IOP Conference Series: Earth and Environmental Science

PURPOSE-LED PUBLISHING™

PAPER • **OPEN ACCESS**

Statistical cost modelling for preliminary stage cost estimation of infrastructure projects

To cite this article: C N Atapattu *et al* 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1101** 052031

Statistical cost modelling for preliminary stage cost estimation of infrastructure projects

C N Atapattu[1], N D Domingo[1] and M Sutrisna[1]

[1]School of Built Environment, Massey University, Auckland, New Zealand

C.Atapattu@massey.ac.nz

Abstract. Reliable and accurate cost estimates are essential to construction projects. They are even more critical in infrastructure projects as they require more time, cost, and public constraints. Therefore, a better cost model is required for infrastructure projects. An extensive literature review was carried out to identify various statistical modelling techniques and models, as well as models developed using these techniques. The literature identified seven statistical modelling techniques. They are; regression analysis, Monte-Carlo simulation, support vector machine, case-based reasoning, reference class forecasting, artificial neural networks, and fuzzy logic. These techniques were all used in various cost models developed for construction projects. According to the analysis of results, neural networks and support vector machine-based models displayed better performance in their cost estimation models. However, it was found that combining several techniques into a hybrid model, for example, the neuro-fuzzy hybrid, can significantly increase these results. Thus, the reliability and accuracy of the current estimation process can be improved with these techniques. Finally, the techniques identified as having better performance can be used to develop a cost estimation model for the preliminary stage. This is because these techniques perform well even though the availability of information is lower. The results of this research are limited to the seven identified techniques and the literature used in the review.

1. Introduction

Cost overrun is considered a global phenomenon in the construction industry, especially in long-run mega-scale projects such as infrastructure projects. [1] researched infrastructure projects in Hong Kong and identified a mean percentage of cost overrun of 32.52% in road projects, 34.83% in rail projects, and 37.48% in bridge and tunnel projects. Considering that infrastructure projects usually experience increased extended construction periods; a proper cost forecasting model is strongly needed for large-scale construction projects worldwide. For example, the project management survey done by KPMG New Zealand observed that the number of projects delivered within budget decreased from 48% to 29% from 2010 to 2017 [2]. It is evident that cost estimation at the pre-contract stage plays a significant role. However, [3] explained that, compared to the other industries, construction projects face difficulty in anticipating the characteristics and the work to be done in the future, considering the time taken for the work. Therefore, it is not straightforward to achieve a reliable and accurate estimate for the project. Furthermore, they emphasized the importance of cost estimation to project development. Researchers also found two types of cost overruns: avoidable and unavoidable. The avoidable costs are the matters that could have been foreseen by the management beforehand. Conversely, unavoidable costs are caused by events that cannot be foreseen beforehand [4]. According to [5], due to this, construction projects must deal with various issues; mainly the waste of funds and investments, losing the expected end-product quality of the outcome, and several other issues that arise due to project delays. Budget estimation at the preliminary stage is commonly developed using historical data. Various traditionally used estimation methods are available at the preliminary stage, where information availability is minor and uncertainty is high [3]. However, the reliability of these traditional models is often questioned because proper judgment is critical in estimation. This statement was emphasized by [6] who suggested both historical trend-based data and competed judgment, based on construction knowledge and experience, are vital components of a proper cost forecasting technique. Thus, in the past few years, researchers have focused on developing new cost models that can incorporate both these components while minimizing the errors and biases of the judgments resulting from the human mind's cognitive behaviour. Hence, researchers adopt various statistical modelling techniques into cost estimation models.

2. Literature review

In recent years there has been an increasing amount of research focused on new cost modelling techniques. This is due to the lack of accuracy in traditional methods. As a result, there were seven basic cost modelling techniques identified through literature: regression analysis, Monte-Carlo simulation, support vector machine (SVM), case-based reasoning (CBR), reference class forecasting (RCF), artificial neural networks, and fuzzy logic [7]. While some researchers developed models based on these basic techniques, it was observed that some studies were performed by combining these techniques or adopting the basics while developing custom-made models. The following sections will review the studies examining each technique.

