



## Particle Swarm Optimization and Tree Models (M5P) as Cost Estimation Tool for Construction Project

Abbas Mahde Abd<sup>ID</sup>, Raquim Nihad Zehawi<sup>\*ID</sup>, Rouwaida Hussein Ali<sup>ID</sup>

Highway and Airport Engineering Department, College of Engineering, University of Diyala, Baquba 32001, Iraq

Corresponding Author Email: [raquim\\_zehawi@uodiyala.edu.iq](mailto:raquim_zehawi@uodiyala.edu.iq)

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.110903>

### ABSTRACT

**Received:** 10 August 2023

**Revised:** 17 December 2023

**Accepted:** 30 December 2023

**Available online:** 29 September 2024

#### Keywords:

*cost estimation, M5P, tree models, PSO, construction management*

A useful tool for non-linear multivariable modelling is the Artificial Neural Network (ANN). Cost-effectiveness has been demonstrated for the use of ANN. Typically, the neural network is trained using the back propagation (BP) algorithm. Even while this technique is particularly successful at training a wide variety of networks, it suffers from slow convergence and easy trapping in local minima. Regression tree models (M5P) were used to suggest the adjustment of neural networks utilizing particle swarm optimization (PSO) as a technique for predicting building costs due to the potential of multiscale modelling and prediction using the incorporated relevant parameters affecting the building cost. The three models were assessed using a variety of metrics, including Mean Square Error (MSE), absolute error, and Pearson's correlation as an accuracy criterion. The MSE was also employed as a parameter to determine the ideal number of neurons for the hidden layer. The results indicate that the PSO model outperforms the other two techniques. The study also concluded that the PSO model can accurately predict the costs of construction projects when considering specified project activities and price indices. A combination of project features and an automated modelling mechanism has the potential to provide a reliable prediction result.

## 1. INTRODUCTION

To prepare estimates of various construction activities and for various purposes, the cost estimating methods have been developed. Many standard and techniques for cost estimating are readily available as listed by Ali and Abd [1]. On the same approach, Akintoye and Fitzgerald [2] in their research stated that the standard estimating procedure was the main technique used by construction organizations, rather than the comparison to similar projects based on documented information, and personal experience. The Rapid technology changes in construction were contributed in processes' management complication of the traditional construction projects [3]. These changes supported by the information technology (IT) may be forced the change in engineering professions toward develops the management skills and support it with a specific tools of IT. This common method may be like the traditional estimation, in which the construction elements costs, including labor, materials, and sub-contractor are founded considering the addition of allowance for profit, overhead, and risk.

Sometimes, Methods of estimation that are conventional or traditional frequently fall short of the realities of the present day, which include aspects of uncertainty. Many researchers mentioned that construction estimation is tedious and time consuming activity. To enhance the estimation process quality, it is necessary to use an assembly pricing approach as pricing module, pricing system, rapid pricing or aggregate pricing [4-6].

It is complicated to adopt a standard or systematic process by which the estimated cost procedure for construction activities can be improved, due to several factors such as the great diversity in construction industry, types of projects, utilized methods, contractors, suppliers and others [7, 8].

Abd and Abd [9] investigated and developed an estimating process rather than the practice of cost estimating, by observing artificial intelligent techniques to simulate and optimize the priced selected items from bills of quantities and the variability associated with the outcomes over a specific time interval. uncertainty. Although a lot of study has been done on contractors' cost estimate procedures for building projects, relatively little of it has taken time effects into account when calculating costs for important construction tasks.

The analysis of existing cost estimate practices for school construction projects is reported in this paper, which also provides a foundation for future research into the creation of a cost estimating model or technique to enhance the forecasting of the productivity of specialized resources.

## 2. LITERATURE REVIEW

### 2.1 Educational building requirements

As stated by different standards, basic and secondary educational buildings projects must be constructed in single

compatible building unit and should not exceed three units consisting of classes and other facilities rooms. In term of school size or capacity, it must have at least six, nine, twelve eighteen, or twenty-four classrooms, obviously with multipurpose labs, and administration office. The classrooms represent the main construction part compared with other facilities in term of the construction area and construction cost, this fact was true for school building and thus, this factor will be considered the key one of construction estimation [10].

Traditionally, most of educational buildings include the three basic labs (biology, physics, and chemistry). All classrooms and educational spaces should follow the international standards in size and student area when considered for primary and final design phases. Some schools have an auditorium or stadium for general purposes. Different standards determine the requirements of the new educational facilities as stated below [10, 11]:

- Suitable entrance width of should not be less than 2.5m if the unique-side classrooms and 3.5m if it's situated from dual-side classroom building.
- Educational building needs to be characterized with simplified design for ease movement can be supervised the entrances and exits of student comfortably.
- Activity classrooms planned and designed for good view for all students.
- Accommodations for students should be provided such as toilets, shops, drinking water fountains, and others.
- Availability of parking area for cars and buses.
- Each student must have a minimum of 2.0 m<sup>2</sup> usable space as building area according to the international standards.

For the purpose of estimation, construction activities as (building area, concrete work, wall construction, and other) can be a reference for the three key elements in the school building requirement (classroom number, administrative area, and different facilities area) and will be considered in this work.

