

# Neural Network-Based Prediction Model for Sites' Overhead in Commercial Projects

Ali Hassan Zeinhom Hassan, Arab Academy for Science, Technology, and Maritime Transport, Egypt

Amira M. Idrees, Faculty of Computers and Information Technology, Future University in Egypt, Egypt\*

 <https://orcid.org/0000-0001-6387-642X>

Ahmed I. B. Elseddawy, Arab Academy for Science, Technology, and Maritime Transport, Egypt

## ABSTRACT

Construction companies need to improve the accuracy of their projects' budgeting to achieve the targeted profit. Site overheads are the expenses related to a project but are not allocated to a specific work package. The main objective of this research is to develop a neural network model for commercial projects to predict and estimate project site overhead costs as a percentage of the direct cost. The focal point of the research is focused on the main factors affecting site overhead costs for commercial projects in Egypt. These factors and their weights were identified by experts through the collected structured data. Cost data for 55 projects in the past seven years was collected with various conditions of company rank, direct cost, project duration, project location, contract type, and type of company ownership. The results have shown that the best model developed consists of six input neurons; two hidden layers with six and five neurons respectively, and one output layer representing the percentage of project site overhead. The model was tested on six projects with accuracy of 84%.

## KEYWORDS

Artificial Intelligence, Commercial Projects, Data Mining, Knowledge Discovery, Machine Learning, Neural Networks, Prediction Models, Projects Overhead

## 1. INTRODUCTION

The construction industry is one of the major sectors in the Egyptian economy, especially real estate & Commercial Buildings (Idrees, ElSeddawy, & Zeidan, 2019). The construction sector has a great impact on the Gross Domestic Product (GDP), it represents about 16.5% of the total Egyptian GDP. However, after the Egyptian revolution in 2011 and the floating exchange rate for the Egyptian pound in November 2016 followed by a high inflation rate, many sectors suffered due to the unstable economic situation accompanied by the political risks in Egypt (Idrees, El Seddawy, & EL Moaaz, 2019) (Khedr, Idrees, & El Seddawy, 2016). With all these factors into consideration, an accurate

DOI: 10.4018/IJeC.318143

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

estimate for the project cost is needed by building construction companies in Egypt, and controlling project costs has become more important and with greater impact than before (Idrees, Alsheref, & ElSeddawy, 2019). Cost is considered one of the main three challenges that face building construction companies, where the success of any project is measured by its completion within the allocated budget, through the planned baseline, and with the desired quality. So an inaccurate estimation can easily lead to a cost overrun of the project, which is reflected on the company's profit. Accordingly, an accurate cost estimate in the early stages is considered a critical stage in producing a project cost, which allows contractors to evaluate the feasibility of the project (Hastak, 2015).

The absence of structured and accurate methods that can assess site overhead costs for commercial projects in Egypt put construction companies at risk of an inaccurate estimate of bid package that may affect the profit margin of the company (Khedr, El Seddawy, & Idrees, 2014). Most of the building construction companies find no difficulty in estimating the direct cost of a project, the inaccuracy appears in the estimating of the overhead costs, producing a cost variance between estimated budget and actual cost either cost over-run or a cost-saving (Hassouna, Khedr, Idrees, & ElSeddawy, 2020).

The objective of this research is to develop an artificial intelligence model based on Artificial Neural Network (ANN) that can enhance the contractor's ability to estimate site overhead cost as a percentage of the project direct cost in the commercial construction market in Egypt. This would improve the companies' performance in predicting overheads for the upcoming projects and increase the company competitive advantage by improving the bid accuracy and also will lead to: Help to control factors affecting site expenditures, create an information system and historical data for projects to improve the predictability of site overheads in future projects and Decrease time and effort spent during the process of site overheads estimation.

The research methodology followed main steps which are more clarified and discussed in details in the following sections, these steps are: discussing the literature review on construction cost categories and identify the main factors influencing site overhead cost in the construction industry, Conducting a survey with industry experts to identify the factors influencing site overhead in the commercial construction industry in Egypt, apply data collection of actual project data according to the concluded factors, perform sensitivity analysis of the collected project data to study the bounding relationship between each factor and site overhead percentage, design and develop an artificial neural network model, test and validate the model, and finally develop a graphical user interface for the developed model to be easily used. The remaining of the research provide more elaboration for each step in the research methodology. The study outline could be stated as follows:

- Problem: The absence of structured and accurate methods that can assess site overhead costs for commercial projects in Egypt put construction companies at risk of an inaccurate estimate of bid package that may affect the profit margin of the company.
- Aim: enhance the contractor's ability to estimate site overhead cost
- Solution: Develop an artificial intelligence model based on Artificial Neural Network (ANN) that can enhance the contractor's ability to estimate site overhead cost as a percentage of the project direct cost in the commercial construction market in Egypt.
- Methodology: Neural Network model
- Benefit: More accurate estimation of the commercial projects' site overhead cost
- Outcome: Improved performance

## **2. PREVIOUS WORK**

The behavior of a project life cycle and its trends can be forecasted before the actual beginning of the project by using simulations and modeling. Using artificial intelligence like expert systems and neural networks has proven to help solve the prediction problems (Cheng, Tsai, & Sudjono, 2010). Artificial Neural Networks (ANNs) are data modeling methods that attempt to solve complicated issues

by formulating data relationships; although there is no simple equation that can map relationships between data variables. ANNs (in methods supposed to be similar to the performance of the human brain) are beneficial where conditions are difficult to identify and have been commonly employed in construction applications like estimating productivity; predicting corporate bankruptcy and measuring project final performance. Indeed, several previous studies have sought to use neural networks to address construction estimation with varying degrees of success.

