methods for estimating project costs are facing growing challenges as demand for more accurate, adaptable strategies that respond to evolving market dynamics and technological progress increases. This study offers insight into supervised ML-based cost estimation techniques, highlighting the models employed, the geographical area of the studies, sample sizes, input and output variables, and property types. The findings indicate that there has been some progress in applying supervised ML for cost estimation. Asia accounts for the most studies (65.96%), followed by Europe (14.89%) and Africa (10.64%). Oceania and North America each account for 4.26%, indicating a restricted research scope in these areas. Additionally, 62% of the studies employed multiple algorithms to enhance the reliability of the results. Moreover, most studies focused on construction costs rather than total project costs or total capital investment (project investment) and on residential and educational property types. The findings suggest that extensive testing and applications are necessary to gain a comprehensive understanding of global perspectives, particularly outside Asia, and in commercial properties such as retail and office buildings.

Keywords: Supervised machine learning, Building projects, Property types, Cost estimation, Construction costs, Project costs

INTRODUCTION

Cost estimation is a crucial approach for decision-making, ensuring success and safeguarding a project's financial stability throughout its various stages (Hakami & Hassan, 2019; Sanni-Anibire, Zin & Olatunji, 2022). In cost estimation, various techniques are employed, each offering unique strengths and applications. However, some of these techniques are often criticised for providing an incomplete account of the overall project costs. Cost estimation methods encounter several challenges. These challenges include low estimation accuracy, errors associated with manual data entry and calculations, and long estimation timeframes (Chen et al., 2025). The limitations of traditional cost estimation techniques necessitate integrating more adaptive, strategic approaches to improve the accuracy and efficiency of cost estimates. As the construction industry increasingly adopts AI, the application of machine

learning (ML) is anticipated to expand in line with Industry 4.0's broader objectives (Baduge et al., 2022). Various ML models have been employed in planning and managing construction projects, including various applications, from creating construction schedules (Holland, Cox, & Worthington, 2017), cost estimation (Chen et al, 2025; Gurmu & Miri, 2023), assessing project risks (Tandukar, Goh, & Chan, 2019), to improving the accuracy and automation of project planning (Pejic et al., 2023). Research on ML highlights that ML can significantly enhance efficiency, reduce costs, boost productivity, improve safety standards, identify potential risks of project delays and budget overruns, and deliver precise, accurate information (Pejic et al., 2023). This study builds on previous research on the application of an ML cost estimation model. It aims to assess the current progress on applying supervised ML techniques and to identify the models employed across various property types and the geographical distribution of the research.

Supervised Machine Learning

Machine learning (ML) is a branch of AI focused on developing algorithms that enable computers to learn from data and generate models from input data (Baduge et al., 2022). The aim of the supervised learning algorithm is to create a predictive function using training datasets that can effectively generalise to accurately classify new input data (Jones, 2008). These algorithms have become well-known for their adaptability, efficiency, and ability to handle diverse data types and challenges. ML can be applied throughout different phases of a construction project, including design, construction, and maintenance. ML has been utilised in several construction areas, including cost estimation, construction monitoring, project performance, and material classification (Alaloul & Qureshi, 2021). According to Kamiri & Mariga (2021), support vector machines (SVMs), random forests (RFs), decision trees (DTs), Naïve Bayes, and artificial neural networks (ANNs) are commonly acknowledged as leading algorithms for classification and prediction tasks in ML.

Supervised ML Cost Estimation Techniques

Supervised ML applications aim to enhance cost estimation and risk assessment, which are crucial for informed decision-making (Tayefeh Hashemi et al., 2020). Arage and Dharwardkar (2017) identified case-based reasoning (CBR), bootstrap resampling techniques, fuzzy logic (FL), rough set theory (RST), multiple regression analysis (MRA), and Neural networks (NN) as ML models utilised for cost estimation. These algorithms have become well-known for their adaptability, efficiency, and ability to handle diverse data types and challenges. Sanni-Anibire et al. (2022) developed and compared 12 machine learning models, and found that MLR, SVM, and ANN yielded better results across all variables, whereas K-Nearest Neighbors (kNN) performed best with missing values, and RST performed best with unclear data. Gurmu and Miri (2023) employed DT, linear regression (LR), Gradient Boosting (GB), RF, and kNN models to assess the cost range of construction projects for buildings, based on parameters such as gross floor area and building type. LR produced the worst results, whereas GB and DT