## 2.2 The definition of building cost

The definition of construction cost is diverse for stakeholders and depends on the scope of the intended cost. It may include direct, indirect costs, contractor profit, short-term or long-term maintenance, and others depending on the opinion of the party concerned with the construction project. The stakeholders in building construction projects; the client, engineer, contractor, end user and community are all together affected by the cost of building in many ways, due to their aims and expectations related to the building projects. Usually, in studies of cost control and cost estimation, the direct cost of each project is highlighted because it is typically relatively high compared to indirect cost throughout the construction projects [12].

For the purpose of investment decisions in building projects, the owners started with the evaluation of the contractor's bids, assess the tender price for each contractor, cost control during the project phases, in order to achieve an accurate estimation of the cost. The cost has long been considered as a scale to the performance of a building. Thus, for appraising building design, it is essential to utilize an appropriate cost estimation model [11, 12]. To use any cost model, the first step is to collect the required key data. Then, classify and analyze these data that must be updated. During the modelling process, quality of data and its level that convenience of the selected model is to be evaluated. When updatable data are acquired

during model application, they must be appended to the initial data set [9, 13, 14].

## 2.3 Computer-aided cost estimation systems

Continuous development in the technologies of information and communications have speed up the performance, accuracy and productivity of construction cost estimation process and made the following jobs very simple [13]:

1. Produce and organize an electronic (digital) bill-of-quantities (BoQ);
2. Establishing the computer-aided cost databases;
3. Setting up systems and software for computer-aided cost estimation.

Moreover, three categories can be used to classify computer-aided cost models for construction cost estimation:

Firstly, the computer aided systems used in initial stage of construction project to develop the design using a specific software as STAAD, ETAB or others, and for the process of construction documentation using MS-Project or Primavera or others.

The second utilize the parametric systems that investigates the relationship between the design variables and cost of building activities to model them numerically via statistical analysis given each variable some ratios or weight of effect. This can be found in the advance building information modelling BIM application using Revit or Archicad or other software.

The third technique is those considering the advantage of high developments in the artificial intelligence domain and knowledge based approach. There are many platforms to apply AI approaches such as machine learning, open source deep learning, or AI Collaborative Design Assistance.

Many approaches take the benefits of neural network, fuzzy logic, and revolutionary algorithms to simulate and model the construction cost estimation. For the purpose of this research, neural network will be considered to implement the cost estimation using the key factors in the building projects. The data sets collected from the BoQs of constructed projects, use to define the cost resource databases beside the reporting facilities to build up the cost estimation systems [14, 15].

## 2.4 Cost estimation under uncertainty

The Uncertainty can be defined as the lack of information needed to the decision that should to be taken at a specific time. Uncertainty occurs when the absence of full knowledge or difficulties in understanding a particular situation. It is associated with the risk and interchangeably used by many researchers. Even both of them related to the current knowledge used for future events, but risk unlike uncertainty, characterized with the possibility to be predicted and estimated the outcomes [16, 17].

Uncertainty affecting the estimation process of construction project cost depending on the stage of project life cycle. Both of the impacts (favourable or adverse) should to be considered in an uncertainty analysis to be comprehensive. Some of these impacts may be categorized as time and cost variability for actions that produced estimating uncertainty. The uncertainty sometime attributed to individual stochastic occurrences or the effect of many estimation elements in the same direction concurrently). Uncertainty analysis process needs to consider all these aspects and their interactions [17, 18].

### 3. METHODOLOGY AND RESEARCH FRAMEWORK

#### 3.1 M5P model tree algorithm

M5P modelling algorithm is the developed version for the M5 algorithm, it was primarily created in a way that allows model trees to solve problems efficiently even when they involve enormous data sets with numerous features and high dimensions [19].

The generation process of M5 algorithm starts by dividing the input data set into a number of subsets, each of which contains data records having similar characteristics. To reduce variation in a specific subset, linear regression modelling will be used in this step. Sequentially, each information generated within the previous step will be used in the creation of new several nodes at the current step using a given attribute for splitting process. This process of new steps generation will build a model tree structure at which, the leaves are at the bottom, with the root of the model at the top (refer to Yaseen et al. [19]). A mathematical logic (called regression tree) at the nodes, that compares a specific value of the basic data set with that of splitted value to find the path all the way to a leaf. The algorithm can generate knowledge from the tree using this technique.

In the M5P method is algorithm segments available data intervals into diverse smaller regions, called subspaces. Multivariate linear regression model is developed for each subspace. Both linear and tree regression are combined to divide data set into numerous categories using linear regression. This segregation technique that introduced for each node will improve final results. Even the M5P tree model sort of complicated instantly, it can be trimmed data down for simplicity with no functionality loses. The error at each node for the linear regression is computed starting from the bottom to the top of the tree, at any specific node, the sub-tree is trimmed until its error becomes less than the model held by the node itself. Following this transformation, the local Multivariate linear regression models in each individual subspace are restructured to enhance their predictive capabilities and overall accuracy for the targeted results [20, 21].

In this study, the collected data sets (data records for implemented 21 projects) are gathered by the researcher were used to develop a model for predicting the construction cost. these data records, were belong to the school's buildings projects. These projects were constructed within the interval of 2007 to 2013, no records were collected for the interval 2014 to 2018 because of the war and risky environment in Iraq.

#### 3.2 Artificial neural network and PSO

Artificial neural network (ANN) amongst others was the most common revolutionary techniques have been used in the construction project management. Many researchers introduce the implementation of ANN and its benefits for the civil engineering field. They support numerous ideas that advocate using this strategy as a productive way to address various difficulties in the building sector. ANN also was incorporated as an early estimation technique in construction of road and tunnel. ANN method was used for costs estimating for the seismic retrofit construction through introducing ANN as an assessment tool [22, 23].

