

Duration Estimation of Steel Framing Activities Using an Artificial Intelligent Algorithm and Top Risk Scores for High-Rise Buildings

Dima Jawad

Associate Professor, Engineering Management Graduate Program, Faculty of Engineering and Applied Science
Ontario Tech University
Oshawa, Ontario, CANADA

Seiyed Hamid Mousavi

Master of Engineering in Engineering Management, Engineering Management Graduate Program, Faculty of Engineering and Applied Science
Ontario Tech University
Oshawa, Ontario, CANADA

Abstract

Underestimation or overestimating activities' duration in construction projects leads to many problems like lack of funding or procrastination. However, estimating duration accurately for construction projects is difficult because of many factors like identified and unidentified negative risks. This paper tried to estimate realistic durations of steel framing activities in a high-rise building using a combination of Artificial Neural Networks (ANNs) and fuzzy logic algorithms with data from five similar projects. Initial planned durations and the top five pure risks are the inputs, and estimated actual durations are the outputs. The data from the first four projects was for training, and the fifth one was for simulation. Then, the estimated expected actual durations under uncertainties are compared to the actual durations of the project. The results showed that about 75% of the fifth project's activities had errors of less than three days compared to the actual durations.

Keywords

Risk, Steel Framing Activities, Neural Networks and Construction.

1. Introduction

More accurate duration estimation in construction projects is vital because time is crucial for all other project activities like cost estimation (time is money) and, most likely, completion date. Making mistakes in estimating activities' duration in construction projects leads to liquidated damages, disputes, lack of funding, lack of resources and some things else for the contractors (Cheung and Chow, 2011; Wambeke and Liu, 2011). Also, duration overestimating will lead to a decrease in the productivity rate of workers and other project team members (Park et al., 2010; Kim et al., 2011), and this prolongation may cause damage to equipment, resources, and materials (Le-Hoai et al., 2008; Chong and O'Connor, 2011; Kanoglu, 2003). Therefore, both under/overestimating would have a negative effect on the project outcomes and initial objectives.

One of the crucial parts of high-rise steel structure buildings is activities related to the installation of columns and beams for floors (Rashid et al. 2016). The weight of this part for progress measurement of the project is varied, but at least 25% of the total weight of projects is for steel framing activities. Also, these activities are in the early phase of the execution phase and are on the critical path of the project; therefore, prolonging the unrealistic planned duration of these activities in the baseline schedule will cause a delay in the completion date and, consequently, the liquidated damages. Consequently, a lack of information may affect the whole project duration in the early phase of construction projects. Based on the contract provisions, contractors may have to create a recovery plan with extra cost. Therefore, this research will focus on the duration of activities related to installing steel framing.

However, an accurate and realistic duration estimation for construction projects is crucial because of many factors like the project's size, location, construction method and especially identified and unidentified risks (Hegazy and Ayed, 1998). To calculate activity duration accurately in construction projects, much research has been done in the past. A discrete-event model using fuzzy logic was presented by (Zhang et al., 2005). A fuzzy algorithm was utilized to determine the risk-associated activity's duration (Ock and Han, 2010). A regression analysis approach for duration estimation was developed by (Waziri et al., 2017) using some relevant predictor variables. In 2011, Maravas and Pantouvakis illustrated a fuzzy project scheduling model for the duration estimation of activities.

Because of some problems and difficulties like fuzzy sets defining, consideration of appropriate rules with difficulty in their interpretation and applying of the non-existence of unique or best fuzzy membership functions and set of rules and complexity of models due to many variables, many researchers, like (Sonmez and Rowings, 1998; Ezeldin et al., 2006; Jung et al., 2022; Isah and Kim, 2022), changed their attentions from fuzzy logic system to the Artificial Neural Networks (ANNs) as a predictor. Some articles used neural networks to forecast the cost and prices of materials and equipment for the projects (Mir et al., 2021; Han et al., 2020).