achieved the best results. These findings are supported by those of Pham et al. (2021). Pham et al. (2021) found that the LR model provides accurate results for construction cost estimation, whereas the DT model provides accurate results for maintenance cost estimation. Kim et al. (2004) evaluated the performance of CBR, NN, and MRA. The findings revealed that NN outperformed MRA and CBR in accuracy and exhibited a smaller error margin. FL was instrumental in cost-estimating situations where precise values were lacking, as it can effectively manage imprecise or uncertain information (Aragé & Dharwardkar, 2017).

METHODOLOGY

To enhance understanding of how supervised ML models are utilised to estimate building project costs, this research followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (as shown in Figure 1) to filter articles from the Scopus and Google Scholar databases. A range of keywords was derived from the title. The following keywords were selected as search terms for the databases: cost estimating, cost estimation, cost prediction, construction costs, building projects, project costs and preliminary costs. The screening process focused on English-language articles in the field of engineering. During the screening process, publications that were not journal articles, peer-reviewed, or conference papers were excluded. The inclusion criteria focused on articles on building projects and on supervised ML estimation models. A review of 226 articles was conducted. The exclusion criteria included studies that focused on predicting material quantities, design costs, railway, highway and road project costs, project site overheads, contractor costs, and engineering services, resulting in the exclusion of 179 articles. Finally, 47 articles were evaluated based on empirical evidence.

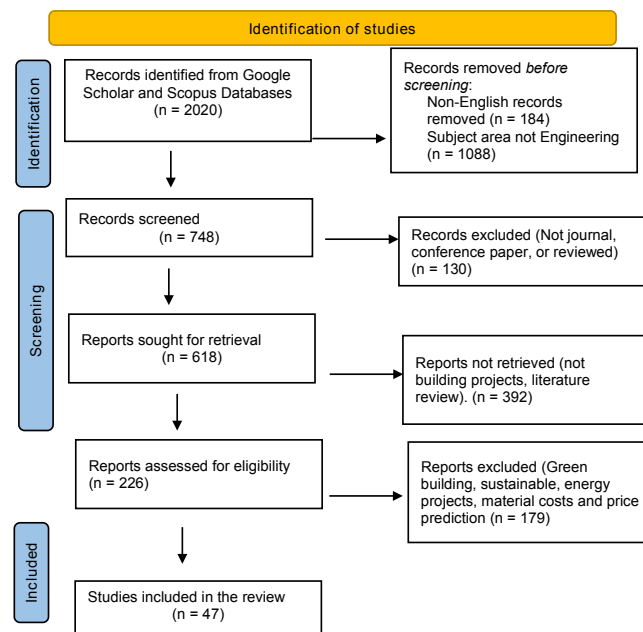


Figure 1: Articles review flow chart.

Results

The articles were systematically reviewed and summarised in Appendix A, Table 1, Table 2, Table 3, and Figure 2. Appendix A overviews the various studies that utilised supervised ML models. Appendix A shows that supervised ML models used an average of seven input variables, ranging from 4 to 64. Studies incorporated both project duration and costs or focused on costs only as the output variable. Sample sizes varied from 14 to 60,000, with residential and educational properties being the most examined types. As summarised in Table 1, 13 articles focused on residential properties, five on educational properties, two on commercial properties, and 10 on more than one property type. Most papers (62%) used multiple algorithms to estimate costs.

Table 1: Summary of supervised ML application for cost estimation.

| Property Type | ML Methods Used | Key Characteristics & Trends | Number of Studies |
|----------------------|-----------------------|----------------------------------------------------------------------|-------------------|
| Residential | ANN, MLR, SVM, CBR | Heavy reliance on ANNs and RA, often with small sample sizes (<200). | 13 |
| Commercial | ANN, SVM, DT, CBR | Includes comparisons between multiple techniques. | 2 |
| Educational | ANN, MLR, SVM, NN | Strong focus on ANNs and MLR. Datasets are small (30–217). | 5 |
| Mixed / Non-Specific | ANN, MLR, RF, XGBoost | Features the largest datasets and ensemble methods | 10 |
| Others | ANN, SVM, LR, LS | Limited studies available. | 4 |

Figure 2 is an illustration of the number of articles published per year. From 1998 to 2008, research on supervised ML was in its infancy, with 11 articles. From 2009 to 2018, the number of publications increased to 16. From 2018 to 2025, although 20 articles were published, the overall publication count remained low.

