IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning–Driven Construction Cost Contingency: A Conceptual Framework and Future Research Directions

[1]Madonna Nabil Roshdy *,[2]Ibrahim Abdel Rashid†
[1] Teaching Assistant and PhD Student, [2] Professor
[1] Department of Structural Engineering, [2] Department of Structural Engineering
[1]Ain Shams University, Cairo, Egypt, [2]Ain Shams University, Cairo, Egypt
*Corresponding Author † Deceased

Abstract: Given the subjective nature of traditional techniques using set percentages or expert judgment, cost contingency estimation remains one of the hardest tasks when it comes to project planning. The influence of project properties, geographical location, contractor capacity, and prior cost overrun tendencies can hardly be taken into consideration by such techniques. To the best of this author's knowledge, there has been no paradigmatic framework developed specifically for cost contingency estimation in Egypt, while contemporary research indicates the potential of artificial intelligence approaches for construction cost estimation. Thirteen predictive factors grouped into project factors, geographic factors, and contractor risk and capacity factors are proposed as the basis of a framework for cost contingency prediction through machine learning. Data gathering on projects, data pre-processing, selection of predictors, development of a machine learning model, and decision support constitute the four steps of the recommended approach. Predictors are used to estimate a cost contingency percentage using a conceptual artificial neural network. In addition, the current paper lists some key shortcomings of existing literature, including insufficient number of samples, varying validation methods, poor generalizability, and absence of interpretability. Finally, the paper suggests some new avenues of research such as artificial intelligence, building information modeling, and Monte Carlo simulations. The proposed paradigm serves as a systematic basis for future empirical research.

***Index Terms* -** Artificial neural network; Conceptual framework; Construction projects; Cost contingency; Machine learning.

I. INTRODUCTION

Despite significant improvements in planning, scheduling, and project control, construction projects suffer from considerable cost overruns. In many developing nations, particularly Egypt, the cost overrun issue is even more serious due to the impact of high inflation rates, fluctuating exchange rates, rise in steel and cement prices, shortage of workers, delayed approval processes, as well as changes in the scope of work (Flyvbjerg et al., 2003; Baccarini, 2005; El-Kholy et al., 2024). Consequently, estimating contingency amounts remains one of the most difficult and important challenges in the context of estimating building costs (Project Management Institute, 2021).

Traditional techniques for contingency estimation depend on pre-calculated ratios, subjective judgments, as well as deterministic estimates. For example, many consultants and contractors usually include 5% to 15% of estimated cost as contingency irrespective of the features and nature of the project (Hulett et al., 2008; Touran & Liu, 2012). Although such methods are rather easy to implement, they may not be sufficiently effective in terms of reflecting the uncertainty related to the project properly. In other words, underestimation and overestimation of contingency occur quite often leading to budget overruns or inflated budget estimates, respectively (Flyvbjerg et al., 2018; Project Management Institute, 2021).

Cost overruns in the construction business are another one of the oldest challenges. Indeed, based on multiple studies conducted internationally, there are many projects with higher costs compared to their initial projections due to improper cost estimation, inadequate risk management, inflation, project design changes, and the contractor's issues (Flyvbjerg et al., 2018).It should be noted that for project teams, it is standard practice to build in some reserve cost or contingency for the project to reduce its uncertainties. In this case, a contingency implies reserves allocated to cover unexpected yet predictable future events that occur during the course of a project (Baccarini, 2004; Hulett et al., 2008).

Conventional contingency estimation methods are primarily based on predetermined percentages or expert judgment. While, in practice, companies often tend to inflate the predicted project cost estimate by around 5–15% without considering the specific characteristics of a project (Baccarini, 2004). Nevertheless, even though these methods are easy to apply, they fail to reflect the actual risk profile, which leads to an insufficient or excessive contingency level. Likewise, expert judgment may be very dependent on knowledge about projects and their cost management, which could lead to inconsistent and subjective conclusions (Hulett et al., 2008; Project Management Institute, 2017).