Petruseva et al., (2012), utilized a neural network approach to estimate the duration of a construction project. From the past data, they could just predict the whole duration of the project. Golizadeh et al., (2015) developed four neural network algorithms for duration estimation of construction projects for projects in tropical region countries. They created a pool of variables from past projects like the cost of the whole project, contractors, working time etc. their model is quite complicated with complex interrelationships. Elfahham (2019), utilized neural networks to predict the cost of construction projects. Matel et al., (2022) tried to estimate the cost of engineering services for construction projects using an artificial neural network approach. They considered some variables like profit margins and market shares as inputs to the network.

Here, for training and simulation, actual data from 5 high-rise residential buildings (four of them for training and one for simulation) will be used. The projects were completed from 2014 to 2022. All five projects to some extent like one another and were completed by the same general contractor. Most of the subcontractors, especially subcontractors of steel framing, were also the same. The location of all projects was in the same zone of the city. Also, most of the work packages in the work breakdown structure and construction methods of the steel structure parts of the projects were exactly the same, but the differences were in size and number of floors and some utilities. These features help us to have a better and more realistic evaluation.

The proposed algorithm selected initial planned durations and the top five pure risks as inputs, and estimated actual durations were outputs. To check the algorithm's performance, the network outputs are compared to the actual durations of the project.

1.1 Objectives

In this study, the main purpose is to predict the duration of the activities near the actual ones using a combination of the above techniques, Artificial Neural Networks (ANNs) and fuzzy logic algorithm, based on the input risk scores and the initial planned duration. This simple method helps companies with good historical data from similar projects prepare better and more realistic schedule baselines based on external and internal conditions.

2. Data Collection

In the current research, inputs are five top risk scores based on past data and expert judgment. The real risk assessment in the past considered projects and initially planned durations for the first four projects and activity ID. The output is

the actual duration of the considered activities. The total number of activities is 563, and they are divided into 79 categories. Each category has the same ID. In other words, the activities like each other from the different projects have one ID number. It helps the network to train more efficiently and have a better prediction. For example, the activity “installation of the seventh stage-zone one-first project” has the same ID as the activity “installation of the seventh stage-zone two-first project” and the activity “installation of the seventh stage-zone one-second project”. Because the nature, installation methodology and the size of activities to some extents are the same. The specification of the five projects is shown in Table 1. Statistical description of the training data for the inputs and the outputs is shown in Table 2. Because of suspension in one of the projects, about 20 activities had abnormal durations; therefore, they were removed from the data.

Table 1. Specification of the five projects

Project Number	Sq.ft.	Number of floors
1	~790000	21
2	~700000	17
3	~630000	17
4	~750000	18
5	~720000	19

Table 2. Statistical description of the training data

Items	Planned Values				Actual Values			
	Min	Max	Mean	Std. Deviation	Min	Max	Mean	Std. Deviation
Installation of the first stage	49	65	55.95	4.45	56	82	69.5	7.54
Installation of the second stage	40	60	50.87	5.46	54	79	64.79	6.95
Installation of the third stage	43	80	57.17	11.42	56	98	73.17	14.38
Installation of the fourth stage	40	56	47.58	4.66	53	72	60.7	6.14
Installation of the fifth stage	40	57	45.53	4.97	50	71	58.42	5.91
Installation of the sixth stage	39	65	47.68	6.71	45	86	60.71	8.93
Installation of the seventh stage	40	80	49.61	10.03	51	98	61.61	12.14
Installation of the eighth stage	39	56	47.45	4.7	51	72	60.81	5.77
Installation of the ninth stage	39	57	48.28	4.44	53	73	62.38	5.49
Installation of the tenth stage	45	63	55.26	4.45	56	82	69.1	7.53
Others	4	27	14.84	7.18	5	31	18.84	7.94

One general contractor constructed the projects, and to some extent, each discipline of the buildings was also completed by the same subcontractor. The subcontractor related to the steel framing activities was a professional company with over twenty-five years of work experience in this field. The general contractor joined it to do the steel framing works fifteen years ago.