2.1. Regression analysis

Regression analysis is recognized as a technique that estimates the relationship between the parameters. It considers the relationship between each dependent variable against one or more independent variables. In other words, the model includes the identification of the impact on a dependent parameter when an independent parameter is varied and all other independent parameters are fixed [7]. In contrast, a large and growing body of literature has investigated the drawbacks of regression analysis and emphasized that the technique lacks a straightforward approach to compare the new project with the historical data to identify the most suitable model for the estimation. Also, certain assumptions will have to be made related to the equations to make them suitable for the regression equations. Furthermore, there is difficulty in using a larger number of variable inputs [8]. Furthermore, surveys such as those conducted by [9], emphasized that the regression technique is based on linear nature while the cost data of construction projects are generally non-linear. However, the researchers highlighted that the technique is comprised of analytical and predictive capabilities. In contrast, [10] argued that the relationship between project parameters could be either linear or non-linear.

Using the regression model concept,[11] developed a model to forecast the costs of highway projects using cost indices and historical cost data and it was stressed that a cost forecasting model should be based on price escalation and other influencing factors. [12] developed a linear regression-based cost model to predict the percentage of cost overruns using 46 highway projects. An average percentage error of 30.42% was found in the model. The same research also included a statistical fuzzy-based model using the same data set. The conclusion was that the regression-based model showed better performance than the fuzzy-based model. In another significant study, [13] highlighted the need to have a custom model based on project type rather than using a generic formula for all. Furthermore, the researchers developed a regression model with a 95% probability of accuracy. This was done using 140 building projects in Jordan, including residential villas, apartments, commercial buildings, plants, military projects, and infrastructure projects.

Another investigation was carried out by drawing on the concept of regression analysis; [14] developed six cost models which differ from each other based on the number of variables used in each model. This was based on cost data from 286 building projects in the United Kingdom. Out of the six models, the model that performed best was the log of cost backward model which had an error of 19.3%. The researchers compared the model against Artificial Neural Network (ANN) using the same data and variables to validate the model. The best outcome of the ANN utilized 41 variables and showed a 16.6% of error This implied that the ANN technique is better than the regression model. [15] emphasized that regression analysis displays poor performance when used with the early-stage design parameters. [8] and [16] compared ANN, regression, and SVM. They emphasized that the ANN models produced better output results than the regression models. However, [8] achieved a significantly lower MAER of 5.68 for their regression model. Unlike the other models, [17] focused on variables significant to cost, with numerical relationships that are challenging to identify. For example, the complexity level, the project's importance, on-time competition, and construction complexity level were numerically comprised in the model. However, the model displayed an error of 23.8%.

2.2. Monte-Carlo simulation
By drawing on the Monte Carlo simulation concept, [18] emphasized that considering factors of uncertainty in a project will lead to quantifying the project cost overruns. The previous studies agreed that the simulation could predict any complicated systems where the uncertainty of available information is higher [4][7] adapted the Monte Carlo simulation to develop a cost estimation model for building projects that forecasts the cost overrun percentage. Simulation models are time-consuming in optimization, but they allow for quicker and easier complex approximations through stage-wise refinements. The model they developed predicted that 65 out of the 500 projects would be expecting a 27.49% cost overrun relative to the contract sum. Rather than providing a single outcome, they proposed a range of cost overrun possibilities against the respective occurrence probability.

2.3. Support vector machine (SVM)
This technique is based on structural risk minimization and statistical learning theory. Additionally, models can be developed based on SVM to identify the factors, data sample collection, and the training process [19]. Researchers identified several benefits of using SVM as a cost model. For example, convex optimization issues can be solved quicker than other models and with a reliable and accurate solution [20]. [20] further emphasized that many research studies that developed SVM-based cost models displayed better performance than traditional cost models. Also, the generalization capabilities and sparse representation of SVM are better than other models such as neural networks. Additionally, SVM comprises six significant properties identified by [20] who resolved that SVM is the best modelling technique in conceptual cost estimates. Finally, the research concluded with a conceptual stage cost estimation model for road projects in the Gaza strip. This was done using seventy projects and with a 95% accuracy performance. Contrarily, in an investigation comparing SVM with other models, [8] identified several disadvantages of SVM. The algorithmic complexity required for the model is significantly higher than other models, and thus a widespread memory is essential. Also, the model needs a trial-and-error period to determine a suitable mathematical function, defined as kernel, and the variables of the selected kernel function. Researchers highlighted that the SVM models performed either equally as well or notably better when comparing its results to neural network and fuzzy system-based models. However, the SVM-based model developed for the school building projects in Korea, in the research described above, achieved a MAER of 7.48. In contrast, the ANN model, developed using the same data, achieved better results.

