ARTICLE IN PRESS

Ain Shams Engineering Journal xxx (xxxx) xxx

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Civil Engineering

Identifying the risk factors affecting the overall cost risk in residential projects at the early stage

Mohamed Badawy [a,*], Fahad Alqahtani [b], Hisham Hafez [c]

[a] Structural Engineering Department, Ain Shams University, Cairo 11517, Egypt
[b] Department of Civil Engineering, College of Engineering, King Saud University, P.O.Box 800, Riyadh 11421, Saudi Arabia
[c] Department of Mechanical and Construction Engineering. University of Northumbria, Newcastle upon Tyne, UK

ARTICLE INFO

Article history:
Received 18 January 2021
Revised 21 August 2021
Accepted 13 September 2021
Available online xxxx

Keywords:
Overall risk
Artificial Neural Network (ANN)
Residential projects
Multilayer perceptron
Data mining

ABSTRACT

Many previous studies have developed models for estimating the total cost, whether in the planning stage or the early stage of the project. However, models for estimating the overall risk were proposed in the planning stage only. This paper identifies the factors affecting the overall risk in residential projects at the early stage. The 43 risk factors at the planning stage were identified using a Delphi technique. Experts summarize the 43 risk factors into four factors that can be used to predict the overall risk in the early stage of the project. A multilayer perceptron model with one hidden layer was proposed. The mean absolute error rate for the proposed model was 10%. Risk factors can be used to develop a model to predict the impact of overall risk on project cost at the early stage. The developed model helps stakeholders decide whether the project should continue or be terminated.

1. Introduction

The early stage may be referred to as the feasibility stage, pre-design stage, or preliminary stage. During this stage, the total cost, risks, and duration of the project are estimated in total based on the experience from actual projects similar in scope. At the end of this stage, stakeholders need to decide whether the project is good for investment so that the necessary design can be completed in addition to more detailed drafting or decide to terminate the project. To determine the total contract value, the cost is calculated from previous projects using an analogous method, after which a percentage of this cost is added to cover the overall financial risks in the project and profit.

Risk is classified into two overall types: negative and positive risk. Negative risks may lead to an increase in schedule and cost overruns [1]. The use of professional risk management will mitigate the effects of negative risks or increase the benefit of positive risks. Identification of all risks is time-consuming and in a case may be counterproductive [2]. The project manager should, therefore, focus on the main risks as much as possible [3]. Risk assessment is the most complex process of all risk management processes and is carried out after the risk identification process [4]. In the risk assessment process, potential risks are assessed and arranged, allowing the project manager to include acceptable risk groups in a watch list and hence identify the most significant risks [5]. Poor estimation of the project cost is one of the main reasons for cost overrun [6]. If the contractor takes a high ratio of the risk, the total price will be higher than the independent estimate and the contractor will lose the bid. If the contractor takes a low ratio of the risk, the total price will be lower than the independent estimate and the contractor may not achieve the required profit or even be subject to loss. Therefore, a risk assessment is necessary at an early stage via an accurate method such as an artificial neural network (ANN). Many studies have focused on studying the cost in the early stage of various projects, whether construction projects [7], residential buildings [8], or power plant projects [9]. Unfortunately, there is a gap in the study of risk factors and their effects on the cost of the project at an early stage. Therefore, this research identifies the risk factors affecting the cost of the residential projects at the early stage in Egypt.

* Corresponding author.
E-mail addresses: mohamed_badawy@eng.asu.edu.eg (M. Badawy), bfahad@ksu.edu.sa (F. Alqahtani), hisham.hafez@northumbria.ac.uk (H. Hafez).
Peer review under responsibility of Ain Shams University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.asej.2021.09.013
2090-4479/

Please cite this article as: M. Badawy, F. Alqahtani and H. Hafez, Identifying the risk factors affecting the overall cost risk in residential projects at the early stage, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2021.09.013

Nomenclature

ANN	Artificial Neural Network
RII	relative importance index
ML_H_1	A multilayer perceptron model with one hidden layer and a hyperbolic activation function
ML_H_2	A multilayer perceptron model with two hidden layers and a hyperbolic activation function
ML_S_1	A multilayer perceptron model with one hidden layer and a sigmoid activation function
ML_S_2	A multilayer perceptron model with two hidden layers and a sigmoid activation function.
RBF	A radial function and the SoftMax function as the activation function.
MSE	Mean Square Error
MAPE	Mean Absolute Percentage Error
ER	Estimated Risk
RS	Risk Score
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
TPR	True Positive Rate
TNR	True Negative Rate
ROC	Receiver Operating Characteristic

This research consists of a literature review to determine the most important risk factors that affect the cost of the residential buildings and then analyzing these factors. The second part concern identification of the most important factors that affect the overall risks in residential buildings at an early stage through expert judgments. The third part concern developing a model for predicting the overall risk at the early stage. Five different models have been developed, based on the data mining of real projects. Finally, the results of the study have been compared with previous studies.

2. Literature review

There are many methods for estimating the dependent variables from independent variables. One of these methods is the ANN, which aims to emulate the human brain. In general, its goal is to benefit from the use of artificial intelligence and reduce related global catastrophic risks to a minimum [10]. Artificial neural networks have many applications in different fields, for example, Mehidi et al. (2014) assessed the risk value in cement industries in Bangladesh using ANNs. The model consisted of ten input units, six neurons in the hidden layer, and two units in the output layer with a backpropagation algorithm. The error in the prediction of results was 5% [11]. O'Halloran & Nowaczyk proposed complete financial systems derived from realistic distributions of bank trading data using an artificial intelligence-based approach by incorporating advanced theoretical models of chart and machine learning to estimate credit risk [12].

Other methods exist for predicting dependent variables such as structural equation modeling (SEM), linear regression, system dynamics (SD), discrete event simulation (DES), and analytic hierarchy process (AHP). Ahmadabadi & Heravi (2019) used SEM to construct a framework for risk assessment of large public–private partnership (PPP) projects focusing on the interaction between risk and stakeholder expectations [13]. Lamptey and Emmanuel (2018) used linear regression to estimate the relationship between the risk of cost overrun and change in project cost. The results indicated that changes in the prices of construction materials and the labor cost represent 97% of the cost overrun of construction projects in the south of Nigeria [14]. Eskander (2018) used the analytic hierarchy process to estimate and classify the probability of a risk occurring in construction projects in Egypt and Saudi Arabia during the bidding and construction phases. The results indicated that financial risk is the most important risk followed by design risks, political and construction risks [15]. In other research, a combination of more than one technique was applied. For example, Xu et al. (2018) developed a hybrid dynamic model that combines system dynamics and discrete event simulation to study the impact of risk factors on the performance of infrastructure project schedules [16]. Al-Sobiei et al. (2005) used a genetic algorithm (GA) and ANN to estimate the risk of contractor shortfalls in construction projects being implemented for the Saudi Armed Forces [17].