Recently, artificial intelligence (AI) methods, especially machine learning (ML) algorithms, have been attracting greater attention among construction management researchers. ML models are preferable when it comes to estimating building costs since they can capture non-linear relationships between variables. In previous studies, artificial neural networks (ANN), support vector machines (SVM), random forests, gradient boosting, and their hybrids were used with promising results (Chakraborty et al., 2020; Kim et al., 2013; Mahmoodzadeh et al., 2022; Sonmez et al., 2007).Nevertheless, there are several limitations regarding ML research in this domain, such as limited data size, inconsistent variables, and low explainability (Bilal & Oyedele, 2020; Gill et al., 2024).

In addition, the Egyptian construction industry also encounters other obstacles stemming from the existence of high inflation rates, the volatility of the exchange rate of the Egyptian pound against foreign currencies, constant changes in the cost of materials, and differences in regions such as Greater Cairo, Upper Egypt, the coast, and new cities. All these issues significantly affect the construction costs and add to the uncertainty involved in contingency estimation (Ammar et al., 2023; Flyvbjerg et al., 2018).Therefore, there is an evident necessity for contingency estimation techniques that could be tailored to specific conditions in local markets. The current study intends to develop a ML-based contingency estimating technique for the Egyptian construction industry.

II. RESEARCH GAP AND NEED FOR A NEW FRAMEWORK

Despite prior research indicating that ML methodologies can improve the predictive accuracy of building cost contingencies and overruns, the current literature reveals significant methodological shortcomings. Numerous published studies depend on very small datasets, utilize inconsistent predictor variables, implement inadequate validation processes, and offer restricted interpretability of the constructed models (Bilal & Oyedele, 2020; Egwim et al., 2021; Gill et al., 2024). The methodological deficiencies diminish the robustness and dependability of ML-based techniques and impede their generalizability, especially when these models are applied in developing countries like Egypt, where project uncertainty and market volatility are significantly elevated.

The primary constraint noted in the literature pertains to the quantity and variety of the dataset. A significant number of prior research utilized ANN, SVM and regression-based models based on fewer than 100 to 300 completed construction projects (El-Kholy et al., 2022; Mahmoodzadeh et al., 2022; Sonmez et al., 2007). Although these datasets may be acceptable for exploratory research, they frequently fall short in encompassing the extensive diversity characteristic of building projects. Moreover, research concentrated on a singular project category or a confined geographical area, so further constraining the generalizability of their results. Thus, whereas these models may attain elevated training accuracy, their predictive efficacy frequently declines when utilized across varying project types or geographical areas. The proposed methodology prioritizes validation using a substantial and varied dataset that accurately reflects the Egyptian construction sector. A second notable issue pertains to the inconsistency in the selection of predictor variables. Prior research included diverse combinations of project cost, duration, procurement method, contractor expertise, project location, and economic variables such as inflation and exchange rates (Gill et al., 2024; Sonmez et al., 2007; Touran & Liu, 2015).The lack of a uniform variable selection framework hinders comparative analysis among studies and restricts the methodical collection of knowledge in this field. The proposed approach addresses this issue by standardizing thirteen predictor variables categorized into three groups: project qualities, geographical considerations, and contractor capability indicators. This methodical approach seeks to improve model openness, consistency, and repeatability.

Inadequate validation processes constitute a significant methodological deficiency in current ML-based cost prediction studies. Numerous investigations have exclusively reported training accuracy or utilized restricted testing methodologies, failing to include independent validation datasets or rigorous resampling approaches (Bilal & Oyedele, 2020; Xu et al., 2021).Such techniques elevate the likelihood of overfitting and produce too positive performance assessments. The suggested approach employs a thorough validation technique that incorporates distinct training, validation, and testing datasets, alongside k-fold cross-validation and early stopping to enhance model generalization.