2.4. Case-based reasoning (CBR)
CBR is the process of solving a new case or a project based on previously resolved cases using the knowledge base of the model. Therefore, this model compares the new project with the old data set and finds a similar case [21][22][23]. Several studies investigating this technique described four steps of the CBR model: retrieve, reuse, revise, and retain [7][10][24]. Furthermore, researchers emphasized that the model is based on the hypothesis that similar cases will have similar problems with similar solutions. However, [10] argued that this technique encounters barriers such as distance measurement, selection

of characteristics, assigning the weightage, and thresholds of cases in reuse. In contrast, [24][25][26] suggested that CBR can override most of the shortcomings that regression and neural networks come across by utilizing the lessons learned from historical data to solve issues in future projects.

This technique was demonstrated by [21] when the cost model was developed. In the study, it was developed a model with a MAER of 14% using the cost data from 143 sports field construction projects in Poland. One of the key focuses of this study was the sustainable development factors accounted for by the model, which most other models were lacking. On the other hand, a broader perspective was adopted by [10]. Firstly, several distance measurement tools were adopted, such as the Mahalanobis concept, Euclidean concept, arithmetic summation-based similarity measure, and fractional function-based similarity measures. Then a cost model was developed using 99 multi-family housing complex projects in Korea as the knowledge base. In their study, 3 test rounds were carried out on the several models developed using the measurement concepts. All the models except the Mahalanobis-based model performed well showing a MAER between 0.085 and 0.096. However, the model based on the Mahalanobis concept showed a MAER of 0.31 to 0.47. In their investigation into optimizing the weights of the cost factors of models, [24][26] adopted a genetic algorithm for the CBR model. The former researchers developed a model for building projects in Korea with a minimal error rate of 5.6%. The latter developed a cost model for river facility construction projects with an average error rate of 15.49% using only eight parameters. These findings were also supported by [25] who demonstrated the high forecasting capability of their model with a low MAER of 47.9% for building projects based on the data from South Korea. In another study, [23] developed a cost model based on CBR for sports field construction works while accounting for a MAER of 5.7%. This study further stressed that the model encounters more significant errors in quantifying the numerical impact of the location and inflation factors. The evidence presented regarding this model collectively suggests that it is essential to have a more extensive knowledge base of cases to achieve the desired results.

2.5. Reference class forecasting (RCF)

The concept of RCF is similar to the CBR model. This technique also analyses future cases and finds solutions using historical data of a similar case (Shabniya & Dilruba, 2017). However, [27] highlighted that RCF and CBR differ to certain degrees as CBR does not consider the irrationality of human behaviour, rather, its outcome depends on expert judgment. Furthermore, [28] recommends RCF at the early stage of estimation when the uncertainty is at its highest. These findings were supported by [28] in their model for estimating cost overruns of public construction projects in Turkey. They found the average cost overrun to be 11.33% but this could vary between 22.94% to 133.48%. Another study conducted by [29] identified that 60% of the Icelandic transportation projects face 95% cost overruns from their initial cost plan. Although the RCF model developed in the study provides a comprehensive insight into this matter, the researchers emphasized that the current method used in Icelandic Road Administration (ICERA) cost forecasting was able to address the matter. This was with a 6% cost overrun for the five years of the study. In contrast, [30] studied RCF model application in a Danish mega railing project. In his study, he concluded that the model did not support the estimation accuracy while the project faced cost and time overruns. The research further emphasized that the model was not able to prevent strategic misrepresentation as well as optimism bias.