As regards top risks and corresponding scores, for each project at execution time, risk review meetings and schedule review analysis were held each month to reassess the identified risks and identify probable new risks. Based on the assessment at that time, due to the real conditions, the five risks were selected as the top risks for framing activities and quantified. These risks were the same for each project, with different risk scores due to different probability and impact for each risk because of different conditions like economic or weather at execution time. The risk statement for the considered top risks is shown in Table 3.

Table. 3. The risk statement for the top risks

No.	Risks	Cause	Impact
1	Delay in place order	Due to lack of funding because of sanctions	Delay in execution and prolongation of activities
2	Delay in manufacturing of columns and beams	Due to lack of proper material in the market because of high inflation rate	Delay in execution and prolongation of activities
3	Rework	Due to make mistakes in welding and alignment	Delay in execution and prolongation of activities
4	Site shut down	Because of HSE issues	Delay in execution and prolongation of activities
5	Site shut down	Because of adverse weather conditions	Delay in execution and prolongation of activities

In this paper, the total number of set data is 563, which includes the number of activities. Each set has six features: initial planned duration and five scores corresponding to each risk.

3. Methods

As mentioned earlier, because of the advantages of ANNs and fuzzy methods, both approaches were considered alternative approaches to predict the probable duration in this paper. To accomplish this, data from four high-rise buildings were considered to train the network. A Schematic of the proposed artificial intelligence algorithm for this study is shown in Figure 1, which comprises two parts. The first part is a direct model comprising a recurrent neural network with eight inputs and one output. The second part is a fuzzy algorithm which works as a regulator (Figure 1).

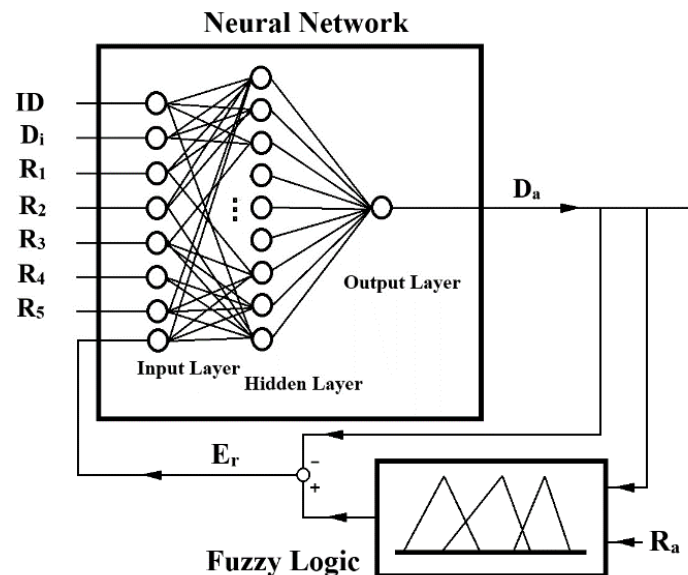


Figure 1. Schematic of the proposed artificial intelligence algorithm.

In the first, the network was trained off-line to update the weights to create the desired model and faster convergence during online training. Inputs to this network are the activity ID (ID), initial planned duration (Di), the score of the first risk (R₁), score of the second risk (R₂), the score of the third risk (R₃), the score of the fourth risk (R₄), score of the fifth risk (R₅) and error (E_r) while the output is probable actual duration (D_a). Similar activities from different projects have the same ID, which helps the network identify the types of activities. The error, feedback to the network, is the difference between the neural network's output and the fuzzy logic output in each iteration.

Here, the input layer has seven nodes, and the output layer has one node. One layer is considered the hidden layer. For multilayer neural networks, finding the appropriate number of nodes for hidden layers to have more accurate predictions is very important and challenging. Both insufficient and many nodes for hidden layers will cause instability and overfitting in the network. Both issues can reduce the neural network's generalization ability (Flores, 2011).