Elhag and Wang (2007) compared ANN and regression analysis to estimate the risk score and category for bridge maintenance projects. They noticed that the performance of neural network models is better than the performance of multiple regression models [18]. Kim et al. (2013) compared the cost of constructing school buildings using three different methods: regression analysis, neural network, and vector support machine. They argued that the results of the neural network model are more accurate than those of the regression analysis model and those of the support vector machine model. Hence the authors chose to use ANN for analysis in this study.

Risk factors in construction projects have been analyzed in many previous studies. For example, Sameh et al. (2018) assessed the risk of sustainable construction projects in the United Arab Emirates (UAE). Thirty factors were divided into five groups: management, technical, green team, green materials, and organizational/economic. They collected forty-four questionnaires, each one containing the probability and impact of each risk factor. The top risks were lack of owner funding, inadequate sustainable design information, design changes, tight schedule, and weak definition of scope in sustainable construction [19]. Choudhry & Iqbal (2012) revealed that financial factors and quality factors are the most important risks [20]. Chan et al. (2011) identified the key risk factors in the target cost contracts and guaranteed maximum price contracts. Their results showed that the most important risk factors are change in the scope of work, errors in the tender document, exchange rate changes, inadequate design, and unexpected design development [21]. Al-Tabtabai and Alex (2000) developed a model using a neural network model to estimate the increase in project cost due to political risks. Their model depended on 50 cases, six input units, 14 hidden neurons, and one output. The mean absolute error rate was 7% [22]. Chenyun and Zichun (2012) used ANNs to estimate the risk index for an expressway construction stage [23]. Table 1 shows the risk factors and their references.

Many researchers have studied risk factors in the construction industry using traditional techniques that can classify risk factors but not estimate the overall risk, such as [2,15,19–21,31,40]. Other studies developed models to estimate the overall risk, but either low-accuracy methods such as Bayesian network [43] or ANNs requiring many input variables that cannot be determined at the early stage [11,17,18].

Several previous studies have developed models for estimating the total cost, both in the planning stage and in the early stage of the project. While for the project risks, extensive studies have been carried out in estimating the risk factors and developing models to

Table 1
The risk factors compared with the previous studies.

Stakeholder	Risk factor	References
The owner (O)	Delayed owner payments, Owner's finance problem, change orders, the undefined scope of working.	[24–27]
Contractor (C)	Actual quantities differ from the contract quantities, use of defective material, quality control, and quality assurance problems, differing site conditions, damage of material on-site, loss of productivity of equipment, errors on surveying works, lack of workers skills, delayed labor disputes resolutions, changes in management ways, poor resource management, work in more than one shift, shortage of equipment, unavailable materials, poor communications between the home and field offices, poor safety procedures, the security of material and equipment.	[28–34]
Designer (D)	Not coordinated design, delay in design, constructability of design, change in design.	[35–37]
Government (G)	Exchange taxes rate, new governmental legislation, unstable security circumstances, bribery, bureaucracy, difficulty getting permits.	[31,38–40]
External environment (E)	Inflation, exchange rate fluctuation, exchange fuel price, construction material price hike, change of labor cost, change in the price of equipment required, adverse weather conditions, difficulty to access the site, catastrophes.	[31,38,40]
Mutual (M)	Legal disputes during the construction phase among the parties of the contract, high competition in bids, poor communication between parties.	[31,41,42]

estimate the overall risk of the project in the planning stage only. Unfortunately, no model has been proposed to estimate the overall risk in the early stage of the project, especially for residential projects that have a special nature. Due to this gap in risk assessment research in the early stage where there is not enough information about the project in this period. This paper discusses risk factors in the early stage, not the planning stage. To ensure the validity of these factors for estimating the overall risk of the project, a model using neural networks was proposed and the error rate of the proposed overall risk model was compared with the error rate of cost models in the early stage in previous studies. Therefore, this paper assesses the risk factors in residential projects in Egypt and develops a model to predict the impact of the overall project cost risks depending on some variables that can be easily identified in the early stage.

3. Methodology

A preliminary list of risk factors was proposed based on the literature review. The Delphi technique was used in this research to obtain a consensus among experts on the final questionnaire. This technique is based on a primary questionnaire answered by experts. After receiving the experts' answers, the researcher reviews answers and removes repeated ones before modifying the questionnaire, which is sent back to all experts for further comments. The review and resending processes are repeated until consensus is reached among experts on the final questionnaire.

In this research, the initial questionnaire was sent anonymously to experts for comments. Expert comments were reviewed and forwarded to the same experts for further comments. The authors collected 200 questionnaires. Data were analyzed and validated for reliability using Alpha Cronbach. Experts identified the most important risk factors that may affect overall risk in residential buildings at the early stage. Real data from housing projects were collected with their corresponding risk factors. Due to time considerations, only 38 projects were collected, thus the data was repeated 5 times to construct models. Five different models were created to forecast the overall risk rating for residential projects. Multilayer perceptron (ML) with one hidden layer was suggested using the hyperbolic activation function in the first model. In the second model, multilayer perceptron was also applied, but with two hidden layers using the hyperbolic activation function. The third model was like the first model but used the sigmoid function as the activation function. Multilayer perceptron with two hidden layers and the Sigmoid activation function was proposed in the fourth model. In the last model, the radial function and the Soft-Max activation function were applied. In all models, 119 cases were identified for training, 34 cases for testing, and 37 cases for the holdout. The final proposed model was the one that contained the minimum mean square error. Ten folds were used for cross-validation. Sensitivity analysis was performed to determine the importance of each factor. The research methodology is illustrated in Fig. 1.

4. Ranking of risk factors at the planning stage

Five experts with at least 15 years of experience in project management were selected. A preliminary list of 51 risk factors that were identified through a literature review was presented to experts and they were asked to add any missing factor, in the first round. The answer was that no significant risk factor was missing. In the second round, experts were asked to rate the importance of each factor on a five-point Likert scale. The median rather than the mean was relied upon to reduce bias. Since the minimum importance value was (1) and the maximum value was (5), so the range was (4). This range was divided into three regions. Hence, if the median of any factor was less than 2.3, which means that this factor has a low impact and is therefore removed from the list. According to Table 2, eight risk factors were omitted from the list. Factors were sent to the experts with the median values. Each expert was asked to indicate whether he agreed with the final list or not. Experts unanimously agreed on the final list of factors in the third round.