In fact, particle swarm optimization (PSO) algorithm is adopted in this work according to the characteristics of this

technique in solving such problems. Moreover, characteristics of the case study encourage to adopt PSO for more investigation to obtain the suitable technique compared with traditional back propagation neural network. Sufficient amount of the data during the time interval of the study and the suitable inter-relationships among various principal parameters have been utilized to determine the precise outcomes of construction cost [24].

PSO is the common method to optimize the continuous nonlinear based problems. Its key approach is that basic solutions are moved through hyperspace to be accelerated towards the best or more optimum solutions. The implementation of this technique using simple codes of evolutionary programming and like genetic algorithms to reach the final solution. As in evolutionary computation systems, the fitness concept employed and suggested solution to the problem are termed particles, each one adjusts the flying path based on the solution experiences of both itself and its companions. These particles or sub-models keeps track of their coordinates in the space associated with its previous best fitness solution. Moreover, each particle considers the counterpart corresponding to the whole system best value acquired of any other particles in the population. Presentation of overall particles vectors are taken, so, most optimization problems are convenient for such variable presentations [23, 25, 26].

The adopted ANN and PSO approach provide a flexible and time-efficient prediction tool to optimize the estimation cost process. Particle swarm optimization (PSO) was proved as an effective optimization tool in problem solving with highly constrained nonlinear problems. This algorithm mainly based on the cooperative behaviour for species like bird swarm. Each particle position (data point) in problem space gives the potential solutions for it. These particles update their locations relative to their own best positions and the local position for the swarm to improve it at each generation reaching to the best solution. This technique suitable for cost estimation as it has many advantages compared to other techniques, such as; to start the iteration process this algorithm does not depend on the initial solution, the parameters may be as fewer number as required for adjustment, minimum computing time with high flexibility in combining process compared with other techniques. Because of the ease implementation of PSO with effective searching speed, this algorithm has been widely used to solve various engineering problems. These three revolutionary techniques have been used to investigate the best modelling technique to solve the estimation of construction project cost.

### 4. RESULTS AND DISCUSSION

#### 4.1 Data classification

Collected data from 21 construction projects was studied and analysed to investigate the most important element in the estimation of project cost. A number of 36 elements were investigated which include the quantities and price of seventeen construction activities beside the project total cost and the year of construction as listed in Table 1.

The develop model to estimate the project cost needs to be simple and practical for the end users with minimal number of effective activities, so the investigated activities that introduced in this work were classified in term of their

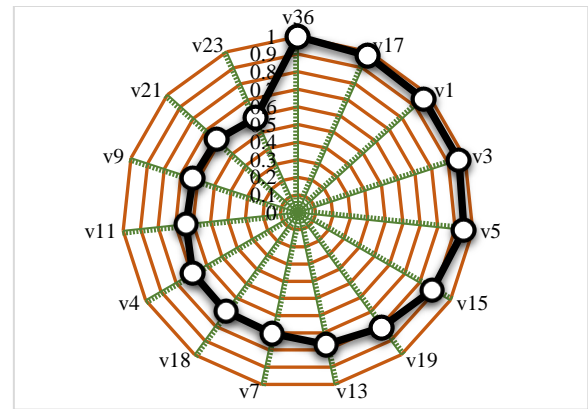
Pearson's correlation coefficient, a threshold of R-value greater than 0.6 was adopted by the authors to exclude the items of low correlation.

**Table 1.** Elements were investigated which include the quantities and price of seventeen construction activities

No.	Var. Code	Item	Corr. Coef.
1	v36	Project price	1
2	v17	Quantity of reinforced concrete for Slabs	0.976
3	v1	Quantity of excavation for foundation	0.963
4	v3	Quantity of reinf. concrete for foundation	0.963
5	v5	Quantity of wall (concrete blocks)	0.95
6	v15	Quantity of Beams and Columns (R. Conc.)	0.883
7	v19	Quantity of roofing work	0.809
8	v13	Quantity of reinf. concrete for cantilevers	0.768
9	v7	Quantity of normal concrete for DPC	0.705
10	v18	Price of reinforced concrete for Slabs	0.695
11	v4	Price of reinf. concrete for foundation	0.691
12	v11	Quantity of wall constructed with Clay bricks	0.64
13	v9	Quantity of Earth filling	0.628
14	v21	Quantity of finishing (gypsum)	0.62
15	v23	Quantity of painting work	0.589
16	v27	Quantity of roofing with tiles	0.586
17	v12	Price of wall constructed with Clay bricks	0.472
18	v20	Price of roofing work	0.449
19	v32	Price of metal windows	0.366
20	v28	Price of roofing with tiles	0.351
21	v6	Price of wall (concrete blocks)	0.34
22	v22	Price of finishing with gypsum	0.32
23	v34	Price of lean concrete	0.315
24	v10	Price of Earth filling	0.309
25	v2	Price of excavation for foundation	0.222
26	v8	Price of normal concrete for DPC	0.188
27	v33	Quantity of lean concrete	0.073
28	v35	Price of Earth filling year	0.027
29	v14	Price of reinf. concrete for cantilevers	-0.019
30	v25	Quantity of rendering work	-0.088
31	v26	Price of rendering work	-0.103
32	v31	Quantity of metal windows	-0.142
33	v29	Quantity of metal doors	-0.152
34	v24	Price of painting work	-0.274
35	v30	Price of metal doors	-0.394
36	v16	Price of Beams and Columns (R. Conc.)	-0.524