Although choosing the number of nodes for hidden layers can be done via trial and error, some heuristic rules are also introduced based on the number of input/output nodes (Panchal et al., 2011). Note that number of nodes can be tuned based on the performance. Two of the heuristic rules are presented:

- Double the number of input nodes plus one can be considered as the number of nodes for hidden layers.
- The sum of the input nodes and output nodes can be considered as the number of nodes for hidden layers.

According to the above rules and trial and error approach, for the network 8, 9 and 10 nodes were considered and the network performance for each number was compared to one another using the Mean Absolute Percentage Error (MAPE). The number corresponding to the best performance was selected as the number of nodes for the hidden layer. MAPE closer to zero was considered as the better performance. The following equation calculates MAPE:

$$MAPE = \frac{1}{n} \sum \left| \frac{A_n - P_n}{A_n} \right| \times 100 \quad (1)$$

where P_n and A_n are the predicted outputs (expected actual duration) and targets, respectively. The net input to node k (net_k) and output value from the node k (O_k) work out via the following formulations.

$$net_k = \sum w_{jk} O_j \quad (2)$$

$$O_k = f(net_k + \theta_k) \quad (3)$$

where w_{jk} stands for the weight among the two nodes in the previous and current layers (jth and kth). θ_k stands for the bias of the kth node. Here, f is the activation function for each node. Activation functions for input and output layers are selected linear functions, while these for the hidden layers' nodes are nonlinear functions. In this structure, logarithmic sigmoid transfer function is selected as the nonlinear function.

$$O_k = f(net_k + \theta_k) = \frac{1}{1 + e^{-(net_k + \theta_k)}} \quad (4)$$

The linear function for the input/output layers is:

$$O_k = f(net_k + \theta_k) = net_k + \theta_k \quad (5)$$

To train the network, it is necessary to feed it with input data. After using functions for calculations, a comparison between the output values and the target values will be done. To have more smooth data and better training, all sets were normalized and then fed to the network. In each iteration, the differences among the output values and the target values will be analyzed and propagated to the network to update the weights and biases. It helps the network to have output values closer to the desired values. This procedure will continue for all the input data sets till satisfying the defined minimum threshold for error convergence. Here, using the Mean Square Error (MSE) the appropriate value for the threshold was defined as 1.5. In this study, the LM method is selected to train the network.

$$w^{i+1} = w^i - \left[\frac{\partial^2 E}{\partial w^{i2}} + \mu_i I \right]^{-1} \frac{\partial E}{\partial w^i} \quad (6)$$

where w and i stand for the weights and iteration counter, respectively. The term ∂E/∂wⁱ expresses the gradient descent of the performance function (E). Here, μ_i ≥ 0 stands for the learning factor, and the variable I stands for the unity matrix. The value of μ is variable and when it increases, the training method converts to the steepest descent algorithm with a small learning rate. It is useful to increase the precision of training. When the value of μ decreases and is closer to

zero, the LM method converts to a Gauss-Newton method. It is useful to increase the speed of training and save training time.

$$\Delta w_i = -[J^T(w_i)J(w_i) + \mu_i I]^{-1} J^T(w_i)v(w_i) \quad (7)$$

where $J(w_i)$ stands for the Jacobian matrix and the term $v(w_i)$ stands as a residual error of the function. Based on the above formulation, the weights will be modified as:

$$w_{i+1} = w_i + \Delta w_i \quad (8)$$

As the second part, the fuzzy algorithm has two inputs, one from the first part and the second one is the average of the risk scores (R_a) that are weighted based on expert judgement. All the fuzzy system inputs were normalized for simplicity and better performance. In this research, Sugeno inference engine is selected with the weighted average method as defuzzification (Zareh, et al., 2012). Inputs are arranged via triangle membership functions because of the simplicity of calculation, and they are more sensitive to inputs than other types of membership functions. The membership functions of the inputs are shown in Figures 2 and 3.