The risk factors were taken from the cited papers and categorization was done by the authors. Risk factors were divided into six main groups: owner (O), the contractor (C), designer (D), government (G), the external environment (E), and mutual (M). The identified risks in the final questionnaire were evaluated by the experts of residential projects in Egypt. The questionnaire consisted of two parts. The first part of the questionnaire concerned general information for the respondents. The second part targeted the probability and impact of the risk factors. Probability and impact scales are divided into five categories: very low, low, medium, high, and very high. The questionnaire was distributed to 230 experts through interviews with experts. Only 200 experts responded for an average response of 87%. The relative importance index (RII) was calculated for each risk using Eq. (1).

$$RII = \frac{\sum_{i=1}^{N} P * I}{N} \tag{1}$$

where (RII) represents the relative importance index, (P) represents the probability of the risk occurrence, (I) represents the impact of the risk factor, and (N) represents the number of respondents to the questionnaire which was 200 in this study.

The maximum probability or impact was (5) which indicated a very high probability, and the minimum was (1), as shown in

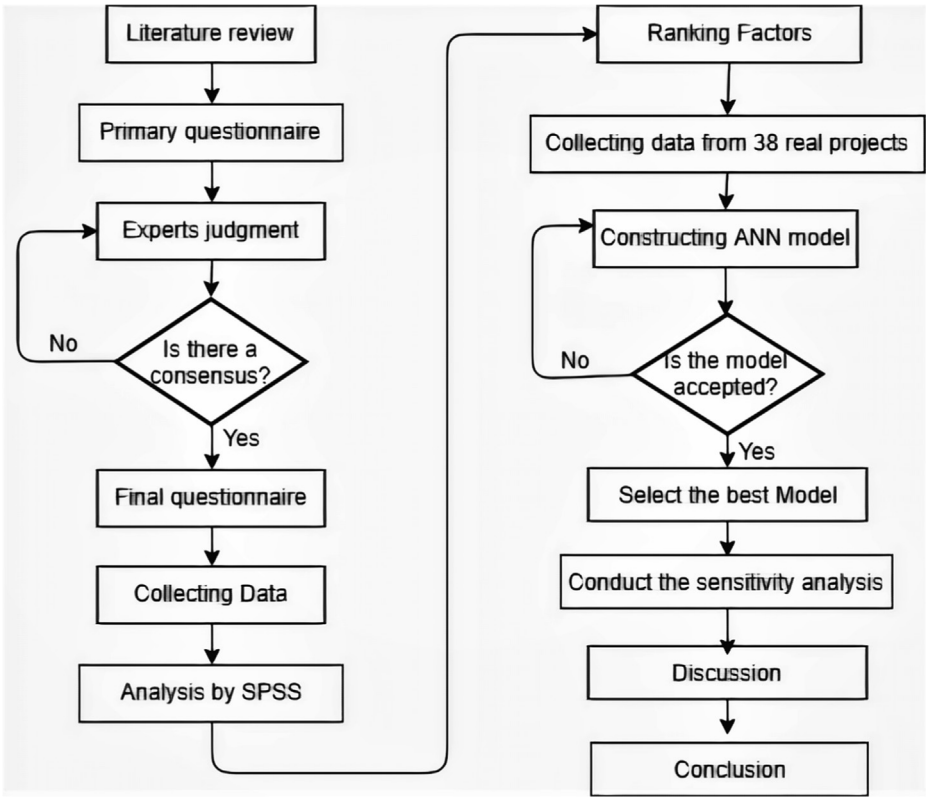

Fig. 1. The methodology charts.

Table 2
The omitted risk factors from the primary list.

	Risk factor	Ex1	Ex2	Ex3	Ex4	Ex5	Mode	Avg
O5	Late handing over of site	2	1	1	2	1	1	1.4
O6	Slow decision-making process	2	1	2	3	2	2	2
C18	Poor site management	1	2	1	1	1	1	1.2
C19	Damage to the structure	1	1	1	1	1	1	1
C20	Poor performance of sub-contractors	1	2	2	2	1	2	1.6
D5	Errors & omission in design drawing	1	2	1	1	1	1	1.2
E10	Culture differences	2	1	1	1	1	1	1.2
M4	Lack of transparency	2	2	2	2	1	2	1.8

Table3. Hence, the maximum relative importance index (RII) was (25) and the minimum RII was (1). The authors classified the risk factors into five categories: very low, low, medium, high, and very high. The range of each category can be calculated using Eq. (2).

$$R = \frac{(A - B)}{m} \tag{2}$$

where R is the range of each category, A is the maximum RII which equals 25, B is the minimum RII which is equal to 1, and m is the number of categories which is equal to 5. The range of each risk category according to the relative importance index is shown in Table 3.

The reliability of the values in the questionnaire was calculated by SPSS software. Cronbach's Alpha for the results was 0.914, which reflects the high reliability of the questionnaire. Furthermore, it indicates a high level of internal consistency concerning the data collected. The ANOVA test was applied to determine whether the results would change due to the difference in experience or not. The results showed that significant values were greater than (0.05), which proved that there was no difference in mean values regarding the years of experience of the respondents. The characteristics of the risk factors are shown in Table 4. The first column contains the symbol of the risk, which indicates the category to which this factor belongs, the second column shows the name of

Table 3
Risk category.

Risk category	Very low	Low	Medium	High	Very high
Probability	1	2	3	4	5
Impact	1	2	3	4	5
RII	1–5.8	5.8–10.6	10.6–15.4	15.5–20.2	20.2–25

each risk factor. The third and fourth columns show the probability and impact of the risk, respectively. The fifth column indicates the relative importance index, the sixth column represents the qualitative risk analysis, and the last column shows the ranking of the risk factor.

Of all 43 risk factors, there was no factor placed in a very high or very low-risk category as shown in Table 4. It is also clear that there were only two factors in the high-risk category. The most important factor was the "exchange rate fluctuation", with a relative importance index of 16.6. The "construction material prices change" factor was rated as the second important factor, with a relative importance index of 15.7. The results showed that there were 17 factors in the medium-risk category. Twenty-seven factors were classified as low-risk factors. The average risk factor for all risks was 10.4, indicating that the risk in the construction of the residential projects in Egypt can be classified as low risk according to Table 3.

5. Risk factors at the early stage

Not all risk factors can be identified and analyzed at the early stage due to limited information about the project at this stage. Five experts reviewed the relative importance index of each risk factor to identify the risk factors that should be taken into consideration at the early stage. Experts unanimously summarized the 43 risk factors into four risk factors that should be considered when predicting the overall risk at the early stage of a project as shown in Table 5. These factors are the contract type, implementation of risk management processes, the contract cost, and total project duration. The factor "contract type" was classified into a fixed-price or cost-plus contract. The factor "implementation of risk management processes" was classified into an "application of risk management processes to the project" or "no of risk management processes applied to the project". Due to time considerations, the authors monitored only 38 residential projects. In each project, the percentage of the real risk and the corresponding four risk factors were recorded. As the population was large and unknown, the population was considered unlimited, so that the sample size can be determined using Eq. (3).

$$SS = \frac{Z2 * p * (1 - p)}{C2} \tag{3}$$

where SS represents the sample size, which was 38, Z was equal to 1.96 according to 95% confidence level, p represents the probability of the choice expressed as a decimal and chosen to be 0.5, and C represents the confidence interval expressed as a decimal. In this study, from Eq. (3) the confidence interval was 0.16 which is less than 0.2, hence the results can be accepted at the early stage [8].

