

**COMPARATIVE ANALYSIS OF MACHINE LEARNING TECHNIQUES USED IN CONSTRUCTION SITES: A SURVEY**

Mohammad Kabir Yaqubi <sup>1</sup>, Sandeep Salhotra <sup>2</sup>

<sup>1</sup>ME Research Scholar, Civil Engineering Department, Chandigarh University, [kabir.yaqubi@yahoo.com](mailto:kabir.yaqubi@yahoo.com),

<sup>2</sup>Professor, Civil Engineering Department, Chandigarh University, Punjab, India, [hod.civil@cumail.in](mailto:hod.civil@cumail.in)

**Abstract:** -Construction Management (CM) should deal with a number of uncertainties related to Time, Cost, Quality, and Safety and so on. Such uncertainties make the entire construction process much unacceptable. Hence there is a use of artificial intelligence came into purview to solve the problem in an effective way. Cost estimation is the most important preliminary process in any construction project. Therefore, construction cost estimation has the lion's share of the research effort in construction management. In this paper, we have analyzed and studied proposals for construction cost estimation for the last 10 years. From the research it has been observed that different researchers have worked on cost estimation using artificial intelligence schemes. It is highlighted that the researchers used mostly SVM, ANN and GA as an artificial intelligence techniques that helps to find an appropriate solution to the given construction problem. From the analysis it has been examine that SVM perform well as compared to other existing techniques.

**Keywords—** Construction sites, Cost, Time, Quality, Artificial Intelligence, ANN, SVM, GA, ACO.

**I. INTRODUCTION**

Before starting any construction work, the owner or the builder must know about the volume of work. He must know about the time as well as the cost required to complete the work. Therefore, it is essential to calculate cost as per the plan and specification of the project. Otherwise, the project might be stopped due to the unavailability of funds / materials [1]. The cost estimation can be defined as the computing process to determine the overall cost of a construction project. The cost is determined by calculating the quantities from the dimensions as per the drawing. The cost estimation can be defined as the process to calculate the cost required to complete a construction project. If an accurate cost is not judged then it leads to numerous issues like, delay in construction. The problem occurs in manually calculating construction cost inspired researchers as well as the construction industries to find out an intelligent solution to resolve the cost estimation problem [2]. Presently information technology plays an essential role to enhance the efficiency of the construction projects. The intelligent approaches are used to handle different issues like choosing the best and qualified contractor, analyzing the performance of the projects at various phases, calculating hazards for cost overruns. Presently, artificial intelligence (AI) technique is used by civil engineers to resolve the above mentioned problems. Initially the AI technique has been used by Holm et al. in 2005[3].

In fact, there are a large number of intelligent strategies available to address building management problems. It encourages scholars to conduct and study the intelligent approaches to addressing the problem of cost building estimations. This paper analyzes the wise solutions used in the last decade and recognizes future development directions. This will help to present more accurate and deeper study of the latest proposals [4]. The investigative process will emphasize the research space in this area. Additionally, it opens a door to identify available opportunities for future research.

TABLE 1  
DIFFERENCE BETWEEN MANUAL COST ESTIMATION AND AUTOMATED COST ESTIMATION

Manual cost estimation	Automated cost estimation
It mainly require experienced person	
It is tedious	Reducing the subjectivity in estimation
Cumbersome and error-prone	Improving the efficiency and the accuracy of cost estimation.
The cost is affected by every estimator's opinion	

This paper is categorized into three subparts, initially; a state-of-art is studied on different intelligent schemes that have been used by various researchers in the field of civil engineering to optimize the construction cost. In the literature survey, best papers like Springer journal, Elsevier, SCI indexing, IEEE journals are selected to study the different cost estimation process. Also, the advantages and disadvantages of intelligent algorithms in the field of construction cost estimation are examined. Also, the cost affecting factors have also been studied to continue a particular benchmarking process [5]. The process followed in this paper is shown in figure 1.