The person correlation coefficient (r) was used to determine the number of 14 construction activities were selected from a set of 36 elements depending as the variables of the highest correlation coefficient. Figure 1 shows these variables in radar

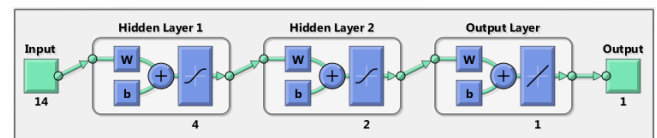
graph for ease presentation relative to their importance. Most of the highest rank was for the quantities of activities (12 item), and only two items in term of activity price. This reflects the big share for the work size compared with the price of activities. Amongst the highest fourteen elements the was two items in term of price and twelve items of quantities, the reason behind that concerned with the degree of fluctuation in activities price along the study period, there was little change in activities price during the time interval of the considered project. So, the main effects were belonged to the quantities more than price for the most activities under consideration.



**Figure 1.** Relative importance of variables

#### 4.2 Neural network modelling

The network used for this analysis was utilize back-propagation process with two layers (4, 2 neurons in each respectively) with sigmoidal transformation function for the hidden layers. Figure 2 shows the pattern configuration for the net used.



**Figure 2.** Pattern configuration

The use of back- propagation and sigmoidal transformation function for the hidden layers was due to its ability that allows the network to introduce non-linearity into the model, which allows the neural network to learn more complex decision boundaries. For the number of layers, it was proved that if the data set is less complex and having fewer dimensions or features then neural networks with 1 to 2 hidden layers would work. If data is having large dimensions or features then to get an optimum solution, 3 to 5 hidden layers can be use, the researchers Testing many scenarios for the networks, the one of two layers of 2 and 4 neurons were found of the best results. It is a matter of time, complexity and qualified accuracy, the simple, and fast the model the best for end user.

Some extreme value can occur due to many reasons, such as the unique situation of each construction project, the weather conditions, the administration, regulations and many others, all these parameters can affect the cost performance, so, the model difficult to follow up all these effects and thus we found some values may be far from the equality line.

Within the model, the weights are the real values that are attached with each input/feature and they convey the

importance of that corresponding feature in predicting the final output.

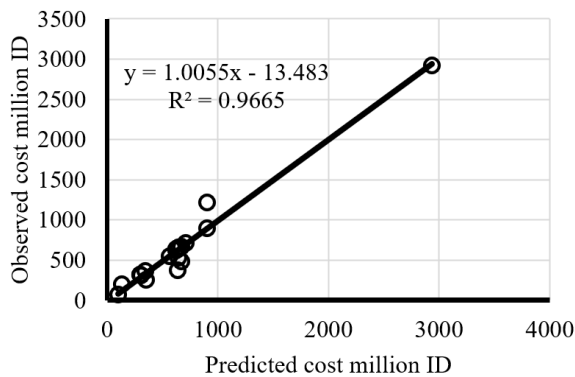
Bias: Bias is used for shifting the activation function towards left or right, you can compare this to y-intercept in the

line equation.

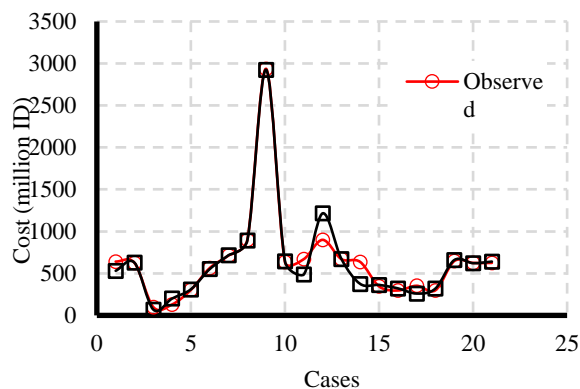
Table 2 shows the base and weights of each variable within the model for both hidden and output layers have been used for the purpose of this network.

**Table 2.** Base and weights of variables within ANN model

Var.	Input Layers						Output Layer
	Layer #1				Layer #2		
	Neur. #1	Neur. #2	Neur. #3	Neur. #4	Neur. #1	Neur. #2	
v0	-1.08	1.254	0.3994	-0.132	1.893	-3.232	-0.265
v1	1.05	-0.875	-0.072	1.897	-3.506	-0.622	1.38
v2	1.81	2.525	-0.695	0.193	-1.299	-3.613	0.118
v3	3.94	-1.163	3.441	-0.349	0.316	-0.034	
v4	-0.25	1.736	2.068	2.892	0.445	1.336	
v5	3.19	-0.362	1.01	1.902			
v6	-0.096	0.965	-0.996	3.853			
v7	2.46	-1.19	1.398	0.2794			
v8	-0.27	3.249	-3.35	1.387			
v9	2.44	0.181	3.369	-0.462			
v10	-1.299	-0.652	-0.389	0.997			
v11	1.437	-1.044	2.496	-0.011			
v12	-0.86	1.219	-1.741	0.502			
v13	1.55	-0.075	4.823	-0.962			
v14	-0.362	-0.099	0.251	1.183			