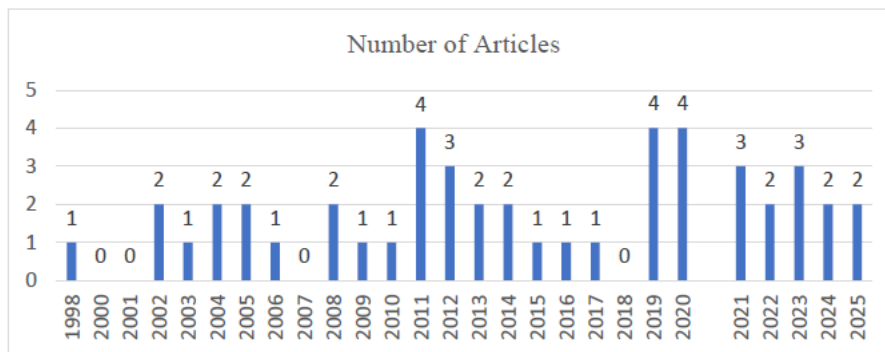


Figure 2: Number of publications per year.

Table 2 presents the distribution of research efforts by continent. Most of the studies (65.96%) are from Asia, with Europe at 14.89% and Africa at 10.64%. North America and Oceania both have 4.26%.

Table 2: Number of papers by geographical area.

| Continents | Frequency | Percentage |
|---------------|-----------|------------|
| N = 47 | | |
| Africa | 5 | 10.64% |
| Asia | 31 | 65.96% |
| Europe | 7 | 14.89% |
| North America | 2 | 4.26% |
| Oceania | 2 | 4.26% |

DISCUSSION

Geographically distribution revealed that Asia dominates the findings, accounting for 65.96%, highlighting the region's leading role in advancing research in cost estimation. Europe at 14.89% and Africa at 10.64%, while Oceania and North America each account for 4.26%, indicating a narrow research focus in these regions. Although the majority of publications on this subject are scholarly articles centred on projects in Asia, the findings demonstrate some progress in the utilisation of supervised ML for cost estimation. These findings suggest that to achieve more balanced global insights, future research should focus on currently underrepresented regions, such as Europe and Africa. Publication activity, although it shows gradual growth over the last two decades, remains modest overall: 11 articles from 1998–2008, 16 from 2009–2018, and 20 from 2019–2025. These suggest that while there is a growing interest in employing supervised ML techniques for cost estimation, the field is still in its early stages, presenting opportunities for further exploration. The literature analysis is consistent with Arage and Dharwardkar (2017), emphasising that the use of ML models for cost estimation has become increasingly popular in recent years. Methodologically, 62% of the studies applied multiple algorithms to improve result reliability, commonly drawing on regression-based methods, neural networks (including ANN variants and hybrids), support vector machines, decision trees, random forests, gradient boosting and kNN. ANN was popular on a small dataset across all property types. Regression models were utilised as a mathematical model to assess the linear relationship between a dependent variable and independent variables. Additionally, regression models were employed as benchmarks for the ANN model. The studies identified cost drivers for various building types, including residential, educational, industrial, health facilities, leisure (hotels), and commercial (retail and office) buildings. However, most studies focused on residential and educational property types, suggesting a potential gap in the exploration of other building types. This focus suggests that while supervised ML techniques have been effectively applied to estimate costs in these property types, there remains a significant opportunity to

extend these methodologies to less-studied building types, such as industrial, healthcare, leisure, and commercial buildings (retail and office).