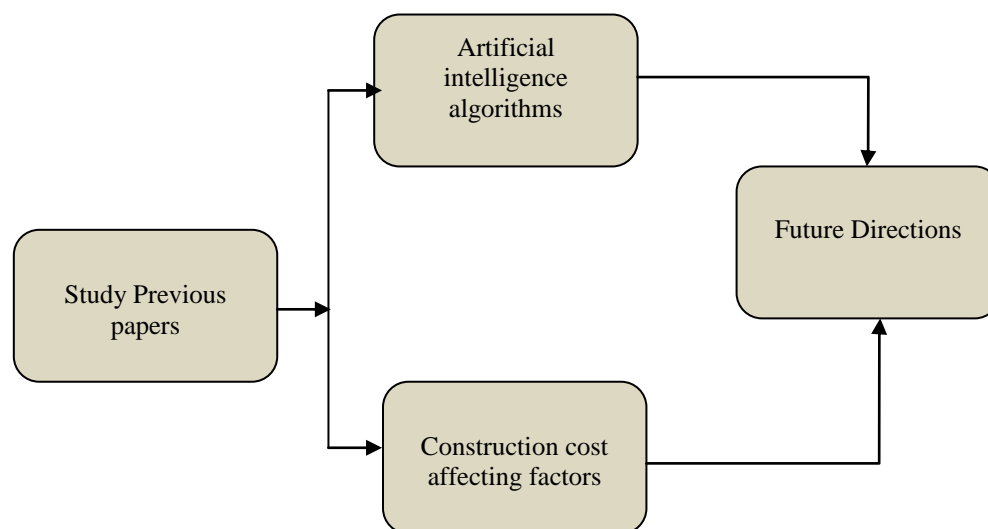


Fig. 1 Designed structure of paper

## II. LITERATURE REVIEW

The importance of cost estimation of building projects has been introduced by Ferry et al [6] in the year 1999. Furthermore, in the year 1991 Newton [7] surveyed over 60 cost structures and apply regression method to enhance the performance of cost model. In both the research, no one can used neural network later in 1988 Hegazy et al. [8] have introduced a neural network (NN) algorithm to administrate construction cost and build up a parametric cost-estimating structure for highway projects. The authors have considered 18 cases of highway projects build in Canada as an input data to the NN. In combination to NN authors have used two optimization algorithm methods such as simplex optimization and genetic algorithm (GA). These techniques are used to optimize the weight and help to determine the best cost. In this paper, author mainly introduced the concept how to use spreadsheet program in the designing of accurate NN structure that can be utilized in construction sites. Later, this concept was used by Al-Tabtabai et al. [9] in the year 1999 to find the relation between dynamic construction cost and the factors that causes to change the construction parameters in highway construction by utilizing the concept of neural network. Here, the researchers have defined a threshold value which is considered as a reference and if the value become more than the defined value then it is considered as a percentage increase in the construction cost. Emsley et al. [10] in 2002 has presented the use of NN in the construction cost estimation. It has been examined that construction cost can be analyzed using two metrics such as (i) Cost per  $m^2$  and (ii)  $\log$  of cost /  $m^2$ . From the test it has been observed that  $\log$  metrics diminish the proportional difference and the later metrics can reduce the square of error. Therefore, the model can be designed by considering these two parameters.

## III. COMPARISON OF EXISTING INTELLIGENT TECHNIQUES

In this section, different machine learning techniques used in the cost estimation process was conducted. The comparison of diverse existing techniques along with justification is defined in the tabular form.

TABLE 2  
 COMPARISON OF EXISTING MACHINE LEARNING TECHNIQUES AND JUSTIFICATION

References	Proposed work	Proposed technique	Justification
[11]	Proposed an automated cost over time in Highway project	ANN	It has been analyzed that the cost of construction in Louisiana has become twice in between the years 1998 to 2015.
[12]	Proposed SVM algorithm to determine the construction cost.	SVM (Support Vector Machine) algorithm	From the results it has been concluded that SVM perform better than existing traditional approach.
[13]	Proposed a cost estimation solution for road tunnel construction sites	ANN Data has been collected from 33 tunnels of 46 km long (1998-2004). Neural network used in two types such as multiplayer feed forward and general regression NN.	Comparison with other existing technique has been provided