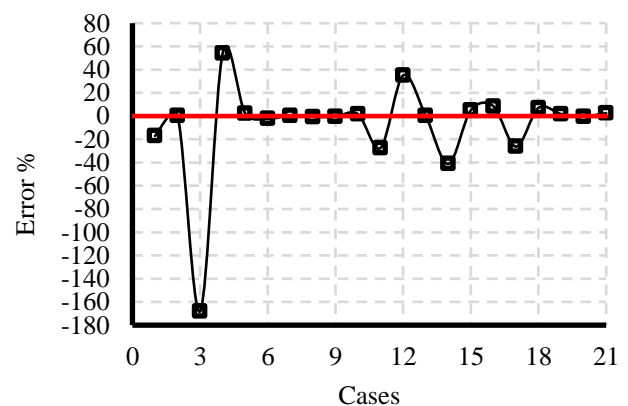


**Figure 3.** The observed cost vs. predicted cost using ANN



**Figure 4.** Comparison of observation and predicted cost using ANN

Figure 3 presents the correlation between the observed and predicted cost. It is reflecting the good prediction even some points far from the equality line, but at general it can be considered as acceptable prediction tool. Figure 4 shows the comparison between both values of observation and predicted cost, and Figure 5 presents the range of percent error for the model.



**Figure 5.** Errors percent for the predicted cost using ANN model

### 4.3 PSO and ANN modeling

In this section, the artificial neural network (ANN) has been trained by particle swarm optimization (PSO). The generalized ANN model can be trained for every input attribute with single target feature.

The ANNs have better generalizability, less susceptibility to noise and outliers than the regression models. The PSO considered as a global optimization technique which has advantages over gradient-based algorithms. The PSO may be easily applicable for differentiable functions and may not be relevant to highly complex problems. The ANN integrated with PSO will be advance, and empirical modelling technique. The ANN-PSO, can deal with datasets of missing points and with nonlinear multi-dimensional dependent and independent features. The ANN-PSO characterized with a high learning ability and potential information processing to be suitable for solving complex and nonlinear problems. The PSO aim to evolve, at the same time, the three principal components of an ANN: the set of synaptic weights, the connections or architecture, and the transfer functions for each neuron. Many

fitness functions were used to evaluate the fitness of different solutions and find the best. These functions are based on the mean square error (MSE) and the classification error (CER) and avoid overtraining and to reduce the number of connections in the ANN and thus, improve the model performance and accuracy.

The suggested training method has been evaluated on the dataset being considered. Figure 6 shows the pattern recognition of the network, which includes a single hidden layer with 5 neurons, and utilize the sigmoidal transformation

function. Table 3 presents the biases and weights for the hall network components.

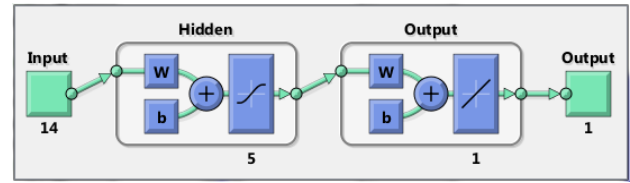


Figure 6. Pattern recognition of PSO-ANN network

Table 3. Biases and weights for PSO-ANN model

Var.	Input Layers (Single Layer)					Output Layer
	Neuron #1	Neuron #2	Neuron #3	Neuron #4	Neuron #5	
v0	-0.62182	-0.51928	0.390898	-0.0929	-0.00248	-0.62182
v1	1.5	0.649233	0.335427	-0.1634	0.634266	0.522557
v2	-1.5	0.713751	-0.16826	0.414538	-0.2492	1.29987
v3	-1.5	-0.01874	0.199267	0.37834	0.109283	0.158839
v4	-1.22736	0.367145	-0.40203	-0.38285	-0.33769	-0.67627
v5	0.477322	-0.54313	0.312497	-0.01552	0.381081	-1.5
v6	0.559833	-0.10794	-0.64589	-0.07507	-0.33787	
v7	-0.58237	0.649067	-0.34828	0.228431	0.0093	
v8	0.529799	0.374504	-0.58164	0.26833	0.255145	
v9	0.161832	0.561809	-0.49258	0.311399	0.477947	
v10	-0.61039	0.236202	0.490568	-0.33969	0.696578	
v11	-0.30559	-0.64054	0.268787	0.247919	0.065139	
v12	0.073194	0.545078	-0.28555	0.242147	-0.5642	
v13	0.586359	0.556223	0.577022	-0.43241	-0.44948	
v14	0.568407	0.218535	-0.56922	-0.46584	-0.29649	