Table 4
The risk assessment at the planning stage.

I.D.	Risk factors	Prob.	Impact	RII	Category	Rank
O1	Delayed owner payments	3.3	3.5	11.7	Medium	9
O2	Owner's finance problem	3.1	3.7	11.7	Medium	10
O3	Undocumented change orders	2.9	3.4	10.3	Low	21
O4	Undefined scope of working	2.4	3.3	8.6	Low	39
C1	Differing between the actual quantities and the contract quantities	3.2	3.6	11.9	Medium	8
C2	Use of defective material	2.7	3.1	8.9	Low	35
C3	Quality control and quality assurance problems	3.0	3.3	10.1	Low	22
C4	Differing site conditions	2.5	3.5	9.2	Low	31
C5	Damage of material on site	2.6	3.4	9.3	Low	30
C6	Loss of productivity of equipment	2.9	3.3	9.9	Low	24
C7	Errors in surveying works	2.5	3.4	8.5	Low	40
C8	Lack of workers skills	3.1	3.3	10.7	Medium	19
C9	Delayed in labor disputes resolutions	2.4	2.6	6.9	Low	42
C10	Changes in management methods	2.9	3.1	9.1	Low	34
C11	Poor resource management	3.0	3.2	10.0	Low	23
C12	Working for more than one shift	3.0	3.4	10.8	Medium	18
C13	Shortage of equipment	2.8	3.3	9.5	Low	27
C14	Unavailable material	3.1	3.5	11.3	Medium	12
C15	Poor communications between the overhead and field offices	2.9	3.1	9.5	Low	28
C16	Poor safety procedures	3.0	3.3	10.4	Low	20
C17	Poor Security of material and equipment	3.0	3.1	9.8	Low	25
D1	Not coordinated design	3.2	3.4	11.0	Medium	15
D2	Delay in design	3.1	3.4	11.1	Medium	14
D3	Non-constructability of design	3.2	3.3	10.9	Medium	16
D4	Change in design	3.3	3.7	12.5	Medium	6
G1	Changing tax rate	3.3	3.6	12.7	Medium	5
G2	New governmental legislations	2.6	3.2	9.1	Low	32
G3	Unstable security circumstances	2.6	3.4	9.4	Low	29
G4	Bribery/corruption	2.6	3.6	9.5	Low	26
G5	Bureaucracy	2.6	3.4	9.1	Low	33
G6	Difficulty to get permits	3.0	3.4	10.8	Medium	17
E1	Inflation	3.5	3.9	14.0	Medium	4
E2	Exchange rate fluctuation	3.9	4.2	16.6	High	1
E3	Changing fuel price	3.7	4.0	15.3	Medium	3
E4	Construction material price hike	3.7	4.1	15.7	High	2
E5	Change of labor cost	3.2	3.4	11.5	Medium	11
E6	A change in equipment price	3.2	3.5	12.0	Medium	7
E7	Adverse weather conditions	2.5	2.6	7.2	Low	41
E8	Difficulty to access the site	2.7	2.9	8.8	Low	37
E9	Catastrophes (floods, earthquakes, fire)	1.9	3.3	6.7	Low	43
M1	Legal disputes during the construction Phase among the parties of the contract	2.6	3.0	8.9	Low	36
M2	High competition in bids	3.3	3.3	11.1	Medium	13
M3	Poor communication between Parties	2.6	3.1	8.6	Low	38

Table 5
Risk factors at the early stage.

I.D.	The contract type	Implementation of risk management processes	The contract cost	Total project duration	Not applicable
O1				√	
O2			√		
O3		√			
O4	√				
C1	√				
C2		√			
C3		√			
C4					√
C5		√			
C6		√			
C7					√
C8		√			
C9					√
C10					√
C11		√			
C12		√			
C13		√			
C14		√			
C15		√			
C16		√			
C17		√			
D1					√
D2					√
D3					√
D4					√
G1				√	
G2					√
G3					√
G4					√
G5					√
G6					√
E1				√	
E2					
E3		√			
E4		√			
E5		√			
E6		√			
E7					√
E8					√
E9					√
M1				√	
M2	√				
M3		√			

6. A model for estimating the overall risk

To ensure that the above-mentioned four risk factors can be used to determine the overall risk at the early stage with an accepted error, an ANN model was developed, and the mean error was calculated. The IBM SPSS software was chosen because of its effectiveness as well as its ease of use and ability to illustrate the results through diagrams. The "Implementation of risk management processes" factor has two options, "yes" if risk management processes are to be implemented in the project, or "no" if risk management processes do not apply to the project. The factor of "the contract type" has two options, fixed-price contract, or cost-plus contract. Hence, these two factors were selected as factors in the input layer. On the other hand, the factors of "the contract value" and "the total duration of the project" are numerical values. Hence, these two factors were selected as covariates in the input layer. The rescaling method of covariates was standardized. The output of the model was "the classification of overall risk" which was divided into three categories, low, medium, and high, according to the risk score shown in Table 6.

Due to time considerations, data from only 38 projects were gathered. The cases were duplicated five times to develop artificial neural network models. Five different ANN models were proposed to forecast the overall risk classification of residential projects, with a different number of hidden layers and activation functions. In the first model (ML_H_1), a multilayer perceptron with one hidden layer and a hyperbolic activation function was used, while the second model (ML_H_2) used a multilayer perceptron with two hidden layers and a hyperbolic function. In the third model (ML_S_1), a multilayer perceptron with one hidden layer and a sigmoid activation function was used, while the fourth model (ML_S_2) used a multilayer perceptron with two hidden layers and a sigmoid function. The fifth model (RBF) used a radial function and the SoftMax function as the activation function. 119 cases were used to train the network, 34 to test it while 37 cases were used for the holdout. For each model, the predicted classifications of the overall risk were estimated. To evaluate the performance of the ANN models, many statistical measures are available to calculate the difference between estimated and actual values. In this study, the performance of each model was measured by calculating the mean square error (MSE) which can be calculated using Eq. (4), and the mean absolute percentage error (MAPE) which can be calculated using Eq. (5) [44].

Table 6
Classification of risk score.