CONCLUSION

Using a PRISMA-guided systematic review, the study examined research papers on the application of supervised ML cost estimation techniques in building projects, motivated by the central role of accurate cost estimation in construction success and the increasing limitations of traditional methods. The growing body of literature, as highlighted by a systematic review, demonstrates a clear trend toward the increased use of supervised ML models for cost estimation across various building types. To develop more balanced, globally relevant evidence, broader testing and application are necessary, especially beyond Asia and across a wider range of property types, including commercial (retail and office), industrial, healthcare, and leisure buildings.

REFERENCES

- Abd, A.M., Jasim, N.A. & Naseef, F.S. 2019. Developing a prediction model to predict the construction project cost by using multiple linear regression technique. *Indian Journal of Science and Technology*, 12(7).
- Alaloul, S. & Qureshi, A.H. 2021. Material classification via machine learning techniques: Construction projects progress monitoring. In: Mazzeo, P.L. & Spagnolo, P. (eds.) *Deep learning applications*. London: Intech Open.
- Ali, Z.H., Kassim, M., Burhan, A.M., Al-Khafaji, Z. & Afan, H.A. 2022. Developing an integrative data intelligence model for construction cost estimation. *Complexity*, 3:1–18.
- Alshemosi, A.M.B. & Alsaad, H.S.H. 2017. Cost estimation process for construction residential projects by using multifactor linear regression technique. *Criterion*, 6(6):151–156.
- Al-Tawal, D.R., Arafah, M. & Sweis, G.J. 2021. A model utilizing the artificial neural network in cost estimation of construction projects in Jordan. *Engineering, Construction and Architectural Management*, 28(9):2466–2488.
- Arafa, M. & Alqedra, M. 2011. Early-stage cost estimation of buildings construction projects using artificial neural networks. *Journal of Artificial Intelligence*, 4(1):63–75.
- Arage, S.S. & Dharwadkar, N.V. 2017. Cost estimation of civil construction projects using machine learning paradigm. *Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 594–599. IEEE.
- Badawy, M. 2020. A hybrid approach for a cost estimate of residential buildings in Egypt at the early stage. *Asian Journal of Civil Engineering*, 21(5):763–774.
- Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., Shringi, A. & Mendis, P. 2022. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141:104440.
- Bala, K., Ahmad Bustani, S. & Shehu Waziri, B. 2014. A computer-based cost prediction model for institutional building projects in Nigeria: An artificial neural network approach. *Journal of Engineering, Design and Technology*, 12(4):519–530.
- Chandanshive, V. & Kambekar, A.R. 2019. Estimation of building construction cost using artificial neural networks. *Journal of Soft Computing in Civil Engineering*, 3(1):91–107.

- Chen, G., Zheng, S., He, X., Liang, X. & Liao, X. 2025. Machine Learning-Based Cost Estimation Models for Office Buildings. *Buildings*, 15(11):1802.
- Cho, H.G., Kim, K.G., Kim, J.Y. & Kim, G.H. 2013. A comparison of construction cost estimation using multiple regression analysis and neural network in elementary school project. *Journal of the Korea Institute of Building Construction*, 13(1):66–74.
- Coffie, G. H., Aigbavboa, C. O. & Thwala, W. D. 2019. Modelling construction completion cost in Ghana public sector building projects. *Asian Journal of Civil Engineering*, 20(7):1063–1070.
- 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.
- Elhag, T.M.S. & Boussabaine, A.H. 1998. An artificial neural system for cost estimation of construction projects. In: W. Hughes (ed.) *Proceedings of the 14th Annual ARCOM Conference*, University of Reading. Association of Researchers in Construction Management, 1:219–26.
- El-Sawalhi, N.I. 2013. Important factors for construction project cost estimating using ANN. *Journal of Civil Engineering and Architecture*, 7(1):90.
- El-Sawalhi, N.I. & Shehatto, O. 2014. A neural network model for building construction projects cost estimating. *Journal of Construction Engineering and Project Management*, 4(4):9–16.
- Emsley, M.W., Lowe, D.J., Duff, A.R., Harding, A. & Hickson, A. 2002. Data modelling and the application of a neural network approach to the prediction of total construction costs. *Construction Management & Economics*, 20(6):465–472.
- Georgy, M. E. & Barsoum, S. H. 2005. Artificial neural networks model for parametric estimating of construction project costs. *Journal Of Engineering and Applied Science-Cairo*, 52(6):1050.
- Günaydin, H.M. & Doğan, S.Z. 2004. A neural network approach for early cost estimation of structural systems of buildings. *International Journal of Project Management*, 22(7):595–602.
- Gurmu, A. & Miri, M.P. 2023. Machine learning regression for estimating the cost range of building projects. *Construction Innovation*, 25(2):577–593.
- Hakami, W. & Hassan, A. 2019. Preliminary construction cost estimate in Yemen by artificial neural network. *Baltic Journal of Real Estate Economics and Construction Management*, 7(1):110–122.
- Hammad, A.A.A., Ali, S.M.A., Sweis, G.J. & Bashir, A. 2008. Prediction model for construction cost and duration in Jordan. *Jordan Journal of Civil Engineering*, 2(3):250–266.
- Holland, A.J., Cox, S.J. & Worthington, S. 2017. Artificial intelligence for construction scheduling. *Automation in Construction*, 81:28–38.
- Jiang, Q. 2020. Estimation of construction project building cost by back-propagation neural network. *Journal of Engineering, Design and Technology*, 18(3):601–609.
- Jin, R., Cho, K., Hyun, C. & Son, M. 2012. MRA-based revised CBR model for cost prediction in the early stage of construction projects. *Expert Systems with Applications*, 39(5): 5214–5222.
- Jones, M.T. 2008. *Artificial intelligence: A systems approach*. New York: Science Press LLC.
- Juszczuk, M. 2020. Development of cost estimation models based on ANN ensembles and the SVM method. *Civil and Environmental Engineering Reports*, 30(3):48–67.
- Kamiri, J. & Mariga, G. 2021. Research methods in machine learning: A content analysis. *International Journal of Computer and Information Technology*, 10(2):78–91.