2.6. Artificial neural network (ANN)

According to [31][32], ANN is one of the most sophisticated models capable of non-linear approximation and modelling complex functions. However, a lot of research has stressed that most of the modelling techniques, such as regression, are not capable of non-linear approximation, thus, most of the previous models are not valid. It was also identified that ANN could use the lessons learned from previously completed cases and generalize the knowledge to future projects [19][32][33]. Additionally, [19] stated that the training process of ANN is overly complex as it is attained through a gradient descent algorithm on the error space. Furthermore, they also found that the genetic algorithm and Particle Swarm Optimisation (PSO) can be used to overcome this issue. The ANN model comprises the input, hidden, and output layers. The input layer may consist of several neurons depending on the input parameters selected for the model. The hidden layer calculates the cost based on the input parameters provided being compared against each parameter's relationship to the cost. Finally, the output layer typically consists of one neuron which provides the project's estimated cost [8][15]. Further, the research emphasized the three stages of model development: the modelling phase, the training phase, and the testing phase. The modelling phase involves analyzing the collected projects' cost data, variable establishment, network relationship design, and rules. The model will be prepared based on the data in the training stage, and then its performance during the testing phase will be evaluated. Considering these facts, [34] developed a cost model which adapted the ANN concept using 18 highway projects in Canada. This predicted the cost with a weighted error of 0.98. Another study was carried out by [33] who developed a model based on ANN using the data from 75 highway projects in Egypt. The study carried out several training, evaluation, and validation tests to measure the model's performance, the results were found to be 4.51%, 5.8%, and 16.0%, respectively. All were within the valid range. Meanwhile, [32] identified fifteen variables in his research developing two models for predicting the time and cost overrun based on the ANN technique through the "NeuroSolutions (6.12) (2012)" software. The model was based on the data from 56 highway projects in Egypt. The average error was concluded to be 39.8%. This is comparatively more accurate than the model developed by [12] using the same data set based on linear regression and fuzzy statistical models.

[35] took a different approach to neural networks. They combined genetic algorithms and neural networks to optimize the model in selecting the number of neurons in the hidden layer. In this study, several models adjusted the number of layers and neurons. The best model consisted of 2 hidden layers, the number of neurons was 12 in the input layer, 18 in the first hidden layer, 12 in the second hidden layer, and 1 in the output layer. This model achieved a MAER of 2.62. The study concluded that the genetic algorithm is better for selecting parameters than the trial-and-error method. As discussed previously, the primary study carried out by [8], comparing three modelling techniques, concluded that the neural network was better for cost forecasting than regression and SVM. This is because it can deal with multifaceted problems while producing user-friendly models. The model developed in the study done by [8] which was based on neural networks achieved a MAER of 5.27. Similarly, [36] compared ANN with regression analysis by developing two models with the same data set from Poland and Thailand. As expected, the ANN-based model exhibited a lower estimation error rate than the regression-based model. However, the model developed by [15] achieved the lowest error of +3.8%. This model was developed for the early-stage cost estimation of apartment buildings; however, this model was focused on design and site-related variables only. Therefore, the contract-related parameters, such as tendering strategy, contract type, procurement methods, and the like, were not considered.

2.7. Fuzzy set theory
Fuzzy logic is recognized as a model used for human communication, reasoning, and decision-making abilities in a formalized way where the information is conflicting, incomplete, and uncertain [31][37]. Nevertheless, [37] further added that the fuzzy models are used in risk assessment, range estimating, forecasting construction performance, contractor selection, working condition assessment, and the determination of cost estimating relationships.

2.8. Neuro-fuzzy hybrid model
In recent years, researchers have started combining several modelling techniques to compare their performance with the mono-technique models. [31] developed a model by combining the ANN and fuzzy set theory and considering the data of 98 water infrastructure projects completed between 2007 to 2011 in Scotland. They adapted ANN’s learning and generalization capabilities and combined them with the positive advantages of fuzzy logic. This includes reasoning and decision-making abilities of the human brain with incomplete information. The model forecasted the final cost with 0.6% of underestimation error and 0.8% of overestimation error, this is a significantly low marginal error.

3. Research methodology

Initially, relevant research articles were gathered from the databases. Then, significant articles were selected from the database. Following this, an in-depth literature review was carried out using significant research articles. Finally, the findings were drafted using the content analysis, and conclusions were made.