Risk category	Low	Medium	High
Risk score	Less than 0.05	0.05–0.15	More than 0.15

$$MSE = \frac{\left(\sum_{i=1}^{N}((ER - RS)^2\right)}{N} \quad (4)$$

$$MAPE = \frac{\left(\sum_{i=1}^{N}\left|\frac{(ER-RS}{RS} * 100\right|\right)}{N} \quad (5)$$

where ER represents the estimated risk, RS represents the risk score, and N is the number of cases. The estimated and actual risk scores for each model were collected in a Microsoft Excel sheet to calculate the MSE and the MAPE in each model.

The model with the minimum error was selected as the proposed model, which was the first model (ML_H_1) with an overall MSE of 15.79% and MAPE of 10.09%. The proposed model consisted of one input layer with six units plus bias, one hidden layer with three neurons in addition to bias, and an output layer with three units. The input, hidden, and output layers of the network diagram are illustrated in Fig. 2. The MSEs and the MAPE of the five developed models were shown in Table 7.

The collected data were repeated 5 times, then divided using 10-fold cross-validation to minimize the risk. Each fold has the same distribution of the risk score; 5 cases with low-risk, 11 cases with medium, and 3 cases with the high-risk score. Ten trials were proposed. In each trial, one-fold was selected to test the network and the remaining folds were used for training. The number of neurons and the errors in training and testing in each trial is shown in Table 8. The maximum error in testing was found in fold-6 with 31.6%, while the average of the 10-fold model in the training stage was 16.2%, which is close to the estimated error in the training stage of the proposed model which was 16.8%. The average of the 10-fold model in the testing stage was 12.6%, which is close to the estimated error in the training stage of the proposed model which was 11.8%.

The classifications of the observed and predicted overall risks were shown in Table 9. The proposed model classified 190 cases, with 99 out of 119 cases in the training stage predicted accurately and 20 predicted incorrectly. In the testing stage, 30 out of 34 cases were perfectly estimated while only 4 cases were incorrect, while in the holdout 31 cases were correct and six cases were incorrect.

7. Discussion

According to the data analysis, the most important risk factors in residential projects at the planning stage were the exchange rate fluctuation, construction material price change, exchange fuel prices, inflation, and exchange rate tax. These factors belong to the external environment and government groups. The factor of exchange rate fluctuation is an important factor and ranks first in this study. This follows in the view of recent economic events in Egypt. This factor was ranked 16th in Chan et al. (2010) [45], 32nd in El-Sayegh (2008) [2], and 6th in Bing et al. (1999) [46]. The risk factor may affect at least one of the objectives of the project, either cost, time, scope, or quality. In this paper, the effects of risk factors were estimated based on their effect on cost only. The ranking of risk factors may change if the impact of risk factors on any other project objective, such as time, is taken into consideration.

Bid price is one of the criteria for the bidding process. Scope of work and technical resources are the main criteria for the early-stage bidding evaluation process [47]. The cost factor was considered in the model proposed in this study. By the scope of work and technical resources, the duration of the activities can be determined, and then the total duration of the project can be estimated. In this research, the scope of work and the technical resources were expressed by the total project duration. The proposed model for predicting the overall risk in construction projects was a multilayer perceptron with one hidden layer and a hyperbolic activation function. The proposed model consisted of one input layer with six units plus bias, one hidden layer with three neurons in addition to

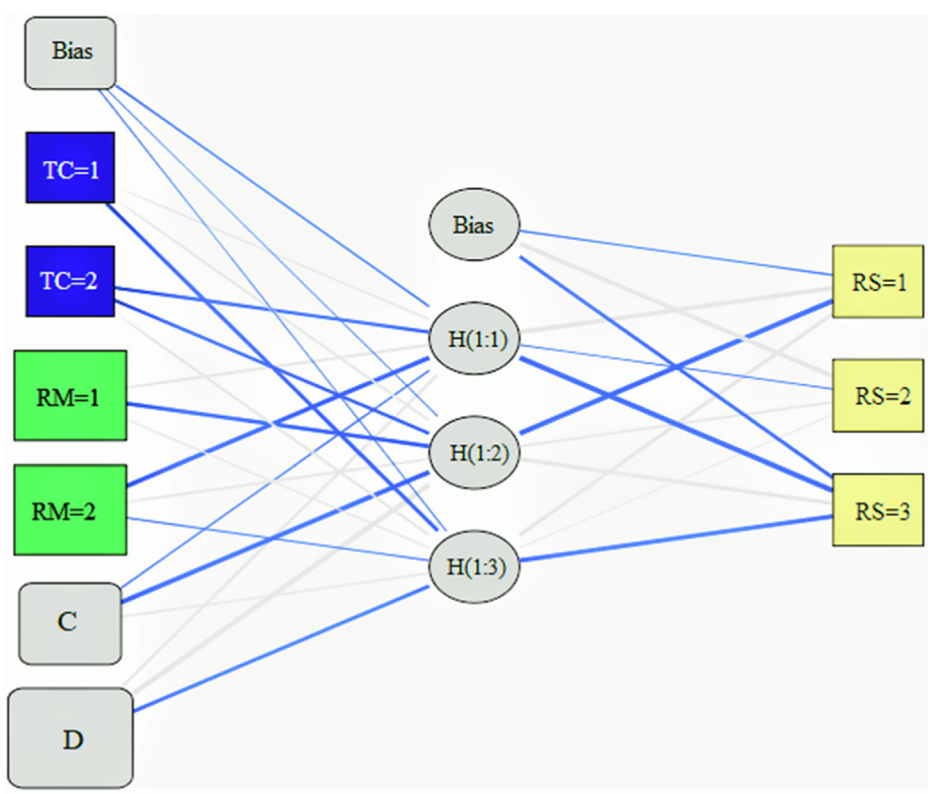

Fig. 2. The architecture of the proposed model.

Table 7
The errors in the five developed models.

	ML_H_1	ML_H_2	ML_S_1	ML_S_2	RBF
Training	16.80%	16%	24.40%	25.20%	19.30%
Testing	11.80%	20.60%	35.30%	26.50%	29.40%
Holdout	16.20%	24.30%	24.30%	29.70%	32.40%
Overall	15.79%	18.42%	26.32%	26.32%	23.68%
MAPE	10.09%	14.47%	17.11%	17.98%	17.11%

Table 8
The errors in K-fold.

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6	Fold-7	Fold-8	Fold-9	Fold-10
No. of neurons	5	4	4	4	4	3	6	8	4	5
training	18.7	8.2	13.5	18.7	22.8	31.6	14.6	11.7	16.4	5.3
testing	15.8	5.3	10.5	15.8	5.3	31.6	26.3	0	10.5	5.3

Table 9
actual and predicted Risk classification.