III. PROPOSED FRAMEWORK AND METHODOLOGY

3.1 Research Design and Conceptual Framework

A systematic methodological framework is needed to develop a reliable and general model to estimate construction cost contingencies, which can combine theoretical understanding and empirical validation. This study used a hybrid research methodology that combines a systematic synthesis of literature with a data-driven artificial intelligence modeling. In this way, the proposed framework is based on well-established scientific principles and verified by empirical data of construction

The conceptual framework was designed as a multi-phase approach that parallels the typical life cycle of machine learning model development in building research. Unlike traditional linear techniques, the framework is not static, but rather iterative, enabling continuous improvement of the model based on performance feedback. The iterative nature is especially important in artificial neural network (ANN) modeling because model accuracy is highly dependent on the quality of the data, the network architecture, and the training method.

The complete framework consists of five stages namely data collection, data preprocessing, model construction, model validation and evaluation and decision-support integration. Figure 1 presents the conceptual research framework developed for this study. The stages are not necessarily sequential but interlinked, allowing the possibility of feedback loops between the assessment of the model performance and earlier stages such as preprocessing and variable selection. The iterative procedure ensures that the final model is not just a single configuration but the outcome of a systematic optimization across several scenarios.

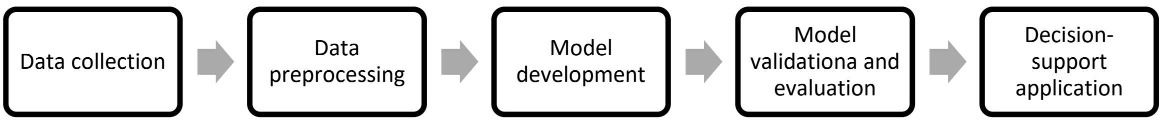

Fig. 1. Proposed research framework for machine learning-based contingency estimation

3.2 Data Collection and Dataset Characteristics

The research is based on a data set of 465 completed building projects in the Egyptian construction sector. The data set was compiled from several sources including contractor records, consultancy archives and historical project paperwork. The dataset was constructed such that a wide variety of project conditions were included to improve the generalizability of the developed model. The collection consists of several types of buildings mainly residential and commercial buildings in different geographic areas in Egypt. This diversity is important, as construction costs and risks vary significantly depending on local factors like labor availability, logistics, infrastructure and market conditions(Bilal & Oyedele, 2020).

The data set's dependent variable is the percentage of cost contingency, which is the additional allocation needed to mitigate the uncertainty and risk in building projects. This variable is based on the real results of the project, not assumptions, so it is a realistic and reliable target for training the model. The use of observed contingency values is consistent with best practice in cost-risk modelling, and enhances the credibility of the predictive framework (Baccarini, 2004; Hulett et al., 2008).

The strength of this research is that it makes use of a large, diverse and experimentally verified data set, unlike many previous studies that have relied upon smaller and more homogenous data sets. Previous studies on machine learning-based cost and contingency prediction have often relied on limited datasets specific to a particular type of project or location, limiting external validity and increasing the risk of overfitting (El-Kholy et al., 2022b; Mahmoodzadeh et al., 2022; Sonmez et al., 2007) . The broad coverage and detailed structure of the current dataset greatly increase the robustness of the model and allow for reliable generalization across different project environments in the Egyptian construction sector.

3.3 Classification and Selection of Input Variables

The selection of input variables is a key factor in determining the performance of a model in machine learning applications. In this study, a comprehensive methodology was followed, which combined literature evaluation, expert consultation, and data availability limitations to identify thirteen input variables. The objective was to determine the selected variables as the main determinants of contingency, without losing consistency and reliability in the data set (Ammar et al., 2023). To enhance interpretability and analytical clarity, the variables were grouped into three main categories, namely project related variables, location and temporal variables and contractor capability variables(Sonmez et al., 2007).