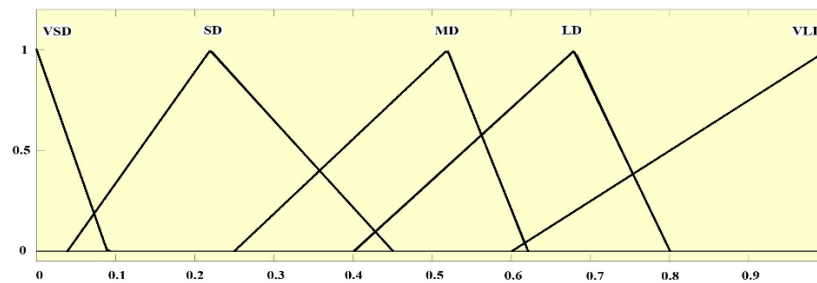


Figure 2. Membership functions for D_a .

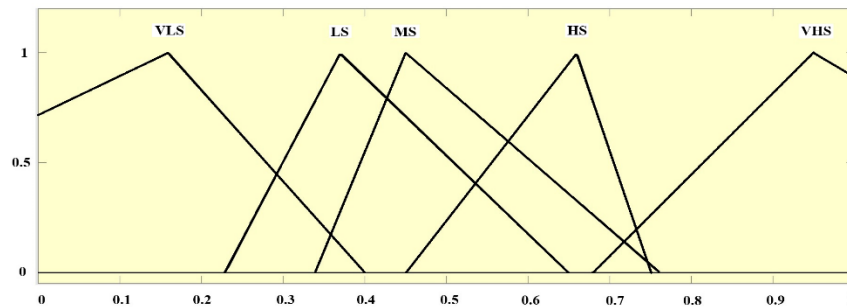


Figure 3. Membership functions for R_a .

The functions were tuned by trial and error to achieve the best responses. For the first input, the abbreviations VSD, SD, MD, LD and VLD stand for very small duration, small duration, medium duration, large duration, and very large duration, respectively. For the second membership functions, abbreviations VLS, LS, MS, HS, and VHS stand for very low score, low score, medium score, high score, and very high score, respectively.

4. Results and Discussion

Different numbers of nodes were used, and the values of MAPE were used to construct the network. The values for MAPE for different numbers of nodes are shown in Table 4.

Table 4. The values of MAPE for the different number of nodes

Number of nodes	Value of MAPE
8	6.18
9	4.07
10	4.43
11	4.47
12	4.13
13	4.43
14	4.83
15	6.84

Therefore, based on the values, 9 nodes had the least value and were selected as the number of nodes for the hidden layer. In this study, to avoid overtraining, training would be stopped if, for 10 iterations the performance error was constant without any changes. The training was done after about 1300 iterations to reach an acceptable performance. For simulation, initial planned durations of 49 activities from the fifth project with their risk scores 49 were sent to the trained network. According to the regression plot (Figure 4), there is a good fit between the outputs of the network and the targets, and the correlation coefficient (R) is 0.993. In other words, a closer value to one for the correlation coefficient shows that predicted outputs are closer to the targets.

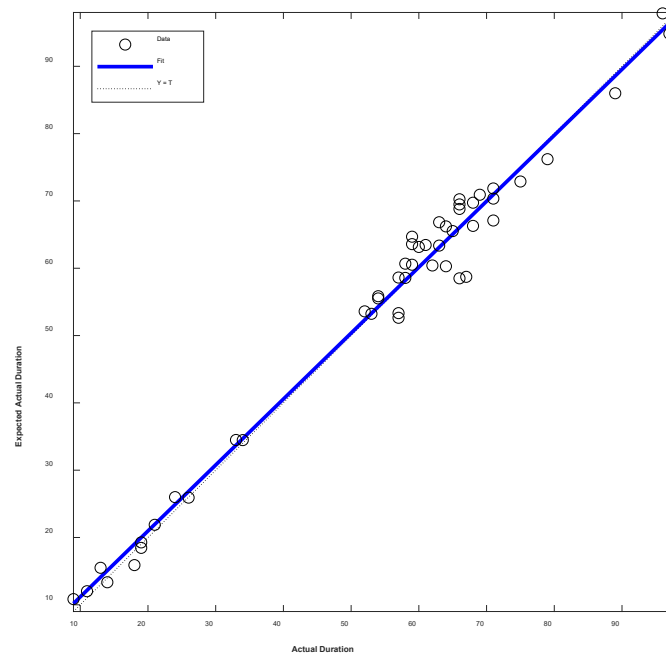


Figure 4. Regression line for the targets (actual durations) versus the outputs (expected actual durations).