[14]	Proposed ANN to judge the seismic retrofit construction cost	The data has been gathered from 158 earthquake private school building Here more number of hidden layers with large number of neurons has been used.	ANN required more computation time and also the complexity is high due to the large number of layers.
[15]	Proposed a framework to examine execution time as well as cost for earthwork.	ANN with different neurons has been used	The execution time and cost obtained for the presented work are 54m-h and 121,159 for two number of neurons
[16]	Proposed an optimization algorithm to resolve the resource problem	Genetic algorithm	By comparing the result with case studies from literature
[17]	Proposed ACO technique to resolve the time-cost Multiobjective optimization problem	Ant colony algorithm	By comparing the results with case studies in construction optimization
[18]	Proposed a cost estimation approach using ANN in concrete structure of 4 to 8 story building.	ANN algorithm	The average accuracy of cost estimation upto 93 % has been achieved. MSE is 0.038
[19]	The model has been designed for estimating the time cost of 75 construction projects from the year 1999 to 2011.	SVM	The average accuracy after validating the structure data up to 97.4 % has been obtained. The percentage error is 2.17
[20]	Develop a model to design an efficient and managed construction space in high tower building in crowded area.	GA	The computation cost has been reduced with the enhanced productivity.
[21]	Work on to minimize time- cost for construction sites	GA	The performance in terms of quality up to 92 % has been obtained. The construction cost has been reduced from 13.387 <sup>€</sup> to 9.133 <sup>€</sup> . The time has been minimized using GA from 280 days to 269 days

#### IV. ARTIFICIAL INTELLIGENCE SCHEMES

There are number of machine learning schemes that can be used to estimate the construction cost automatically. The scheme includes: SVM, ANN, Linear regression, and GA (Genetic algorithm). Some of them are described below:

##### A. Artificial Neural Network

Neural network (NN) is a computer structure that simulates and work similar to that of the human brain learning process. NN is mostly used in many industries, including construction area. Application of NNs to construction has been widely studied. In addition, researchers have examined the application of NNs to increase the accuracy of the cost estimation rather than the regression model. The results of several studies have shown that NNs are superior to the regression model while estimating cost.

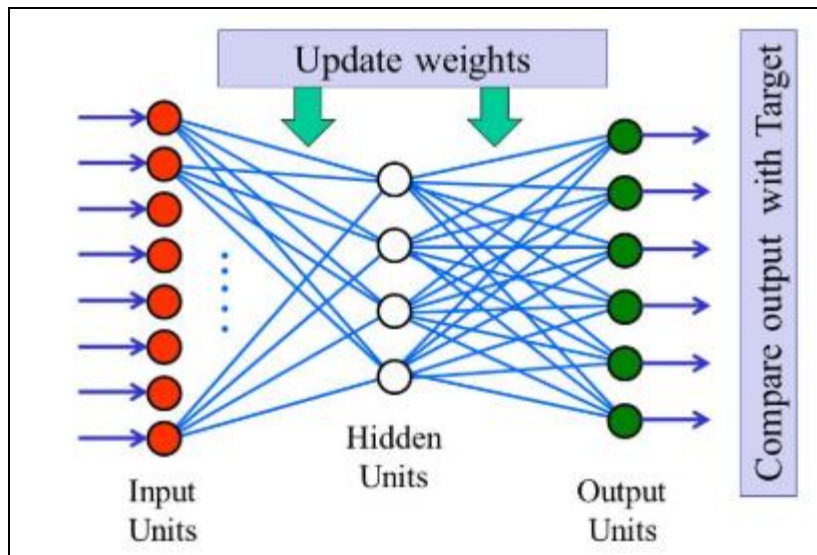


Fig. 2 ANN architecture

The ANN structure comprises of three layers that consists of number of nodes represented by different colors. Input layer includes sensor nodes denoted by red color, output layer neurons are represented by green color. The nodes are interconnected with each other to transmit data from one node to other node. Weight function is added to modify the output as per the desired input. The weight function increase/ decrease the signal strength which regulate the learning process. ANN simply resolves the problem in the similar way as that of human brain [22].