Project variables are the characteristics of construction projects. The base cost, project duration, project type, built up area, contract type, procurement method and design maturity were the selected project variables. The characteristics are commonly recognized to be the main determinants of the complexity and uncertainty of a project as per interviews conducted with the industry's experts. It is more vulnerable to economic volatility in projects with longer durations and scope changes and cost overruns in projects with low design maturity (Flyvbjerg et al., 2018).

Geographical and time variables were included to account for external effects on building costs. Project location was used to account for regional variations in construction practices and market conditions. The year of the start of the project can be a proxy for the state of the economy during the implementation of the project e.g. inflation, changes in prices of materials (Hoseini et al., 2020).

Contractor competence variables are the execution and managerial efficacy hazards. These include the contractor's rank which was measured by the grade of the contractor according to the Egyptian Engineering Syndicate, the number of such projects carried out, past overruns of cost and reliance on sub-contractors. The integration of these three categories guarantees that the model considers the entire spectrum of contingency factors, integrating technical, economic, and managerial aspects into a single analytical framework (Gill et al., 2024).

3.4 Data Preprocessing and Preparation

The preparation of data is a crucial step in the development of machine learning models, particularly in the construction industry where data sets are usually heterogeneous, incomplete and influenced by real uncertainties. Previous research has demonstrated that the quality and uniformity of the input data significantly affect the accuracy, stability, and generalization capacity of ANNs (Bilal & Oyedele, 2020). In the present study, the dataset was pretreated in a thorough and systematic way to check its suitability for ANN-based modeling and its potential to yield reliable predictions.

The first step of the preparation stage was data cleansing for data integrity improvement and bias reduction. Duplicate recordings were detected and removed to avoid skewing of the learning process by repetitive observations. The dataset was then examined for missing or incomplete entries. Each case was reviewed, instead of imputing for missing values automatically. Records without critical variables to train the model were excluded, and small discrepancies were solved with logical validation against project documentation. This selective methodology follows the best practices in construction data analytics as inaccurate imputation may create spurious patterns and influence model reliability (Bilal&Oyedele, 2020).

After data cleansing, outlier detection was performed to identify extreme values that may affect the model efficacy. Outliers are prevalent in construction datasets because of variations in project size, complexity and exposure to risk. However, past work has shown that not all extreme events are indicative of data errors as some may be associated with truly high risk or abnormal operations (Hulett et al., 2008). Therefore, each outlier was considered individually. Valid extreme observations were kept, while erroneous ones due to data entry errors were corrected or deleted. Genuine outliers are kept so that the model can learn from the whole spectrum f project conditions, which is particularly important for contingency estimates. Another important preprocessing step was to convert the category variables into numerical formats. ANNs can only accept numerical inputs, so categorical variables such as project type, contract type and project location are encoded by suitable methodologies.

Encoding method was applied to preserve the inherent characteristics of each variable and to avoid the generation of spurious ordinal correlations. Previous work on machine learning for buildings has shown that encoding categorical variables in an inappropriate way can bias learning and degrade model interpretability (Xu et al., 2021). Then all numerical variables were standardized with the min-max normalization method. This technique normalizes the input variables to a scale of 0 to 1. This avoids the dominance of aspects with large numerical values (e.g. base cost) over aspects with smaller values (e.g. contractor experience or design maturity. The improvement of the convergence velocity, numerical stability, and prediction accuracy of artificial neural networks for regression-based applications following normalization has been well documented.

Normalization and scaling enable you to compare variables meaningfully, by getting them on the same scale. This becomes important when you want to bring together disparate sources of construction data. Ensuring that all input variables are equally important to the learning process helps in model resilience and reliable prediction outcomes (Gill et al., 2024). The preprocessing methods used in this work aimed at striking a balance between data integrity and model flexibility. The data was carefully cleaned for preparing the dataset for ANN-based modeling, outliers were appropriately treated, categorical encoding was optimized and normalization was done thoroughly. The constructed model needs systematic preprocessing to learn the complex relationships between project attributes and contingency requirements, which enhances the predictive accuracy and generalization ability.