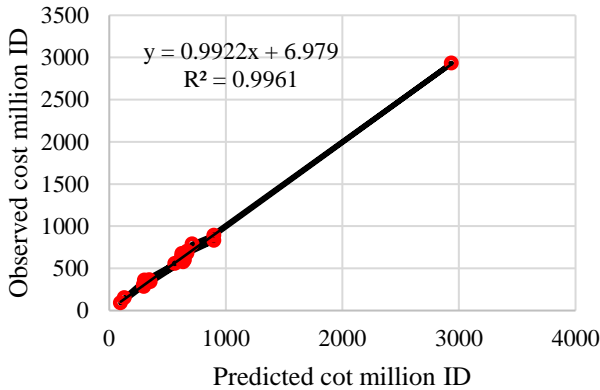


Figure 7. The observed cost vs. predicted cost using PSO-ANN model

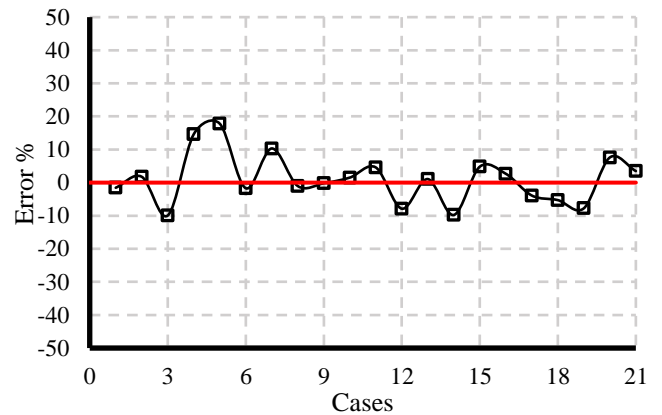


Figure 9. Errors percent for the predicted cost using PSO-ANN model

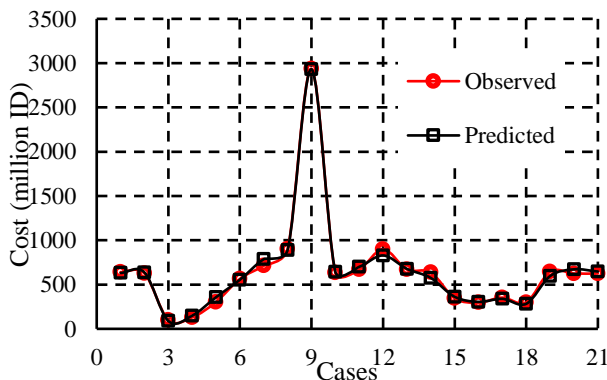


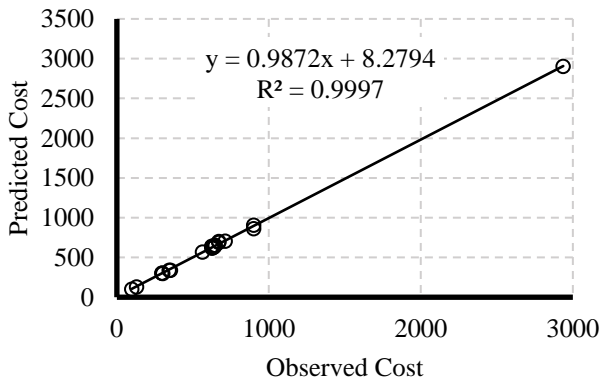
Figure 8. Comparison of observation and predicted cost using PSO-ANN model

This model characterized with high correlation coefficient of 99.8 % with smooth distribution of predicted data along the equality line as shown in Figure 7. Figure 8 shows the comparison between both values of observation and predicted cost, and Figure 9 presents the range of percent error for the model, it is reflecting an acceptable model with some fluctuation along the equality line. This occurs as a result of some missing data or reliability of collected data for the cases under consideration.

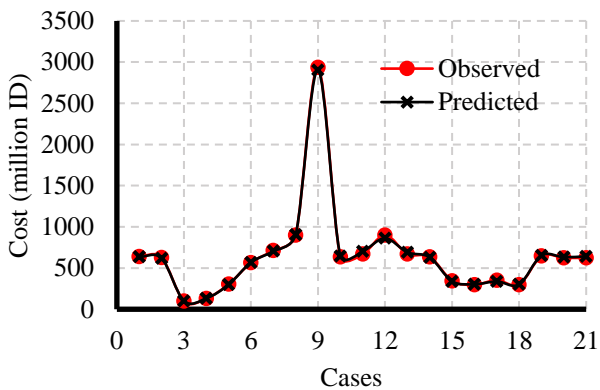
Overfitting can occur in ANN model due to random error in the data or the data set includes numerous highly correlated features, the relationships between variables can lead to this problem when the model is too complex (as in ANN models). In this work, the high correlated data ( $R > 0.6$ ) introduced to simplify the model and make it easy to implement by end users, so this high correlated model was expected.

#### 4.4 Cost estimation using M5P model

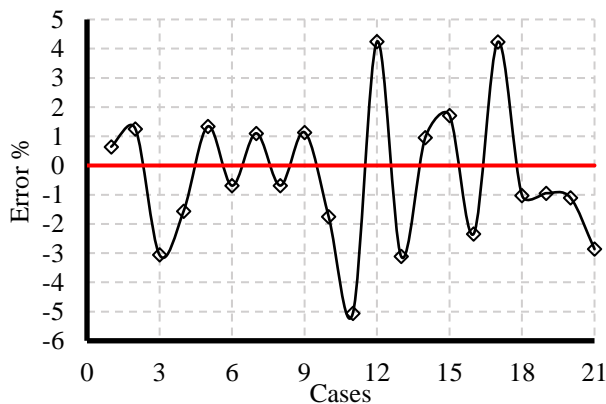
Both techniques, ANN and PSO-ANN were compared with the revolutionary method utilized the M5P modelling. This technique even it was very simple in modelling and application, but it was result in high accuracy and very good presentation for the cost estimation. The correlation coefficient was 99.97 % for the predicted results. The data was distributed uniformly along the equality line as shown in Figure 10. Also, in Figure 11, the predicted data were very consistent in its behavior compared with the observation. The percent error for the predicted cost was in narrow range (-0.05 to 0.04) as shown in Figure 12.