- 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.
- Kim, G.-H., Shin, J.-M., Kim, S. & Shin, Y. 2013. Comparison of school building construction costs estimation methods using regression analysis, neural network, and support vector machine. *JBCPR Journal of Building Construction and Planning Research*, 1(1):1–7.
- Koo, C., Hong, T. & Hyun, C. 2011. The development of a construction cost prediction model with improved prediction capacity using the advanced CBR approach. *Expert Systems with Applications*, 38(7).
- Lathong, K. & Wisaeng, K. 2024. The prediction of low-rise building construction cost estimation using extreme learning machine. *Advances in Technology Innovation*, 9(1):12–27.
- Liu, Z. & Lin, J. 2025. Combinatorial machine learning approaches for high-rise building cost prediction and their interpretability analysis. *Journal of Asian Architecture and Building Engineering*, 24(4):2817–2828.
- Lowe, D.J., Emsley, M.W. & Harding, A., 2006. Predicting construction cost using multiple regression techniques. *Journal of Construction Engineering and Management*, 132(7):750–758.
- Luu, V.T. & Kim, S.Y. 2009. Neural network model for construction cost prediction of apartment projects in Vietnam. *Korean Journal of Construction Engineering and Management*, 10(3):139–147.
- Pejić, M.S., Terzić, M., Stanojević, D., Peško, I. & Mučenski, V. 2023. Improving construction projects and reducing risk by using artificial intelligence. *Social Informatics Journal*, 2(1):33–40.
- Petruseva, S., Zileska-Pancovska, V., Žujo, V. & Brkan-Vejzović, A. 2017. Construction costs forecasting comparison of the accuracy of linear regression and support vector machine models. *Technical Gazette*, 24(5):1431–1438
- Pham, T.Q.D., Quang, N.H., Vo, N.D., Bui, V.S. & Tran, V.X. 2021. Fast and Accurate Estimation of Building Cost Using Machine Learning. In: R. Kumar, N.H. Quang, V. Kumar Solanki, M. Cardona & P.K. Pattnaik (eds.) *Proceedings of the Research in intelligent and computing in engineering: Advances in intelligent systems and computing conference*, 1254:515–525. Singapore: Springer.
- Sanni-Anibire, M.O., Zin, R.M. & Olatunji, S.O. 2021. Developing a preliminary cost estimation model for tall buildings based on machine learning. *International Journal of Management Science and Engineering Management*, 16(2):134–142.
- Setyawati, B.R., Sahirman, S. & Creese, R.C. 2002. Neural networks for cost estimation. *AACE International Transactions*, 131.
- Shi, H. & Li, W. 2008. The integrated methodology of rough set theory and artificial neural network for construction project cost prediction. *2nd international symposium on intelligent information technology application*, 2:60–64.
- Sitthikankun, S., Rinchumphu, D., Buachart, C. & Pacharawongsakda, E. 2021. Construction cost estimation for government building using artificial neural network. *Int. Trans. J. Eng. Manag. Appl. Sci. Technol*, 12(6):1–12.
- Skitmore, R.M. & Ng, S.T. 2003. Forecast models for actual construction time and cost. *Building and Environment*, 38(8):1075–1083.
- 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.
- Sonmez, R. 2004. Conceptual cost estimation of building projects with regression analysis and neural networks. *Canadian Journal of Civil Engineering*, 31(4):677–683.