3.5 Design of Artificial Neural Network Model

ANNs were selected as the primary modelling method in this study because they have been demonstrated to be able to describe complex non-linear interactions between multiple variables. Traditional regression models are constrained by a priori linear assumptions while ANNs are data-driven and can find complex patterns from the data themselves. This makes them particularly suitable for estimating construction cost contingencies where interactions of project variables are nonlinear, complex and highly uncertain (Kim et al., 2013). This study assessed different ANN architectures on their suitability for contingency estimation. The first architecture studied was a two-layer feed-forward neural network, comprising a hidden and an output layer. One important point to note is that conventionally the input layer is not considered to be a computational layer as it does not perform any weighted transformations or learning, but allows input variables to be fed into the network. Hence, a network with one hidden layer and one output layer is often called a two-layer feedforward network (Goodfellow et al., 2016).

Together with this construction a three-layer feedforward neural network was also studied. This design has two hidden layers and one output layer. This allows for deeper models and better representational power. Deep feedforward networks learn complex hierarchical features by progressively transforming data through a large number of nonlinear layers. However, increasing the depth of the network increases the risk of overfitting, especially when the dataset is limited, and may result in higher computational costs without proportional gains in predictive power. Taking these points into account, the two network architectures were systematically evaluated in terms of the trade-off between model complexity and generalization capability. Thus, the comparison study led to the selection of the best ANN architecture for estimating the construction cost contingency which shows resilience and computing efficiency. To overcome this trade-off, different network architectures were evaluated to find the most suitable architecture.

The baseline architecture chosen for this study is a 13–14–1 configuration which means there are 13 neuros in the input layer, 14 neurons in the hidden layer and one neuron in the output layer. The 13-14-1 architecture provided a good trade-off between model accuracy and computational efficiency. This architecture follows the accepted design principles of ANN, which implies that the number of hidden neurons should be similar to the number of input variables for moderately big datasets (Haykin, 2009).

In the hidden layer, a tangent sigmoid (tansig) activation function is used to add nonlinearity and enhance the network's ability to model complex interactions. The output layer has a linear (purelin) activation function as the expected contingency percentage is a continuous value. This combination of activation functions is widely used in ANN-based regression applications and is well accepted in the literature of construction cost estimation (Juszczyk, 2017). The ANN architecture was purposefully constructed to possess high predictive ability and maintain generalization ability. This enabled the precise modeling of different factors of contingency estimations in construction projects.

3.6 Model Optimization and Training Algorithms

The efficacy of predictions of an artificial neural network is heavily influenced by the training technique used in the optimization of network weights (Xu et al., 2021). The study considers three common training algorithms, Levenberg–Marquardt (trainlm), Bayesian Regularization (trainbr) and Scaled Conjugate Gradient (trainscg). The comparison of different training methods allowed a comprehensive comparison and reduced the possible bias by relying on a single optimization approach.

The Levenberg-Marquardt algorithm is known to converge rapidly and very accurately for feedforward neural networks trained on medium size data sets. Trainlm is a popular ANN training algorithm in engineering applications (Hagan & Menhaj, 1994) that combines the benefits of gradient descent and Gauss-Newton. However, it can be more prone to noise and therefore more likely to overfit on datasets with lots of fluctuations (Bilal & Oyedele, 2020).

Bayesian Regularization incorporates probabilistic ideas into the training process, thus adding a penalty on network complexity. This approach constrains the weight from expanding too much, which helps to relieve overfitting and improve the model generalization. Bayesian Regularization has proved to be very effective in noisy or uncertain context such as building cost models, where there is a great variability in data (Ammar et al., 2023).

The Scaled Conjugate Gradient algorithm requires less memory and has better computing performance than Levenberg-Marquardt. Trainscg is useful for larger datasets or limited computational resources, though it typically converges slower (Xu et al., 2021).
The optimal model configuration was identified using an optimization method based on scenarios. A number of ANN models were developed systematically varying network architecture, number of hidden neurons and training methodology (Mahmoodzadeh et al., 2022). All scenarios are evaluated on the same datasets and performance measures for an objective comparison. This systematic and experimental approach is in keeping with best practice in machine learning model development, where model selection is informed by empirical data rather than subjective opinion (Goodfellow et al., 2016).