**Figure 10.** The observed cost vs. predicted cost using M5P model



**Figure 11.** Comparison of observation and predicted cost using M5P model



**Figure 12.** Errors percent for the predicted cost using M5P model

**Table 4.** Statistical comparison of the three models

Measures	ANN	PSO-ANN	M5P
MSE	1322.225	1202.292	250.549
MAE	64.301	15.005	11.636
r	0.983	0.998	0.9997

It is obvious that the results are more closely correlated with the line of best fit when the means squared error and absolute error are smaller. Depending on the dataset and the homogeneity of observations, it might be impossible to obtain a mean squared error and absolute error value that is very tiny. This is the reason makes the data scattered around the regression line, so the high correlation beside the lower range of errors necessary to select the best modelling approach. In this case, the third model (M5P model) have the best performance in modelling the cost prediction for the construction projects under consideration. Table 4 summarize the mean square error, absolute error and correlation coefficient for the three models.

M5P, a model-tree-based algorithm, presents a promising alternative as it combines the simplicity of decision trees with the power of linear regression, allowing for a greater interpretability without sacrificing predictive performance.

Limitations of this model: even the M5P model demonstrated a high level of accuracy, with a low correlation between input dependent features and model error. Overfitting problem may be occurring in case of highly correlated input variables, as in this research. This is an essential characteristic for a power predictive model, as it ensures that the predictions process requires the inclusion of more independent variables or minor factors to avoid the overfitting problem. Also, dataset used for the development and validation of the M5P model was limited in size and scope, which may affect the generalizability of the results and cause the potential overfitting.

The percent error for the predicted cost was in a narrow range (-0.05 to 0.04), this narrow range prove that the M5P model is more reliable and accurate than the other models.

#### 5. CONCLUSIONS

Developing a cost estimation method based on elements related to quantities and price indices. In this work; twelve factors related to quantities and two related price indices have been investigated. The factors related to the quantity of reinforced concrete for Slabs and excavation for foundation were of the most correlation to the project cost. Then, the reinforced concrete for foundation and the wall construction work followed in the rank of high correlation.

The second approach of this work deals with the automated techniques to model the prediction of construction cost. Considering the fourteen factors, three of revolutionary modelling techniques have been investigated. ANN proved as a good prediction tool in term of result precision but with significant rate of mean square error and absolute error. This mean there are over estimation for some factors or under estimation. Using PSO to improve the mechanism of neural network results in more precise result and minimize both errors (MSE and MAR). The third modelling technique, the regression tree modelling (M5P) has been proved that it has the best precise results of the highest correlation and the minimal errors.

These conclusions were developed by creating a database

containing data from earlier projects. Additionally, to give a meaningful benchmark for how precisely such models can predict the project cost, any future cost estimating models should take into account the database appropriate for the location of the projects and the surrounding environment. Utilizing scientific benchmarks may aid in categorizing and evaluating cost estimating proposal submissions.

This work can support the cost estimation using the most critical items of the of the construction project. Even in case of limited items included in this modelling technique, the construction firms can develop a specific model for projects of similar field exploiting the benefits of this modelling approach effectively. In spite of the developed model in this work utilize specific construction item, it is possible to expand the items and size (in term of long time interval and number of cases) of data set for more generalization and reliability.

In light of the concluded results, a further study can be achieved to raise the benefit of this work, such as the using of genetic algorithm to optimize the number of layers and neurons in ANN modelling for more precise and accurate results, also there is a possibility to study the utilize the M5P modelling in cost control and monitoring during the project procurement phase.