**B. Linear Regression Approach**

This technique is used to find the relation between a scalar dependent variable “a” and more than one descriptive variables denoted by “b”. If there is only one descriptive variable then it is known as “simple linear regression” whereas for multiple descriptive variables the technique is named as multiple linear regressions.

In a linear regression scheme, linear function is used to set the data points. Mathematically, linear function can be represented as:

$$A = \beta_0 + \beta_1 b_1 + \beta_2 b_2 + \dots + \beta_n b_n \quad (1)$$

Here, A is the target variable

$b_1, b_2, \dots, b_n$  are the predictor variables with  $\beta_0, \beta_1, \dots, \beta_n$  coefficient.

This technique mainly works on linear data and do not work on the real world data (non-linear data). Linear model have advantages such as: (i) Simple (ii) Require small memory space.

In case of construction sites to determine the construction time one can use “Bromilow’s time cost structure as written below.

$$T = k \cdot P^B \quad (2)$$

Here, T is the contracted Time

P signify the contracted price

K is known as the model parameter which is used to determine the productivity.

B represents the time dependence of change in construction cost [23]

Linear regression an also be represented in logarithmic form and mathematically can be written as

$$\ln T = \ln K + B \ln P \quad (3)$$

**C. SVM (Support Vector Machine)**

In recent years, civil engineering has focused on artificial intelligence. SVM is a very new, rapidly developing area of research. SVMs are exemplified by the sample classification, functional adaptation and regression issues. SVM is sub type of machine learning schemes that have been proofed to be a better classification algorithm as well as regression tasks over the traditional ANN method in recent years. SVM was initially developed by vapnik that mainly consists of two categories such as classification and regression. SVM is a supervised learning approach that is used to reduce the structure risk in construction field. The most appropriate representation is recognized as the feature selection feature. A feature set describing a situation (that is, a predictive value) is called a vector. SVM refers to kernel methods, in which the transformation of nonlinear problem is converted into linear problem by utilizing the kernel function. This is done by entering the input field into the multifunctional access area. Whether the problem is a type of classification or regression problem will be addressed in the new multipurpose space as shown in fig.3.

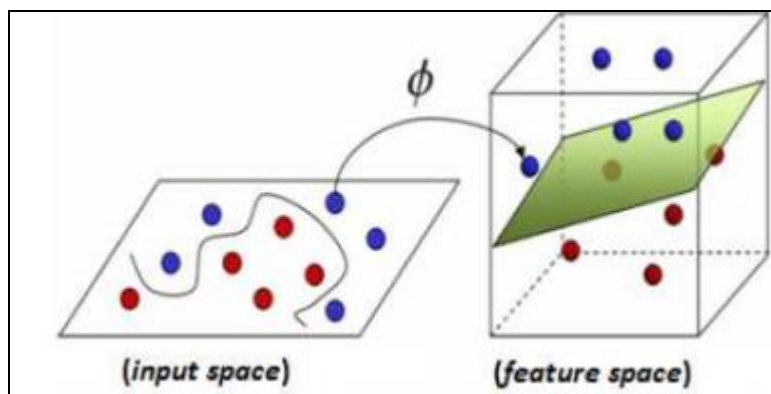


Fig. 3 Mapping in multidimensional space

The problem of differentiating the input data is shown in fig. 3. From the figure it is clear that input space is non-linear that is the data is available randomly. After applying SVM technique the data is formed into two groups separated by Green Square.