3.7 Model validation and performance evaluation

To ensure robustness and generalization capabilities, the dataset was split into training, validation and testing subsets. The training dataset was used to tune the network weights, while the validation dataset was used to evaluate the performance during training and to reduce overfitting. The testing dataset was used only for the final model testing, providing an unbiased estimate of predicted accuracy. The three-way data partitioning technique is a common approach in machine learning applications because it increases objectivity and facilitates reliable evaluation of generalization (Goodfellow et al., 2016).

Model performance was evaluated using various statistical metrics such as correlation coefficient (R), mean squared error (MSE), root mean squared error (RMSE) and mean absolute percentage error (MAPE). R measures the strength of the relationship between expected and actual values, while the error metrics measure the magnitude and distribution of prediction errors. Different metrics give a complete picture of how well the model is performing. Each metric is a measure of different aspects of accuracy. MSE penalizes large errors more. MAPE gives practitioners a comprehensible percent error (El-Kholy et al., 2022).

A good model should have high correlation coefficients, low error magnitudes and similar performance on training, validation and testing datasets. The agreement between different datasets is important. It shows that the model has learned generalizable trends, not simply memorizing the training data. Model performance and possible bias were examined visually using regression plots and error distribution analyses (Chakraborty et al., 2020).

3.8 Decision Support Integration

The last stage of the proposed framework is the integration of the developed ANN into a decision support system suitable for practical use in construction project management. The main aim of this integration is to offer contractors a data-driven tool that can evaluate the contingency values of projects based on certain project characteristics. The proposed ANN model allows for dynamic contingency estimation by considering multiple contributing factors simultaneously, in contrast to traditional approaches of contingency estimation based on fixed percentages or subjective judgments.

The user inputs important project parameters such as cost, length, location and contractor competency and receives a contingency recommendation that is aligned with the inherent risk profile of the project. Along with direct prediction, the model can be used for sensitivity analysis to determine the effect of specific variables on the expected outcome. This aspect increases transparency and allows decision makers to grasp the importance of critical factors and increases confidence in model outputs (Gill et al., 2024).

AI in integration of the building decision-support system is a huge step forward in cost management practices (Hammad et al., 2016; Hoseini et al., 2020; Fateminia & Fayek, 2023).. Due to the accuracy of cost estimation, better risk assessment, and informed decision making, AI-based technologies reduce cost overruns and improve project performance (Ammar et al., 2023).The suggested approach serves as an effective bridge between academic research and professional practice, transforming sophisticated prediction models into useful decision support tools for building applications.

IV. ACKNOWLEDGMENT

The authors respectfully acknowledge that Prof. Ibrahim Abdel Rashid contributed substantially to the conceptual development and supervision of this work during its early stages. Prof. Ibrahim Abdel Rashid passed away prior to the submission of this manuscript. The authors further confirm that consent for publication of this work has been obtained from the family of the late Prof. Ibrahim Abdel Rashid. The remaining author takes full responsibility for the content of this paper.