The maximum difference between the targets (actual durations) and the outputs (expected actual durations) was 8 days for the installation of the ninth stage for thirteen to fifteen floors of the first zone. The initial planned duration, the actual duration and the expected actual duration for the activity were 44, 65 and 57 days, respectively. This comparison for some other activities is shown in Table 5. About 75% of differences between the outputs and targets were less than 3 days. History of differences between the actual durations and expected actual durations are shown in Figure 5.

Table 5. Comparison between actual and expected durations

Activity Name	(1) Initial Planned Duration (day)	(2) Actual Duration (Day)	(3) Expected Actual Duration (Day)	Difference between (1) and (3)	Difference between (2) and (3)
Installation of the seventh stage- Zone one	75	96	97	22	1
Installation of the eighth stage- Zone one	53	66	64	11	2
Installation of the sixth stage- Zone two	48	63	60	12	3
Installation of the tenth stage- Zone two	50	64	62	12	2

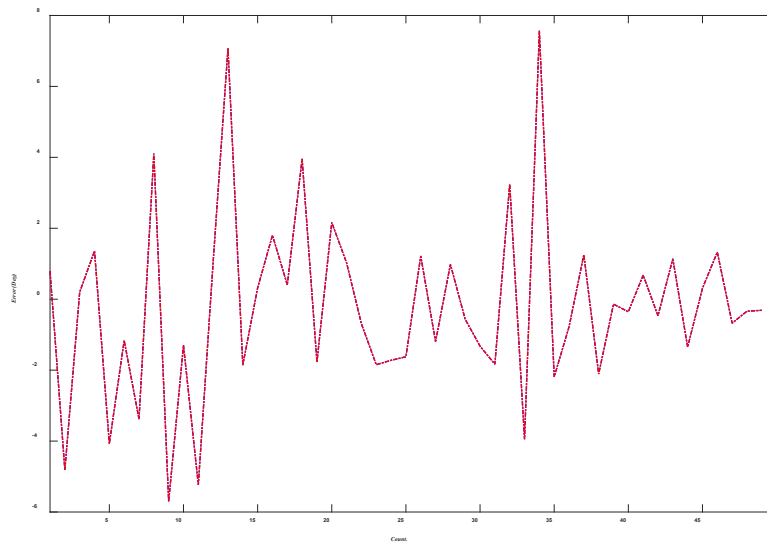


Figure 5. History of differences between the actual durations and expected actual durations.

The negative values are for the cases where the expected actual durations (outputs) are greater than the actual durations (targets). The trends for the initial planned durations, the actual durations and the expected actual durations for the activities are shown in Figure 6.

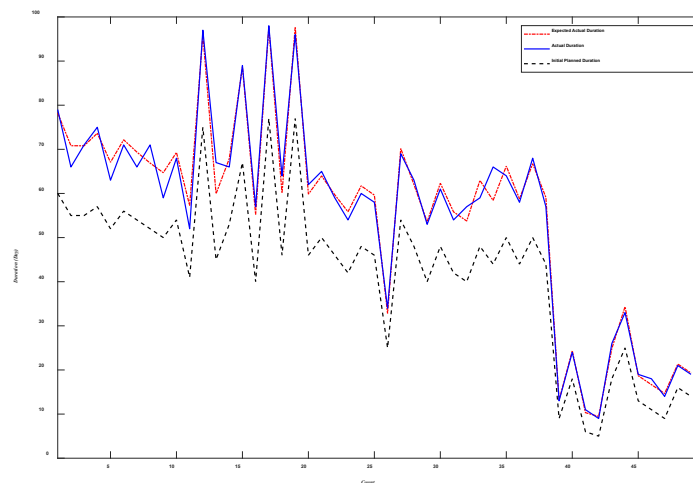


Figure 6. Duration comparisons between the activities of the fifth project.