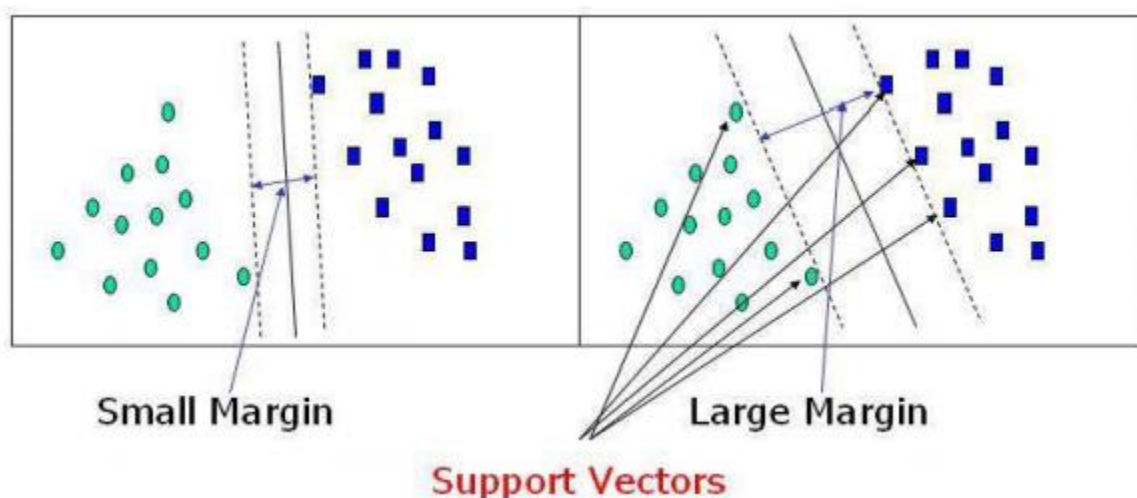


Fig. 4 Margin and support vector

Fig. 4 signify the margin along with the support vector. The left hand side graph represents the support vector with small margin whereas the right hand side graph denotes the data separated by large margin. More is the margin, more will be the support vector and hence increase the accuracy of the system [24].

D. Genetic Algorithm

Genetic algorithm is a helpful tool to plan and monitor the project's activity. GA is a search and optimization mechanisms that help decision makers to find out optimal solutions in large search area. The algorithm is inspired by the mechanism of evolution, and the structural change of genetic material between the members of the population as a basic mechanism for the search of consecutive generations (Goldberg, 1989). Thus, the model is implemented in three key stages: 1) Initialization phase, which creates an initial set of possible solutions for the problem of using the resource; 2) an evaluation phase that calculates the value, time, cost and quality of each created solution; and 3) population creation stage in which the fitness function is improved over the successive generations. Thus, GAs makes a random, yet still directed search for a global optimal solution. Usually, GA encodes the solution into a linear string which is known as chromosome. This can be viewed in a linearly organized box as depicted in figure 5. Every box contains a number known as gene that has to be worked on the project. The flow of GA is shown in fig. 6.

0	1	0	1	0	1	0
1	0	0	1	1	0	1
0	1	0	1	0	1	0

Fig. 5 Representation of genes in binary form

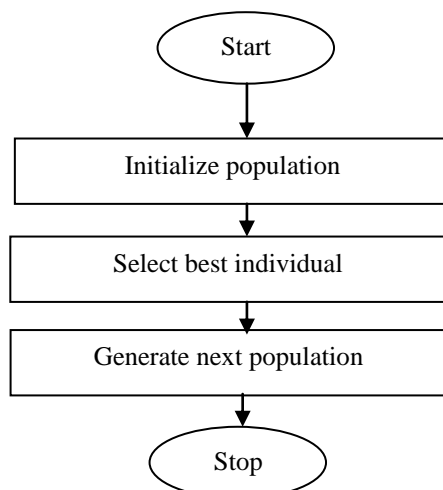


Fig. 6 Flow of GA algorithm

### V. COMPARISON OF EXISTING TECHNIQUES

In this section, the comparison of accuracy to estimate the construction cost as well as time using different artificial intelligence scheme such as artificial neural network, support vector machine and genetic algorithm have been illustrated in table 2 and fig. 7.