## REFERENCES

- [1] Ali, M.H., Abd, A.M. (2021). Extreme learning machines (ELM) as smart and successful tools in prediction cost and delay in construction projects management. *IOP Conference Series: Earth and Environmental Science*, 856(1): 012041. <https://doi.org/10.1088/17551315/856/1/012041>
- [2] Akintoye, A., Fitzgerald, E. (2000). A survey of current cost estimating practices in the UK. *Construction Management & Economics*, 18(2): 161-172. <https://doi.org/10.1080/014461900370799>
- [3] Omar, A.M., Omar, N.H. (2022). Factors affecting construction projects' cost estimating. *Fayoum University Journal of Engineering*, 5(2): 26-41, <https://doi.org/10.21608/fuje.2022.168316.1027>
- [4] Bode, J. (2000). Neural networks for cost estimation: Simulations and pilot application. *International Journal of Production Research*, 38(6): 1231-1254. <https://doi.org/10.1080/002075400188825>
- [5] Akintoye, A. (2000). Analysis of factors influencing project cost estimating practice. *Construction Management & Economics*, 18(1): 77-89. <https://doi.org/10.1080/014461900370979>
- [6] Oberlender, G.D., Trost, S.M. (2001). Predicting accuracy of early cost estimates based on estimate quality. *Journal of Construction Engineering and Management*, 127(3): 173-182. [https://doi.org/10.1061/\(ASCE\)0733-9364\(2001\)127:3\(173\)](https://doi.org/10.1061/(ASCE)0733-9364(2001)127:3(173))
- [7] Staub-French, S., Fischer, M., Kunz, J., Paulson, B. (2003). An ontology for relating features with activities to calculate costs. *Journal of Computing in Civil Engineering*, 17(4): 243-254. [https://doi.org/10.1061/\(ASCE\)0887-3801\(2003\)17:4\(243\)](https://doi.org/10.1061/(ASCE)0887-3801(2003)17:4(243))
- [8] Qian, L., Ben-Arieh, D. (2008). Parametric cost estimation based on activity-based costing: A case study for design and development of rotational parts. *International Journal of Production Economics*, 113(2): 805-818. <https://doi.org/10.1016/j.ijpe.2007.08.010>
- [9] Abd, A.M., Abd, S.M. (2012). Resources sustainability planning model using hierarchical approach for construction project. *Diyala Journal of Engineering Sciences*, 5(2): 1-19. <https://doi.org/10.24237/djes.2012.05201>
- [10] Ibrahim, N.M., Osman, M.M., Bachok, S., Mohamed, M.Z. (2016). Assessment on the condition of school facilities: Case study of the selected public schools in Gombak district. *Procedia-Social and Behavioral Sciences*, 222: 228-234. <https://doi.org/10.1016/j.sbspro.2016.05.151>
- [11] Nurhayati, K., Nurul S.M.Y., Abdullah, S.S.I., Mentaza, K.P.A., Azizah, I. (2018). Building performance evaluation techniques. *Advanced Science Letters*, 24(6): 4425-4428. <https://doi.org/10.1166/asl.2018.11618>
- [12] Love, P.E., Sing, C.P., Carey, B., Kim, J.T. (2015). Estimating construction contingency: Accommodating the potential for cost overruns in road construction projects. *Journal of Infrastructure Systems*, 21(2): 04014035. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000221](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000221)
- [13] Chrysafis, K.A., Papadopoulos, B.K. (2009). Cost-volume-profit analysis under uncertainty: A model with fuzzy estimators based on confidence intervals. *International Journal of Production Research*, 47(21): 5977-5999. <https://doi.org/10.1080/00207540802112660>
- [14] Zehawi, R.N., Hameed, A.H. (2022). The impact of urban growth pattern on local road network: A system dynamics study. *International Journal of Design & Nature and Ecodynamics*, 17(2): 221-231. <https://doi.org/10.18280/ijdne.170208>
- [15] Ghadbhan Abed, Y., Hasan, T.M., Zehawi, R.N. (2022). Machine learning algorithms for constructions cost prediction: A systematic review. *International Journal of Nonlinear Analysis and Applications*, 13(2): 2205-2218. <https://doi.org/10.22075/ijnaa.2022.27673.3684>
- [16] Beltramo, M.N. (1988). Beyond parametrics: The role of subjectivity in cost models. *Engineering Costs and Production Economics*, 14(2): 131-136. [https://doi.org/10.1016/0167-188X\(90\)90115-X](https://doi.org/10.1016/0167-188X(90)90115-X)
- [17] Feng, K.L., Wang, S., Lu, W.Z., Liu, C.Y., Wang, Y.W. (2022). Planning construction projects in deep uncertainty: A data-driven uncertainty analysis approach. *Journal of Construction Engineering and Management*, 148(8). [https://doi.org/10.1061/\(asce\)co.1943-7862.0002315](https://doi.org/10.1061/(asce)co.1943-7862.0002315)
- [18] Albayati, R.S., Zehawi, R.N. (2022). System dynamic model for simulating aviation demand: Baghdad international airport as a case study. *Mathematical Modelling of Engineering Problems*, 9(5): 1289-1297. <https://doi.org/10.18280/mmep.090517>
- [19] Yaseen, Z.M., Deo, R.C., Hilal, A., Abd, A.M., Bueno, L.C., Salcedo-Sanz, S., Nehdi, M.L. (2018). Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Advances in Engineering Software*, 115: 112-125. <https://doi.org/10.1016/j.advengsoft.2017.09.004>
- [20] Behnood, A., Behnood, V., Gharehveran, M.M., Alyamac, K.E. (2017). Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Construction and Building Materials*, 142: 199-207.



- <https://doi.org/10.1016/j.conbuildmat.2017.03.061>
- [21] Yang, L., Liu, S., Tsoka, S., Papageorgiou, L.G. (2017). A regression tree approach using mathematical programming. *Expert Systems with Applications*, 78: 347-357. <https://doi.org/10.1016/j.eswa.2017.02.013>
- [22] Petroutsatou, K., Georgopoulos, E., Lambropoulos, S., Pantouvakis, J.P. (2012). Early cost estimating of road tunnel construction using neural networks. *Journal of Construction Engineering and Management*, 138(6): 679-687. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000479](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000479)
- [23] Gopalakrishnan, K. (2010). Neural network-swarm intelligence hybrid nonlinear optimization algorithm for pavement moduli back-calculation. *Journal of Transportation Engineering*, 136(6): 528-536. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000128](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000128)
- [24] Kannan, S.K., Diwekar, U. (2024). An enhanced particle swarm optimization (PSO) algorithm employing quasi-random numbers. *Algorithms*, 17(5): 195. <https://doi.org/10.3390/a17050195>
- [25] Cavalieri, S., Maccarrone, P., Pinto, R. (2004). Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry. *International Journal of Production Economics*, 91(2): 165-177. <https://doi.org/10.1016/j.ijpe.2003.08.005>
- [26] Shahbazi, M., Heidari, M., Ahmadzadeh, M. (2024). Optimization of dynamic parameter design of Stewart platform with Particle Swarm Optimization (PSO) algorithm. *Advances in Mechanical Engineering*, 16(6): 1-16. <https://doi.org/10.1177/16878132241263940>

## NOMENCLATURE

ANN	Artificial Neural Network
BP	Back Propagation Algorithm
M5P	Regression Tree Model
PSO	Particle Swarm Optimization
MSE	Mean Square Error
BoQ	Bill-of-Quantities