4. Findings and discussion

Table 1 displays the finding of the previous studies. It summarises the models used for infrastructure projects identified in the study, the variables considered in the models, and their performances. According to the results, ANN's show lower performance when the model considers more variables. For example, the two ANN models developed by [38] reflect six primary and five secondary variables. However, the models show a higher rate of errors ranging from 33% – 50%. However, [33] created a model that achieved significantly better performance using only six technical variables. Similar behaviour was noticed with regression analysis and SVM. The regression model with fewer variables achieved ±5% marginal error [39]. Conversely, [36] model shows a +24% error, although it considers only seven numerical variables. Meanwhile, it was observed that the two hybrid models achieve better results than all the other models. For example, one model combines ANN and fuzzy logic, while the other combines regression analysis, SVM, and data mining techniques. Furthermore, it was noticed that combining several modelling techniques helps to mitigate the disadvantages of one technique with the advantages of the other technique [31][40]. This research is expected to extend toward developing a reliable model for road/ highway projects in New Zealand. This literature review emphasised several vital areas that need to be addressed by future models. Firstly, the selection of variables is crucial. A thorough study is required to identify specific variables that will significantly contribute to the project's cost. The second step is to select the proper techniques. According to the results of this review, it is better to combine several techniques into one model, however, it is not easy to select which techniques should be combined. Therefore, an in-depth study of each technique will be required to identify the best methods suitable for the model.

Table 1. Models developed for infrastructure projects

Model	Reference	Project type	Variables	Performance
ANN	[33]	Highways	Project scope, project duration, year of construction, project region, mainline length, mainline classification	Training - +4.51%; Evaluation- +5.8%; Validation - +16.0%
	[32]	Highways	<u>*Independent variables –*</u> Inadequate project planning and execution, inadequate cost planning and monitoring, lack of communication between construction parties, price fluctuations, lack of proper technical study before the tender by the contractor, errors in project quantities measurements, slow decision-making process, equipment failures, lack of adequate field visits before tendering by the contractor, inappropriate use of project site, improper use of materials, inaccurate drawings and contract documents, material monopoly by suppliers, inflation, and rework. <u>*Dependent variables -*</u>	+39.8%

			Percentage of cost overruns of the constructed highway projects.	
	[36]	Roads	Predominant work activity, work duration, pavement width, shoulder width, ground rise fall, average site clear, earthwork volume, surface class, and the base material.	+24% to +26%
	[38]	Urban railways	*Primary parameters* Percentage of tunnel section over the total length of the rail, percentage of total length of elevated stations over the total rail length, percentage of total length of at-grade stations over the rail length, percentage of total length of cut-and-fill method over the main line length, supply and installation of the rails, and number of underground stations. *Secondary parameters* Contract type, number of at-grade stations, number of elevated stations, the main line length, and percentage of total length of depressed-open sections (ramps) to the total rail length.	Model 1 - +49.8%; Model 2 - +33.3%
Regression analysis	[39]	Highway projects	Final cost, awarded bid, days used, actual contract time, and initial duration.	±5%
	[12]	Highway projects	Same as [32]	+30.42%
	[36]	Road projects	As mentioned under ANN.	+30% to +36%
	[38]	Urban railway projects	As mentioned under ANN.	+35.2%
SVM	[20]	Road projects	Road area, road surface type, base course type, base course thickness, interlock thickness, asphalt thickness, pipe diameter, manhole depth, cut quantity, fill quantity, curb length, electrical	-5% average error
Fuzzy logic	[12]	Highway projects	Same as [32]	+40.37%
ANN & Fuzzy hybrids	[31]	Water infrastructure	Tendering strategy, site access, type of location, project type, Contractor's need, soil type, initial cost, the initial duration.	+0.6% to +0.8%
Regression, SVM & data mining hybrids	[40]	Water infrastructure	Tendering strategy, procurement option, ground condition, soil type, delivery partner, scope, the purpose of the project, and operating region.	+2.33% to -3.83%