TABLE 2  
 COMPARISON OF EXISTING TECHNIQUES

References	Technique used	Accuracy (%)
[18]	ANN	93
[19]	SVM	97.4
[21]	GA	92

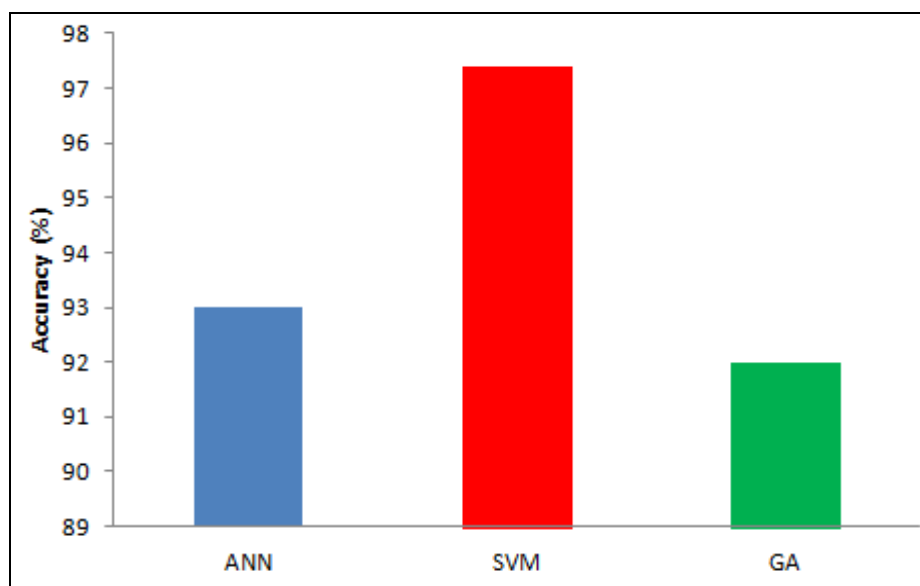


Figure 7 Comparison of exiting techniques

Fig. 7 demonstrates the comparison of artificial intelligence schemes in terms of accuracy parameters while using in the construction sites. The accuracy in terms of percentage has been measured by authors Günaydın et al [18], Petrusseva et al. [19] and Sorrentino & M [21] and represented by the blue color, the red color and by the green color respectively. From the graph it is clear that SVM perform better as compared to ANN and GA algorithm. The accuracy rate of SVM has been increased by 4.73 % from the ANN algorithm and from GA the accuracy has been increased by 5.87%.



*VI. CONCLUSION*

In this paper, a study and analysis were conducted in various proposals to address the development cost of building cost estimation based on intelligent techniques. The first part was about a literary study to examine the current state of intelligent solutions in the construction industry. Regarding this issue, we have exclusively selected magazines specializing in information technology and building management over twenty-year period from 1998 to 2018. In the context of the research, a twenty-year period is sufficient to surround research directions in a given field. From the analysis it has been concluded that SVM perform well among GA and ANN technique. In future, we can use ANN or SVM as a classification algorithm to estimate construction cost. Or GA can also be used for the optimization of construction cost.