As can be seen (Figures 5 and 6), the proposed algorithm could predict the expected actual durations based on the risk scores close to the actual durations of the activities.

5. Conclusions

Trying to have accurate durations for activities in construction projects helps contractors avoid many problems like lack of funding or procrastination effect. Although precise estimation of the durations, because of many factors like identified and unidentified negative risks, also are difficult always to try to have better estimation using past data. In this paper, artificial intelligence algorithms using neural networks and fuzzy logic methods have been constructed to estimate steel framing activities. Data from four similar high-rise buildings were used to train the network. Data, as inputs included initial planned durations and actual scores of the five top risks for each project, and the outputs, were the actual durations of those activities. The network had eight input nodes with one output node. MAPE criteria were utilized to identify the number of nodes for the hidden layer. Based on the criteria, 9 nodes had the best performance and were consequently selected. Two membership functions were considered for the fuzzy part to calculate the error and send it to the neural network as feedback for online training and output improvement. For simulation, data from the fifth project was used. The inputs to the trained network were initially planned durations of the fifth project with their scores of five top risks. After that, the outputs (expected actual durations of the activities) were obtained. The obtained results were compared to the actual durations of the activities of the fifth project. According to the comparison, about 75% of 49 steel framing activities of the project had differences with the actual durations of less than 3 days. And the maximum difference was 8 days. The results showed the proposed algorithm can be used for other similar projects with the same characteristics.

References

- Cheung, S. O., and Chow, P. T., Withdrawal in construction project dispute negotiation. *Journal of Construction Engineering and Management*, vol. 137, no 12, pp. 1071-1079, 2011.
- Chong, W. K., Lee, S. H., and O'Connor, J. T., Estimating highway construction production rates during design: Elements of a useful estimation tool. *Leadership and Management in Engineering*, vol. 11, no. 3, pp. 258-266, 2011.
- Elfahham, Y., Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alexandria Engineering Journal*, vol. 58, pp. 499-506, 2019.
- Ezeldin, A. S., and Sharara, L. M., Neural networks for estimating the productivity of concreting activities. *Journal of Construction Engineering and Management*, vol. 132, no. 6, pp. 650-656, 2006.
- Flores, J. A., *Focus on artificial neural networks*, Nova Science Publishers, 2011.
- Golizadeh, H., Sadeghifam, A. N., Aadal, H., and Abd-Majid, M. Z., Automated tool for predicting duration of construction activities in tropical countries. *KSCE Journal of Civil Engineering*, vol. 20, pp. 12-22, 2016.
- Han, C., and Zhang, Q., Optimization of supply chain efficiency management based on machine learning and neural network. *Neural Computing and Applications*, vol. 33, pp. 1419-1433, 2020.
- Hegazy, T., and Ayed, A., Neural network model for parametric cost estimation of highway projects. *ASCE Journal of Construction Engineering and Management*, vol. 124, no. 3, pp. 210-218, 1998.
- Jung, W., Han, S. H., Park, C., Lee, C., and Baek, S., Three-phased risk-management benchmark for internationalization of small and medium-sized construction companies. *KSCE Journal of Civil Engineering*, vol. 26, no. 3, pp. 1024-1039, 2022.
- Kanoglu, A., An integrated system for duration estimation in design/build projects and organizations. *Engineering, Construction and Architectural Management*, vol. 10, no. 4, pp. 272-282, 2003.
- Kim, T. W., Lee, H. S., Park, M., and Yu, J. H., Productivity management methodology using productivity achievement ratio. *KSCE Journal of Civil Engineering*, vol. 15, no. 1, pp. 23-31, 2011.
- Le-Hoai, L., Dai-Lee, Y., and Lee, J. Y., Delay and cost overruns in Vietnam large construction projects: A comparison with other selected countries. *KSCE Journal of Civil Engineering*, vol. 12, no. 6, pp. 367-377, 2008.
- Maravas, A., and Pantouvakis, J., A process for the estimation of the duration of activities in fuzzy project scheduling. *International Conference on Vulnerability and Risk Analysis and Management*, April 11-13, 2011, Hyattsville, Maryland, USA, 2011.
- Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., and Voordijk, H., An artificial neural network approach for cost estimation of engineering services. *International Journal of Construction Management*, vol. 7, pp. 1274-1287, 2022.