5. Conclusion and further research

A construction project's cost estimation must be accurate and reliable, especially when the project's uncertainty is high and little information is available. Therefore, it is crucial to develop a realistic budget as it affects most decisions related to design, construction, and technology. Cost overrun in construction projects is a global phenomenon due to a lack of accuracy in traditional estimation techniques. Therefore, in the past few year's researchers have focused on developing cost models using statistical techniques. Seven powerful techniques have been utilized: regression analysis, Monte Carlo simulation, CBR, RCF, SVM, ANN's, and fuzzy logic. These techniques were used either as a single technique-based model or multi-modelling techniques combined as hybrids. It was found that these techniques could improve estimation accuracy compared to traditional methods. Moreover, the most reliable modelling techniques with better performance were ANN's, SVM's, regression analysis, and hybrids. Although CBR and RCF are also good modelling techniques, they must have a larger number of project cost databases to provide more reliable output. Based on the performance of these techniques against the infrastructure projects, the Neuro-Fuzzy hybrid can be identified as the most accurate model overall, while ANN and SVM techniques also showed better results as mono-technique-based models. Therefore, to mitigate the issues in initial budgeting and estimation of infrastructure projects, it is recommended to incorporate these techniques in estimation.

References

[1] Huo T, Ren H, Cai W, Shen G Q, Liu B, Zhu M and Wu H 2018 Measurement and Dependence Analysis of Cost Overruns in Megatransport Infrastructure Projects: Case Study in Hong Kong *Journal of Construction Engineering and Management.* **144**

[2] KPMG NZ 2017 Driving Business Performance - Project Management survey.

[3] Challal A and Tkiouat M 2012 The design of cost estimating model of construction project: Application and simulation *Open Journal of Accounting.* **1** 15-26

[4] Shanmugam M, Amaratunga R and Zainudeen N 2006 Simulation modelling of cost overruns in building projects. Working paper

[5] Shehu Z, Endut I R, Akintoye A and Holt G D 2014 Cost overrun in the Malaysian construction industry projects: A deeper insight *International Journal of Project Management.* **32** 1471-1480

[6] Skitmore M and Marston V 2005 *Cost Modelling*. Routledge

[7] Shabniya V and Dilruba K M 2017 A review on construction cost forecasting techniques *International research journal of Engineering and Technology*. **4** 3333-3339

[8] Kim G H, Shin J M, Kim S and 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.* **1** 1-7

[9] Vahdani B, Mousavi S M, Mousakhani M, Sharifi M and Hashemi H 2012 A neural network model based on support vector machine for conceptual cost estimation in construction projects *Journal of Optimization in Industrial Engineering.* **6** 11-18

[10] Ahn J, Park M, Lee H S, Ahn S J, Ji S H, Song K and Son B S 2017 Covariance effect analysis of similarity measurement methods for early construction cost estimation using case-based reasoning *Automation in Construction.* **81** 254-266

[11] Herbsman Z 1986 Model for forecasting highway construction cost. *Transportation Research Record.* **1056** 47-54

[12] El-Maaty A E A, El-Kholy A M and Akal A Y 2017 Modelling schedule overrun and cost escalation percentages of highway projects using fuzzy approach *Engineering, Construction and Architectural Management.* **24** 809-827

[13] Hammad A A A, Ali S A, Sweis G J and Bashir A 2008 Prediction model for construction cost and duration in Jordan *Jordan Journal of Civil Engineering.***2** 250-266

[14] Lowe D J, Emsley M W and Harding A 2006 Predicting construction cost using multiple regression techniques *Journal of construction engineering and management.* **132** 750-758

[15] Günaydın H M and Doğan S Z 2004 A neural network approach for early cost estimation of structural systems of buildings *International Journal of Project Management.* **22** 595-602

[16] Wang Y R, Yu C Y and Chan H H 2012 Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models *International Journal of Project Management.* **30** 470-478