- Stoy, C., Pollalis, S. & Schalcher, H.R. 2008. Drivers for cost estimating in early design: Case study of residential construction. *Journal of Construction Engineering and Management*, 134(1):32–39.
- Tayefeh Hashemi, S., Ebadati, O.M. & Kaur, H. 2020. Cost estimation and prediction in construction projects: A systematic review on machine learning techniques. *SN Applied Sciences* 2(1703):1–27.
- Tandukar, D., Goh, C.J. & Chan, A.P. 2019. Artificial intelligence-driven project risk assessment. *Automation in Construction*, 99:333–345.
- Trijeti, T., Irwanto, R., Rahayu, T. & Panudju, A.T. 2023. Artificial neural networks for construction project cost and duration estimation. *Revue d'Intelligence Artificielle*, 37(6).
- Zhang, Y. & Fang, S. 2019. RSVRs based on feature extraction: a novel method for prediction of construction projects' costs. *KSCE Journal of Civil Engineering*, 23(4): 1436–1441.

Appendix A: Supervised ML for building project publications.

| Author | Technique | Project Type |
|----------------------------|----------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| Arafa & Alqedra 2011 | ANN (sample = 71, inputs = 7, output = cost) | Educational, Residential, & Public Buildings |
| Gurmu & Miri 2023 | DT, LR, KNN, GB, RF (sample = 60000, inputs = 5, output = cost) | Retail, Industrial, Public, Hospital, Residential & Commercial |
| Hakami & Hassan, 2019 | ANN (sample = 136, inputs = 17, output = cost) | Administration, Commercial, Educational, Residential, & Health Centre |
| Hammad et al., 2008 | MLR, LS (sample = 140, inputs = 5, output = cost & duration) | Hotel, Residential, Educational & Industrial |
| Jin et al., 2005 | MRA, CBR (sample = 150, inputs = 17, output = cost) | Commercial & Residential |
| Sanni-Anibire et al., 2021 | MLR, KNN, ANN, SVM (sample = 30, inputs = 8, output = cost) | Office, Residential & Hotels |
| Skitmore & Ng 2003 | RA (sample = 93, inputs = 6, output = cost & duration) | Residential, Educational, Industrial & Recreational |
| Ujong et al., 2022 | MLR, ANN (sample = 78, inputs = 6, output = cost & duration) | Residential, Commercial, Religious, Educational, Agricultural & Industrial |
| Wang 2012 | ANN, SVM (sample = 92, inputs = 3, output = cost & schedule) | Residential, Religious, Educational, Retail, Hospitals, Offices, Hotels & Recreational |
| Zhang & Fang 2019 | RF, SVM (sample = 30, inputs = 11, output = cost) | Commercial & Residential |
| Chen et al., 2025 | BPNN, SVM (sample = 110, inputs = 17, output = cost) | Commercial |
| Son et al., 2012 | SVM, RA, PCA, DT, ANN, MLR (sample = 84, inputs = 64, output = cost) | Commercial |

(Continued)

Appendix A: Continued.