- Mir, M., Dipu-Kabir, H. M., Nasirzadeh, F., and Khosravi, A. Neural network-based interval forecasting of construction material prices. *Journal of Building Engineering*, vol. 39, pp. 1022-1035, 2021.
- Isah, M. A., and Kim, B. S., Assessment of risk impact on road project using deep neural network, *KSCE Journal of Civil Engineering*, vol. 26, no. 3, pp. 1014-1023, 2022.
- Ock, J. H., and Han, S. H., Measuring risk-associated activity's duration: A fuzzy set theory application. *KSCE Journal of Civil Engineering*, vol. 14, no. 5, pp. 663-671, 2010.
- Park, M., Kim, W., Yoon, Y., and Nepal, M. P., Scheduling decisions and their dynamic consequences on construction performance. *KSCE Journal of Civil Engineering*, vol. 14, no. 3, pp. 251-259, 2010.
- Petruseva, S., Zujo, V., and Zileska-Pancovska, V., Neural network prediction model for construction project duration. *International Journal of Engineering Research and Technology*, vol. 1, no. 2, pp. 1-13, 2012.
- Rashid, I. A., Aboul-Haggag, S. Y., Mahdi, I., and Elhegazy, H. M., Construction performance control in steel structures projects. *Industrial Engineering and Management*, vol. 5, no. 4, pp. 1-11, 2016.
- Sonmez, R., and Rowings, J. E., Construction labor productivity modeling with neural networks. *Journal of Construction Engineering and Management*, vol. 124, no. 6, pp. 498-504, 1998.
- Wambeke, B. W., Hsiang, S. M., and Liu, M., Causes of variation in construction project task starting times and duration. *Journal of Construction Engineering and Management*, vol. 137, no. 9, pp. 663-677, 2011.
- Waziri, B. S., Jibrin, A. T., and Kadaï, B., Functional duration models for highway construction projects in Nigeria. *Arid Zone Journal of Engineering, Technology and Environment*, vol. 10, pp. 13-23, 2014.
- Zareh, S. H., Sarrafan, A., Khayyat, A. A. A., and Zabihollah, Z., Intelligent semi-active vibration control of eleven degrees of freedom suspension system using magnetorheological dampers. *Journal of Mechanical Science and Technology*, vol. 26, no. 2, pp. 323-334, 2012.
- Zhang, H., Tam, C., and Li, H., Modeling uncertain activity duration by fuzzy number and discrete-event simulation. *European Journal of Operational Research*, vol. 164, pp. 715-729, 2005.

Biographies

Dima Jawad, PEng, PMP, is an Associate Professor and Faculty of Engineering and Applied Science at Ontario Tech University, ON, Canada. She earned her Ph.D. in Civil and Environmental Engineering from Rutgers University, USA, in 2003. Dr. Jawad was the Nakheel Chair for Sustainable Infrastructure, Nakheel, UAE in 2009. Her research interests include quality management, operation management, sustainable development planning, risk management and advanced construction management.

Seiyed Hamid Mousavi, PMP, PMI-RMP, PMI-SP, PSM, earned his Master of Engineering in Engineering Management in 2023 at Ontario Tech University, ON, Canada. Mr. Mousavi has work experience as a scheduler in construction projects.