[17] Ganiyu B and Zubairu I 2010 Project cost prediction model using principal component regression for public building projects in Nigeria *Journal of Building Performance ISSN.* **1** 21-28
[18] Touran A and Lopez R 2006 Modelling cost escalation in large infrastructure projects *Journal of construction engineering and management.* **132** 853-860
[19] Cheng M Y and Hoang N D 2014 Interval estimation of construction cost at completion using least squares support vector machine *Journal of Civil Engineering and Management.* **20** 223-236
[20] El-Sawalhi N I 2015 Support vector machine cost estimation model for road projects *Journal of Civil Engineering and Architecture.* ***9***
[21] Agnieszka L and Krzysztof Z 2018 Cost Calculation of Construction Projects Including Sustainability Factors Using the Case Based Reasoning (CBR) Method *Sustainability.* **10** 1608-1608
[22] Tah J H M, Carr V and Howes R 1998 An application of case-based reasoning to the planning of highway bridge construction *Engineering, Construction and Architectural Management.* **5** 327-338
[23] Zima K 2015 The Case-based Reasoning Model of Cost Estimation at the Preliminary Stage of a Construction Project *Procedia Engineering.* **122** 57-64
[24] Kim S and Shim J H 2014 Combining case-based reasoning with genetic algorithm optimization for preliminary cost estimation in construction industry *Canadian Journal of Civil Engineering.* **41** 65-73
[25] Won-Gil H, Sangyong K and Jung-Kyu J 2019 Improved similarity measure in case-based reasoning: a case study of construction cost estimation *Engineering, Construction and Architectural Management.* **27** 561-578
[26] Lee S, Jin Y, Woo S and Shin D H 2013 Approximate cost estimating model of eco-type trade for river facility construction using case-based reasoning and genetic algorithms *KSCE Journal of Civil Engineering.* **17** 292-300
[27] Awojobi O and Jenkins G P 2016 Managing the cost overrun risks of hydroelectric dams: An application of reference class forecasting techniques *Renewable and Sustainable Energy Reviews.* **63** 19-32
[28] Bayram S and Al-Jibouri S 2016 Application of Reference Class Forecasting in Turkish Public Construction Projects: Contractor Perspective *Journal of Management in Engineering.* **32**
[29] Fridgeirsson T V 2016 Reference class forecasting in Icelandic transport infrastructure projects *Transport Problems.* **11** 103-115
[30] Themsen T N 2019 The processes of public megaproject cost estimation: The inaccuracy of reference class forecasting *Financial Accountability and Management.* **35** 337-352
[31] Ahiaga-Dagbui D D, Tokede O, Smith S D and Wamuziri S 2013 A neuro-fuzzy hybrid model for predicting final cost of water infrastructure projects In *29th Annual ARCOM Conference* (pp. 181-190) Association of Researchers in Construction Management Reading, UK
[32] El-Kholy A 2019 Exploring the best ANN model based on four paradigms to predict delay and cost overrun percentages of highway projects *International Journal of Construction Management.* **21** 694-712
[33] Adel K, Elyamany A, Belal A M and Kotb A S 2016 Developing parametric model for conceptual cost estimate of highway projects *International Journal of Engineering Science.* **6** 1728-1734
[34] Tarek H and Ayed A 1998 Neural network model for parametric cost estimation of highway projects *Journal of Construction Engineering & Management.* **124** 210-218
[35] Kim G H, An S H and Kang K I 2004 Comparison of construction cost estimation models based on regression analysis, neural networks, and case-based reasoning *Journal of building and environment.* **39** 1235-1242

[36] Sodikov J 2005 Cost estimation of highway projects in developing countries: artificial neural network approach *Journal of the Eastern Asia Society for Transportation Studies*. **6** 1036-1047
[37] Fayek A R and Rodriguez Flores J R 2010 Application of fuzzy logic to quality assessment of infrastructure projects at conceptual cost estimating stage *Canadian Journal of Civil Engineering*. **37** 1137-1147
[38] Sonmez R and Ontepeli B 2009 Predesign cost estimation of urban railway projects with parametric modeling *Journal of Civil Engineering and Management.* **15** 405-409
[39] Shr J F and Chen W T 2006 Functional model of cost and time for highway construction projects *Journal of Marine Science and Technology.* **14** 127-138
[40] Ahiaga-Dagbui D D and Smith S D 2014 Dealing with construction cost overruns using data mining *Construction management and Economics.* **32** 682-694