			Predicted			
	Sample		Low	Medium	High	Correct %
Actual	Training	Low	22	7	0	75.9%
		Medium	2	61	6	88.4%
		High	0	5	16	76.2%
	Testing	Low	6	2	0	75.0%
		Medium	1	18	1	90.0%
		High	0	0	6	100.0%
	Holdout	Low	12	1	0	92.3%
		Medium	2	16	3	76.2%
		High	0	0	3	100.0%

bias, and an output layer with three units. The inputs of the proposed model were "the implementation of risk management processes is yes", "the implementation of risk management processes is no", "a contract type is a fixed-price contract", "a contract type is a cost-plus contract", "the contract value" and "the total duration of the project". The outputs of the proposed model were the classifications of the overall risk which were low, medium, and high. The overall mean square error in the proposed model was 15.79% and the mean absolute percentage error was 10.09%.

Hyari et al. accept an average accuracy rate of 26% and an absolute error rate of 28% in their model [48]. The error can be accepted in the early stage if it is less than 20% [8]. The differences in cost between the estimated cost of the projects at the feasibility studies and the contractual cost of public projects in the UAE range between −28.5% and + 36% [49]. The error in the proposed model for estimating the overall risk in this research was 16%, which was less than the accepted error for the above-mentioned studies. Hence, the ANN proposed model can be considered acceptable and the four input factors are enough to estimate the overall risk in the early stage.

The importance of the four input variables in the prediction of the overall risk in the early stage of the project is illustrated in Fig. 3, which indicates that the contract duration is the most

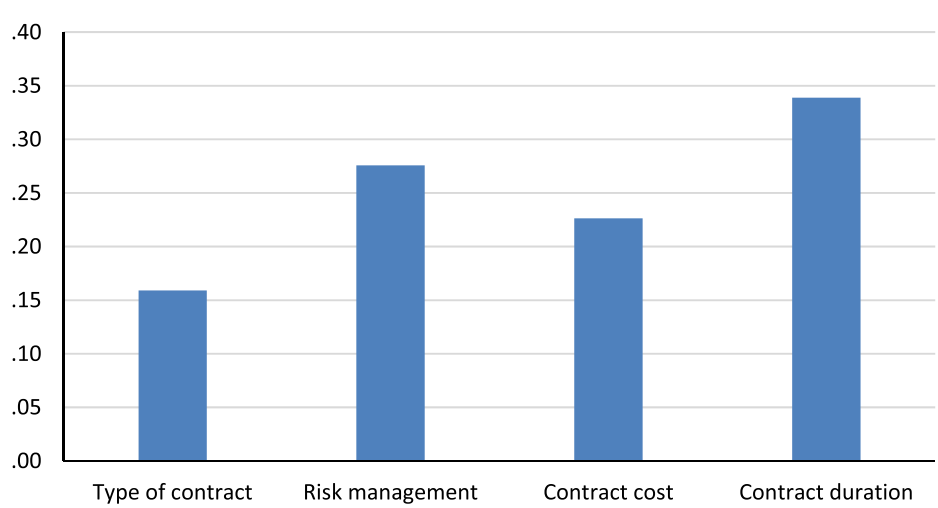

Fig. 3. The importance of independent variables.

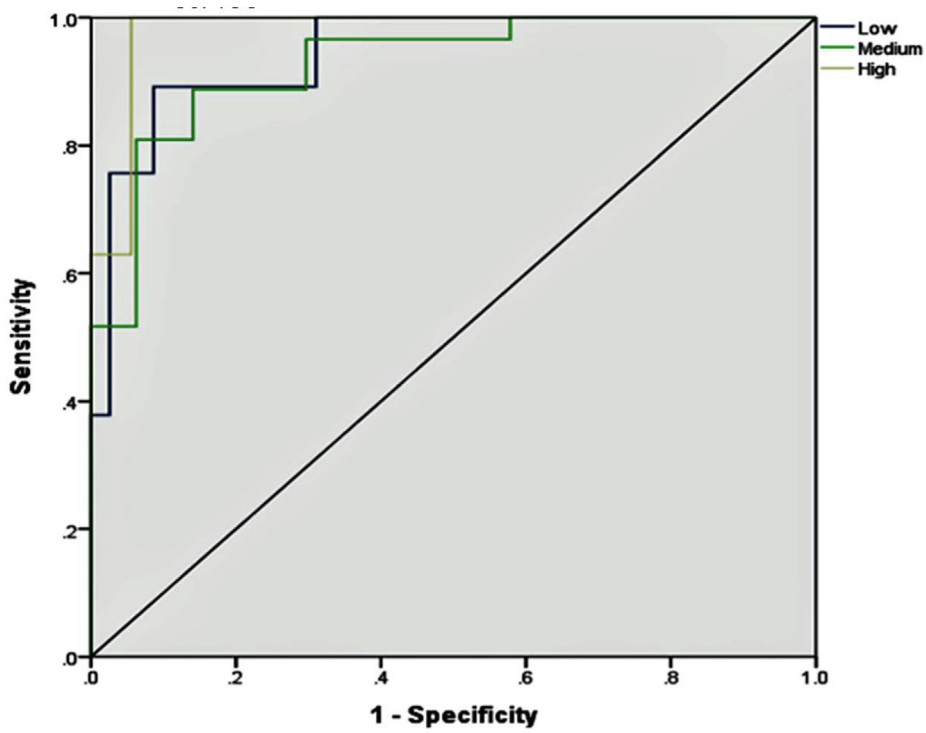

Fig. 4. The receiver operating characteristic.

important factor in affecting the overall risk in residential buildings with the importance of 0.34. The next important factor is the implementation of risk management processes in the project, which have the importance of 0.28, followed by the contract cost of the importance of 0.23. The final factor was the contract type with the importance of 0.16. The sensitivity representing the true positive rate can be estimated using Eq. (6). The specificity representing the true negative rate can be estimated from Eq. (7). The receiver operating characteristic (ROC) is illustrated in Fig. 4 which illustrates the relationship between sensitivity versus specificity. The accuracy in predicting the low-risk or high-risk categories was higher than the accuracy of estimating the overall risk in the medium-risk category.

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{7}$$

8. Conclusion

The main risk factors at the planning stage were determined, via a questionnaire designed using the Delphi technique. 200 responses to questionnaires were received. The results indicated that the most important risk factors in residential projects at the planning stage were the exchange rate fluctuation, construction material price change, exchange fuel price, and inflation. The risk factors at the planning stage were identified and analyzed which were 43 factors to help to determine the risk factors at the early stage which were four factors only. This research aimed to identify the factors that can be used to predict the overall risk of residential projects in Egypt at an early stage. Unfortunately, not all the risk factors identified at the planning stage can be analyzed at the early stage due to the lack of data. Four risk factors affecting the overall risk were identified at an early stage. These factors were the implementation of risk management processes, the contract cost, contract type, and the project duration. Data mining for 38 real residential projects in Egypt was performed. The cases were duplicated five times to develop the ANN models. Five different ANN models were proposed to forecast the overall risk of the project, using different numbers of hidden layers and activation functions. The proposed model contained one hidden layer with three neurons. The overall MSE of the proposed ANN model was 15.79% and the value of MAER was 10.09%. The contract duration and the implementation of risk management were the most important factors that affect the overall risk of the residential buildings in Egypt at the early stage. The main challenge facing this study is performing an accurate risk analysis with limited information available. The results of this study indicated that the error rate was less than the required threshold, hence the model can be accepted. Hence, the four input factors can be used to identify the overall risk at the early stage.