| Author | Technique | Project Type |
|--------------------------|--------------------------------------------------------------|---------------|
| Cho et al., 2012 | MLR, ANN (sample = 96, inputs = 6, output = cost) | Educational |
| Elhag & Boussabaine 1998 | ANN (sample = 30, inputs 3&13, output = cost) | Educational |
| Georgy & Barsoum 2005 | ANN (sample = 30, inputs = 19, output = cost) | Educational |
| Kim et al., 2013 | RA, NN, SVM (sample = 217, inputs = 10, output = cost) | Educational |
| Setyawati et al., 2002 | NN (sample = 41, inputs = 7, output = cost) | Educational, |
| Alshemoni & Alsaad 2015 | MLR (sample = 20, inputs = 10, output = cost) | Residential |
| Badawy 2020 | ANN, MLR (sample = 174, inputs = 4, output = cost) | Residential |
| Chandanshive 2019 | ANN (sample = 78, inputs =11, output = cost) | Residential |
| Emsley et al., 2002 | RA, ANN, MLP, RBF (sample = 286, inputs = 25, output = cost) | Residential |
| Gunaydin & Dogan 2004 | ANN (sample = 30, inputs = 8, output = cost) | Residential |
| Ji et al., 2011 | CBR (sample = 164, inputs = 12, output = cost) | Residential |
| Jiang 2020 | MLR, ANN (sample = 10, inputs = 6, output = cost) | Residential |
| Juszczyk 2020 | ANN, SVM (sample = 162, inputs = 9, output = cost) | Residential |
| Kim et al., 2004 | MRA, NN, CBR (sample = 530, inputs = 12, output = cost) | Residential |
| Koo et al., 2011 | CBR, MRA, ANN (sample = 101, inputs = 8, output = cost) | Residential |
| Liu & Lin 2025 | DT, SVM, BPNN (sample = 135, inputs = 13, output = cost) | Residential |
| Luu & Kim 2009 | ANN (sample = 14, inputs = 7, output cost) | Residential |
| Stoy et al., 2008 | RA (sample = 70, inputs = 5, output = cost) | Residential |
| Dang-Trinh et al., 2023 | SVM, LR, ANN (sample = 35, inputs = 10, output = cost) | Industrial |
| Bala et al., 2014 | ANN (sample = 206, inputs = 9, output = cost) | Institutional |
| Waziri 2010 | ANN (sample = 510, inputs =9, output = cost) | Institutional |

(Continued)

Appendix A: Continued.

| Author | Technique | Project Type |
|----------------------------|----------------------------------------------------------------------------------------------|-----------------------|
| Lathong & Wisaeng 2023 | ANN, MLR, DT, Ensembles, SVM, RF (sample = not specified, inputs = 15, output = cost) | Low-Rise Buildings |
| Ali et al., 2022 | ANN, SVM, RF (sample = 90, inputs = 7, output = cost) | Public Buildings |
| Sonmez, 2004 | RA, ANN (sample = 20, inputs = 11, output = cost) | Retirement Facilities |
| Abd et al., 2019 | MLR, WLS (sample = 501, inputs = 25, output = cost) | Not Specific |
| Al-Tawal et al., 2020 | ANN (sample = 104, inputs = 27, 41, 53, output = cost) | Not Specific |
| El Sawalhi 2013 | ANN (sample = 106, inputs = 14, outputs = cost) | Not Specific |
| El-Sawalhi & Shehatto 2014 | ANN (sample = 169, inputs = 11, output = cost) | Not Specific |
| Lowe et al., 2006 | MLR (sample = 286, inputs = 41, output = cost) | Not Specific |
| Narbaev et al., 2024 | XGBOOST, LR, SVR, RF, LightGBM, CatBOOST, GGM, LGM (sample = 110, inputs = 4, output = cost) | Not Specific |
| Petruseva et al., 2017 | LR, SVM (sample = 75, inputs = 5, output = cost) | Not Specific |
| Shi & Li 2008 | ANN, RS (sample = 14, inputs = 6, output = cost) | Not Specific |
| Sitthikankun et al., 2021 | ANN (sample = 50, inputs = 11, output = cost) | Not Specific |
| Trijeti et al., 2023 | MLR, ANN (sample = 78, inputs = 6, outputs = cost & duration) | Not Specific |
| Pham et al., 2021 | ANN, RA, XGBoost, DT, KNN, SVM, RF, MLR (sample = 10000, inputs = 24, output = cost) | Not Specific |