REFERENCES

1. Ammar, T., Abdel-Monem, M., & El-Dash, K. (2023). Appropriate budget contingency determination for construction projects: State-of-the-art. *Alexandria Engineering Journal*, *78*, 88–103. https://doi.org/10.1016/j.aej.2023.07.035
2. Baccarini, D. (2004). Estimating project cost contingency-a model and exploration of research questions. In *Heriot Watt University. Association of Researchers in Construction Management* (Vol. 1).
3. Bilal, M., & Oyedele, L. O. (2020). Guidelines for applied machine learning in construction industry—A case of profit margins estimation. *Advanced Engineering Informatics*, *43*, 101013. https://doi.org/10.1016/j.aei.2019.101013
4. 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*. https://doi.org/10.1016/j.aei.2020.101201
5. Egwim, C. N., Alaka, H., Toriola-Coker, L. O., Balogun, H., & Sunmola, F. (2021). Applied artificial intelligence for predicting construction projects delay. *Machine Learning with Applications*, *6*, 100166. https://doi.org/10.1016/j.mlwa.2021.100166
6. El-Kholy, A. M., Tahwia, A. M., & Elsayed, M. M. (2022). Prediction of simulated cost contingency for steel reinforcement in building projects: ANN versus regression-based models. *International Journal of Construction Management*, *22*(9), 1675–1689. https://doi.org/10.1080/15623599.2020.1741492
7. Fateminia, S. H., & Fayek, A. R. (2023). Hybrid fuzzy arithmetic-based model for determining contingency reserve. *Automation in Construction*, *151*. https://doi.org/10.1016/j.autcon.2023.104858
8. Flyvbjerg, B., Ansar, A., Budzier, A., Buhl, S., Cantarelli, C., Garbuio, M., Glenting, C., Holm, M. S., Lovallo, D., Lunn, D., Molin, E., Rønnest, A., Stewart, A., & van Wee, B. (2018). Five things you should know about cost overrun. *Transportation Research Part A: Policy and Practice*, *118*, 174–190. https://doi.org/10.1016/j.tra.2018.07.013
9. Gill, E. Z., Cardone, D., & Amelio, A. (2024). Revolutionizing the construction industry by cutting edge artificial intelligence approaches: a review. In *Frontiers in Artificial Intelligence* (Vol. 7). Frontiers Media SA. https://doi.org/10.3389/frai.2024.1474932
10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://www.deeplearningbook.org/
11. Hammad, M. W., Abbasi, A., & Ryan, M. J. (2016). Allocation and Management of Cost Contingency in Projects. *Journal of Management in Engineering*, *32*(6). https://doi.org/10.1061/(asce)me.1943-5479.0000447
12. Haykin, S. (2009).*Neural networks and learning machines* (3rd ed.). Pearson Education.
13. Hoseini, E., Bosch-Rekveldt, M., & Hertogh, M. (2020). Cost Contingency and Cost Evolvement of Construction Projects in the Preconstruction Phase. *Journal of Construction Engineering and Management*, *146*(6). https://doi.org/10.1061/(asce)co.1943-7862.0001842
14. Hulett, D., Hornbacher, K., & Whitehead, M. B. A. W. T. (2008). Project cost risk analysis: The risk driver approach prioritizing project risks and evaluating risk responses. *Hulett & Associates*.
15. Juszczyk, M. (2017). The use of artificial neural networks for residential building conceptual cost estimation. *Journal of Civil Engineering and Management*, *23*(4), 483–493.
16. 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. *Journal of Building Construction and Planning Research*, *01*(01), 1–7. https://doi.org/10.4236/jbcpr.2013.11001
17. Mahmoodzadeh, A., Nejati, H. R., & Mohammadi, M. (2022). Optimized machine learning modelling for predicting the construction cost and duration of tunnelling projects. *Automation in Construction*, *139*, 104305. https://doi.org/10.1016/j.autcon.2022.104305
18. Sonmez, R., Arif Ergin, ;, & Talat Birgonul, M. (2007). *Quantitative Methodology for Determination of Cost Contingency in International Projects*. https://doi.org/10.1061/ASCE0742-597X200723:135
19. Project Management Institute. (2017, May). Project Manager Competency Development Framework–Third Edition. Project Management Institute.

20. Touran, A., & Liu, J. (2015). A Method for Estimating Contingency Based on Project Complexity. *Procedia Engineering*, *123*, 574–580. https://doi.org/10.1016/j.proeng.2015.10.110
21. Xu, Y., Zhou, Y., Sekula, P., & Ding, L. (2021). Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment*, *6*(6), 100045. https://doi.org/10.1016/j.dibe.2021.100045