9. Limitations of research

The impact of the risk factors and the overall risk were estimated only based on the impact on the cost of the residential buildings in this study. The four input variables used in this research can be used to estimate the overall risk of construction projects only in Egypt and should be reviewed before using in any other country. However, because the data used to develop the model was obtained from residential buildings in Egypt constructed between 2018 and 2019, the user must adjust the weights of variables to apply it to later times.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors extend their appreciation to the Research Supporting Project number (RSP-2021/264), King Saud University, Riyadh, Saudi Arabia for funding this work.

References

[1] Smith NJ, Merna T, Jobling P. Managing risk in construction projects. Wiley; 2014.
[2] El-Sayegh S. Risk assessment and allocation in the UAE construction industry. Int J Proj Manag 2008;26:431–8. doi: https://doi.org/10.1016/j.ijproman.2007.07.004.
[3] Barber RB. Understanding internally generated risks in projects. Int J Proj Manag 2005;23:584–90. doi: https://doi.org/10.1016/j.ijproman.2005.05.006.
[4] Thomas AV, Kalidindi SN, Ganesh LS. Modelling and assessment of critical risks in BOT road projects. Constr Manag Econ 2006;24:407–24. doi: https://doi.org/10.1080/01446190500435275.
[5] Hashemi H, Mousavi SM, Tavakkoli-Moghaddam R, Gholipour Y. Compromise ranking approach with bootstrap confidence intervals for risk assessment in port management projects. J Manag Eng 2013;29:334–44. doi: https://doi.org/10.1061/(asce)me.1943-5479.0000167.
[6] Ali Shaikh F, Odhano N, Kaliannan S. Performance and management of cost in the construction industry. Civ Eng J 2020;6(1368–74). doi: https://doi.org/10.28991/cej-2020-03091554.
[7] Magdum SK, Adamuthe AC. Neural network approach for early cost estimation in construction projects. 3rd natl conf eng post grad students RIT NConPG-17, 2017.
[8] Badawy M. A hybrid approach for a cost estimate of residential buildings in Egypt at the early stage. Asian J Civ Eng 2020;21:763–74. doi: https://doi.org/10.1007/s42107-020-00237-z.
[9] Gunduz M, Sahin HB. An early cost estimation model for hydroelectric power plant projects using neural networks and multiple regression analysis. J Civ Eng Manag 2015;21:470–7. doi: https://doi.org/10.3846/13923730.2014.890657.
[10] Boyd M, Wilson N. Catastrophic risk from rapid developments in artificial intelligence. Policy Q 2020;16. doi: https://doi.org/10.26686/pq.v16i1.6355.
[11] Mehidi SH, Chakrabarty N, Mohiuddin HM. An application of artificial neural network (ANN) process to assess risk in cement industries in bangladesh. Ind Eng Manag 2014;3:1–6.
[12] O'Halloran S, Nowaczyk N. An artificial intelligence approach to regulating systemic risk; 2020.
[13] Akbari Ahmadabadi A, Heravi G. Risk assessment framework of PPP-megaprojects focusing on risk interaction and project success. Transp Res Part A Policy Pract 2019;124:169–88. doi: https://doi.org/10.1016/j.tra.2019.03.011.
[14] Lamptey-Puddicombe AD, Emmanuel AT. Evaluation of risk of fluctuation claim on cost of construction projects in the south-south zone of Nigeria. Civ Eng Archit 2018;6:252–6.
[15] Eskander RFA. Risk assessment influencing factors for Arabian construction projects using analytic hierarchy process. Alexandria Eng J 2018;57:4207–18. doi: https://doi.org/10.1016/j.aej.2018.10.018.
[16] Xu X, Wang J, Li CZ, Huang W, Xia N. Schedule risk analysis of infrastructure projects: a hybrid dynamic approach. Autom Constr 2018;95:20–34. doi: https://doi.org/10.1016/j.autcon.2018.07.026.
[17] Al-Sobiei OS, Arditi D, Polat G. Predicting the risk of contractor default in Saudi Arabia utilizing artificial neural network (ANN) and genetic algorithm (GA) techniques. Constr Manag Econ 2005;23:423–30. doi: https://doi.org/10.1080/01446190500041578.
[18] Elhag TMS, Wang Y-M. Risk assessment for bridge maintenance projects: neural networks versus regression techniques. J Comput Civ Eng 2007;21:402–9. doi: https://doi.org/10.1061/(ASCE)0887-3801(2007)21:6(402).
[19] El-Sayegh SM, Manjikian S, Ibrahim A, Abouelyousr A, Jabbour R. Risk identification and assessment in sustainable construction projects in the UAE. Int J Constr Manag 2018. doi: https://doi.org/10.1080/15623599.2018.1536963.
[20] Choudhry RM, Iqbal K. Identification of risk management system in construction industry in Pakistan. J Manag Eng 2013;29:42–9. doi: https://doi.org/10.1061/(asce)me.1943-5479.0000122.
[21] Chan DWM, Chan APC, Lam PTI, Yeung JFY, Chan JHL. Risk ranking and analysis in target cost contracts: empirical evidence from the construction industry. Int J Proj Manag 2011;29:751–63. doi: https://doi.org/10.1016/j.ijproman.2010.08.003.
[22] Al-Tabtabai H, Alex AP. Modeling the cost of political risk in international construction projects. Proj Manag J 2000;31:4–13. doi: https://doi.org/10.1177/875697280003100302.
[23] Chenyun Zichun Y. The BP artificial neural network model on expressway construction phase risk. Syst Eng Procedia 2012;4:409–15. doi: https://doi.org/10.1016/j.sepro.2012.01.004.
[24] Assaf SA, Al-Hejji S. Causes of delay in large construction projects. Int J Proj Manag 2006;24:349–57. doi: https://doi.org/10.1016/j.ijproman.2005.11.010.
[25] Sun Y, Fang D, Wang S, Dai M, Lv X. Safety risk identification and assessment for Beijing Olympic venues construction. J Manag Eng 2008;24:40–7. doi: https://doi.org/10.1061/(ASCE)0742-597X(2008)24:1(40).
[26] El-Sayegh S, Mansour M. Risk assessment and allocation in highway construction projects in the UAE. J Manag Eng 2015;31:04015004. doi: https://doi.org/10.1061/(asce)me.1943-5479.0000365.
[27] Badawy M. Second-order confirmatory factor analysis model for estimating the overall risk of change orders in road projects ahead-of-p. J Eng Des Technol 2021. doi: https://doi.org/10.1108/JEDT-09-2020-0383.