REFERENCES

- [1] Barg, S., Flager, F., & Fischer, M. (2018). An analytical method to estimate the total installed cost of structural steel building frames during early design. *Journal of Building Engineering*, 15, 41-50.
- [2] Volk, R., Luu, T. H., Mueller-Roemer, J. S., Sevilimis, N., & Schultmann, F. (2018). Deconstruction project planning of existing buildings based on automated acquisition and reconstruction of building information. *Automation in construction*, 91, 226-245.
- [3] Holm, L., Schaufelberger, J. E., Griffin, D., & Cole, T. (2005). *Construction cost estimating: process and practices* (pp. 3-12). Upper Saddle River, New Jersey: Pearson Prentice Hall.
- [4] Saieg, P., Sotelino, E. D., Nascimento, D., & Caiado, R. G. G. (2018). Interactions of building information modeling, lean and sustainability on the architectural, engineering and construction industry: a systematic review. *Journal of Cleaner Production*, 174, 788-806.
- [5] Ganbat, T., Chong, H. Y., Liao, P. C., & Wu, Y. D. (2018). A Bibliometric Review on Risk Management and Building Information Modeling for International Construction. *Advances in Civil Engineering*, 2018.
- [6] Ferry, D. J., Brandon, P. S., & Ferry, J. D. (1999). Cost planning of buildings.
- [7] Newton, S. (1991). An agenda for cost modelling research. *Construction Management and Economics*, 9(2), 97-112.
- [8] Hegazy, T., & Ayed, A. (1998). Neural network model for parametric cost estimation of highway projects. *Journal of Construction Engineering and Management*, 124(3), 210-218.
- [9] Al-Tabtabai, H., Alex, A. P., & Tantash, M. (1999). Preliminary cost estimation of highway construction using neural networks. *Cost Engineering*, 41(3), 19.
- [10] Emsley, M. W., Lowe, D. J., Duff, A. R., Harding, A., & Hickson, A. (2002). Data modelling and the application of a neural network approach to the prediction of total construction costs. *Construction Management & Economics*, 20(6), 465-472.
- [11] Wilmot, C. G., & Mei, B. (2005). Neural network modeling of highway construction costs. *Journal of construction engineering and management*, 131(7), 765-771.
- [12] An, S. H., Park, U. Y., Kang, K. I., Cho, M. Y., & Cho, H. H. (2007). Application of support vector machines in assessing conceptual cost estimates. *Journal of Computing in Civil Engineering*, 21(4), 259-264.
- [13] Petroutsatou, K., Georgopoulos, E., Lambropoulos, S., & Pantouvakis, J. P. (2011). Early cost estimating of road tunnel construction using neural networks. *Journal of construction engineering and management*, 138(6), 679-687.
- [14] Jafarzadeh, R., Ingham, J. M., Wilkinson, S., González, V., & Aghakouchak, A. A. (2013). Application of artificial neural network methodology for predicting seismic retrofit construction costs. *Journal of Construction Engineering and Management*, 140(2), 04013044.
- [15] Hola, B., & Schabowicz, K. (2010). Estimation of earthworks execution time cost by means of artificial neural networks. *Automation in Construction*, 19(5), 570-579.
- [16] Ghoddousi, P., Eshtehardian, E., Jooybanpour, S., & Javanmardi, A. (2013). Multi-mode resource-constrained discrete time-cost-resource optimization in project scheduling using non-dominated sorting genetic algorithm. *Automation in construction*, 30, 216-227.
- [17] Afshar, A., Ziaraty, A. K., Kaveh, A., & Sharifi, F. (2009). Nondominated archiving multicolony ant algorithm in time-cost trade-off optimization. *Journal of Construction Engineering and Management*, 135(7), 668-674.
- [18] Günaydin, H. M., & Doğan, S. Z. (2004). A neural network approach for early cost estimation of structural systems of buildings. *International Journal of Project Management*, 22(7), 595-602.
- [19] Petruşeva, S., Zileska-Pancovska, V., & Zujo, V. (2013). Predicting construction project duration with support vector machine. *International Journal of research in Engineering and Technology*, 11(2), 12-24.
- [20] Jang, H. S. (2004). Genetic algorithm for construction space management. *KSCE Journal of Civil Engineering*, 8(4), 365-369.
- [21] Sorrentino, M. (2013). Genetic Algorithms for Construction Time-Cost-Quality Trade-Off: A Road Project Case Study. IN *BO-RICERCHE E PROGETTI PER IL TERRITORIO LA CITTA EL ARCHITETTURA*, 4(2), 163-176.
- [22] Kulkarni, P., Londhe, S., & Deo, M. (2017). Artificial neural networks for construction management: A review. *Soft Computing in Civil Engineering*, 1(2), 70-88.
- [23] Leung, A. W., Tam, C. M., & Liu, D. K. (2001). Comparative study of artificial neural networks and multiple regression analysis for predicting hoisting times of tower cranes. *Building and Environment*, 36(4), 457-467.

- [24] Petruseva, S., Zileska-Pancovska, V., & Zujo, V. (2013). Predicting construction project duration with support vector machine. *International Journal of research in Engineering and Technology*, 11(2), 12-24.
- [25] Magalhães-Mendes, J. (2015). Multiobjective Approach for Time-Cost Optimization Using a Multi-mode Hybrid Genetic Algorithm. In *Evolutionary Algorithms and Metaheuristics in Civil Engineering and Construction Management* (pp. 33-49). Springer, Cham.