[28] Aibinu AA, Odeyinka HA. Construction delays and their causative factors in nigeria. J Constr Eng Manag 2006;132:667–77. doi: https://doi.org/10.1061/(ASCE)0733-9364(2006)132:7(667).
[29] Choudhry RM, Aslam MA, Hinze JW, Arain FM. Cost and schedule risk analysis of bridge construction in pakistan: establishing risk guidelines. J Constr Eng Manag 2014;140:04014020. doi: https://doi.org/10.1061/(asce)co.1943-7862.0000857.
[30] Ke Y, Wang S, Chan APC. Risk allocation in public-private partnership infrastructure projects: comparative study. J Infrastruct Syst 2010;16:343–51. doi: https://doi.org/10.1061/(asce)is.1943-555x.0000030.
[31] Kishan P, Bhavsar JJ, Bhatt R. A study of risk factors affecting building construction projects. Int J Eng Res Technol 2014;3:831–5.
[32] Khodeir LM, Nabawy M. Identifying key risks in infrastructure projects – case study of Cairo Festival City project in Egypt. Ain Shams Eng J 2019;10:613–21. doi: https://doi.org/10.1016/j.asej.2018.11.003.
[33] Abuaddous M, Al-Btoosh JAA, Al-Btoush MAKA, Alkherret AJ. Building information modeling strategy in mitigating variation orders in roads projects. Civ Eng J 2020;6(1974–82). doi: https://doi.org/10.28991/cej-2020-03091596.
[34] Nabawy M, Ofori G, Morcos M, Egbu C. Risk identification framework in construction of Egyptian mega housing projects. Ain Shams Eng J 2021;12:2047–56. doi: https://doi.org/10.1016/J.ASEJ.2020.09.016.
[35] Shen LY. Project risk management in Hong Kong. Int J Proj Manag 1997;15:101–5. doi: https://doi.org/10.1016/S0263-7863(96)00045-2.
[36] Yuan JF, Deng XP, Li QM. Critical risks identification of public private partnerships in China and the analysis on questionnaire survey. In: 2008 int conf wirel commun netw mob comput WiCOM, 2008. doi: https://doi.org/10.1109/WiCom.2008.2453.
[37] Tsai T-C, Yang M-L. Risk assessment of design-bid-build and design-build building projects. J Oper Res Soc Japan 2010;53(20–39). doi: https://doi.org/10.15807/jorsj.53.20.
[38] Sarvari H, Valipour A, Yahaya N, Noor NM. Risk identification and assessment in malaysian public-private partnership projects. Vulnerability, Uncertainty, Risk; 2019. p. 436–45. https://doi.org/10.1061/9780784413609.045.
[39] Aziz RF. Ranking of delay factors in construction projects after Egyptian revolution. Alexandria Eng J 2013;52:387–406. doi: https://doi.org/10.1016/j.aej.2013.03.002.
[40] Zou PXW, Zhang G, Wang J. Understanding the key risks in construction projects in China. Int J Proj Manag 2007;25:601–14. doi: https://doi.org/10.1016/j.ijproman.2007.03.001.
[41] Odeh AM, Battaineh HT. Causes of construction delay: traditional contracts. Int J Proj Manag 2002;20:67–73. doi: https://doi.org/10.1016/S0263-7863(00)00037-5.
[42] Tang W, Qiang M, Duffield CF, Young DM, Lu Y. Risk management in the Chinese construction industry. J Constr Eng Manag 2007;133:944–56. doi: https://doi.org/10.1061/(ASCE)0733-9364(2007)133:12(944).
[43] Osama M, Sherif A, Badawy M. Risk analysis of construction of administration projects using Bayesian networks ahead-of-print. J Eng Des Technol 2021. doi: https://doi.org/10.1108/JEDT-04-2021-0222.
[44] Kim G-H, Shin J-M, Kim S, Shin Y. Comparison of school building construction costs estimation methods using regression analysis, neural network, and support vector machine. J Build Constr Plan Res 2013;01:1–7. doi: https://doi.org/10.4236/jbcpr.2013.11001.
[45] Chan APC, Yeung JFY, Yu CCP, Wang SQ, Ke Y. Empirical study of risk assessment and allocation of public-private partnership projects in China. J Manag Eng 2011;27:136–48. doi: https://doi.org/10.1061/(asce)me.1943-5479.0000049.
[46] Bing L, Tiong RL-K, Fan WW, Chew DA-S. Risk management in international construction joint ventures. J Constr Eng Manag 1999;125:277–84. doi: https://doi.org/10.1061/(ASCE)0733-9364(1999)125:4(277).
[47] Oad PK, Kajewski S, Kumar A, Xia B. Bid evaluation and assessment of innovation in road construction industry: a systematic literature review. Civ Eng J 2021;7(179–96). doi: https://doi.org/10.28991/cej-2021-03091646.
[48] Hyari KH, Al-Daraiseh A, El-Mashaleh M. Conceptual cost estimation model for engineering services in public construction projects. J Manag Eng 2016;32. doi: https://doi.org/10.1061/(ASCE)ME.1943-5479.0000381.
[49] Al-Zarooni, Abdou S. Risk management in pre-design stage and its potential benefits for UAE public projects. In: Proc 28th world congr hous challenges 21st century, Abu Dhabi, UAE; n.d.

Dr. Mohamed Badawy has finished his B.Sc. in the Civil engineering department, Cairo University, Cairo Egypt in 1995. He has finished his M.Sc. in the civil engineering department, Ain Shams University, Cairo Egypt in 2005. His Ph.D. in the civil engineering department, Ain Shams University, Cairo Egypt in 2010 He is currently an associate professor at Ain Shams University in the field of Construction Engineering and Project Management. He participated in 16 published papers until 2021.

Dr. Fahad Bin Welah obtained his BS in Civil Engineering in 2007 from King Saud University in Riyadh, Kingdom of Saudi Arabia. In 2009 he joined the University of Birmingham where he obtained his MSc in Construction Management in 2010. He then continued at the same university, where he then obtained his PhD degree in 2017. He is currently an assistant professor at King Saud University in the field of Engineering Sustainability, Construction Engineering and Project Management.

Hisham Hafez is a post-doctorate researcher at the Universitat Politècnica de Catalunya in the research domain of sustainable building materials with a specialization in concrete durability, supplementary cementitious materials, and life cycle assessment. He completed his PhD at Northumbria University, UK in 2021 and has managed during the past 3 years to publish 6 journal articles and 1 conference article in high impact platforms.