A study was conducted in 2011 discussing the application of neural networks for a parametric cost estimate of construction overhead costs in Egypt. The study used 52 projects that were executed through the period from 2002 to 2009 considering 10 factors affecting overhead costs. The developed network consisted of 10 input neurons and a single hidden layer with 13 hidden neurons and a sigmoid transfer function. This model has an accuracy of 80%. (ElSawy, Hosny, & AbdElRazik, 2011). In Egypt, Georgy and Barsoum developed an ANN model for the parametric cost estimate of school construction projects. They used statistical and neural network models for estimating costs; the research found was a neural network of a single hidden layer with a number of neurons equal to two-thirds of the number of neurons in the input layer would be sufficient. (Georgy M., 2005). In the Philippines, Lyne and Maximino developed an AAN to estimate the total structural cost of building construction projects in the Philippines based on six parameters which are: number of floors, number of basements, area of floors, concrete volume, area of formwork, and weight of reinforcement steel. The data used for the learning process consists of data from 30 different building construction projects; this data set was divided into 60%, 20%, and 20% for training, validation, and testing respectively. The resulted model architecture was a network of six input neurons (one for each input parameter), one hidden layer with seven neurons, and one output neuron representing the structural cost. The model was developed using MatLab with a feed-forward backpropagation technique (Roxas, Lyne, & Maximino, 2014). Other studies have been conducted (see summary in table 1).

More recent research as in (Mostofi, To˘gan, Ay˘zen, & Tokdemir, 2022) which compared a set of machine learning algorithms in predicting the impact of construction rework cost targeting to propose the highest performance algorithm. The research focused on construction field attributes which are design and construction related. Another research with both the same target and field was presented in (Car-Pusic, Petrusseva, Pancovska, & Zafirovski, 2020) using neural networks for construction cost prediction. The research confirmed the time saving contribution and higher accuracy by applying neural networks. Neural network has also been applied to identify the affecting factors on the project performance in (elhegazy, badra, aboul haggag, & abdel rashid, 2022). The research identified that estimating team and schedule were two of the most effective parameters. Moreover, neural network proved its effectiveness in production cost prediction with high performance in (Wei, 2022). The previous studies confirms the positive contribution of applying neural networks in the cost prediction task in different fields, however, there are lack of research considering the commercial projects which have their specific attributes. Therefore, the current study shall cover the estimating process of site overheads for commercial projects field and in Egypt community in specific aiming to focus on its affecting attributes. As mentioned in (Chen, Yu, Yang, & Shao, 2022), affecting attributes strongly varies according to the applied field as well as targeted community. The previous studies discussing the estimating process of overheads had training data for various project types and taking into consideration many factors that might lack accuracy and the focus on certain project types. The accuracy of a neural network is based on the accuracy and amount of the training data. So this research will focus on the commercial projects only that would increase its accuracy dealing with this type of projects as the collected project data will be for commercial projects only.

### **3. DATA COLLECTION AND ANALYSIS**

Collecting the essential project data for some commercial projects, which were executed by the first four categories of contractors in Egypt in the past 7 years is accomplished. This data is required for

Table 1. Site overhead factors concluded from literature review

Reference Factor	Project Location	Project Size	Project Duration	Project Complexity	Payment Schedule	Contract Type	Tendering Method	Class of Company	Cash Availability	Firm need of work	Assign work to sub-cont.	Inflation Rate	Number of Competitors
(S. A, & M, Project overhead costs in Saudi Arabia, 1999)	●	●		●	●	●			●	●			●
(C & C, 2002)	●	●		●	●	●				●	●		
(S. A, & M, The management of construction of company overhead costs, 2001)	●	●		●	●	●				●	●		●
(A & A, 2008)	●			●	●			●	●	●	●		●
(L.C., 2010)	●	●	●			●							
(C.T., 2012)		●	●	●		●	●			●		●	
(EISawy, Hosny, & AbdElRazik, 2011)	●	●	●	●		●		●					

the training of the neural network model for estimating the site overhead percentage. The results in table 2 from the previous studies conducted on this specific area reflects the main factors influencing site overhead costs for building construction in general. These findings will support the survey to find the factors concerning commercial construction projects in Egypt (see table 2).

From the analysis of the collected data, it was concluded that there is a difference between the factors studied in the literature review and those factors that affect the site overhead cost in commercial projects in Egypt. Limiting the research study to commercial projects narrows the factors that may affect the site overhead percentage for a project; moreover, some factors are not taken into consideration in the Egyptian construction industry especially in the commercial construction projects. So after the analysis of the data collected from experts and comparing it with the previously studied factors in the literature review, a final list was concluded that represents the factors that contribute to the commercial construction projects in Egypt (see figure 1).

The most stated factor was the ownership type of the company. That is whether the contracting company is publically or privately owned.

The ownership type of the company should be considered in the Egyptian construction industry as there are construction firms that are owned by the public sector, and the management style of these companies is different from those of private ownership (see table 2).

#### 4. CASE STUDY DATA

Data for 55 projects were collected according to the listed factors in the previous section; these data are for commercial projects that were executed in the past 7 years in Egypt. The collected project data was scanned and studied to examine the influence of each factor on the total site overhead percentage. This study will provide a comparative analysis for each factor and how it may influence the site overhead percentage. Also, it will help in identifying the critical factor

Figure 1. Factors affecting site overhead percentage

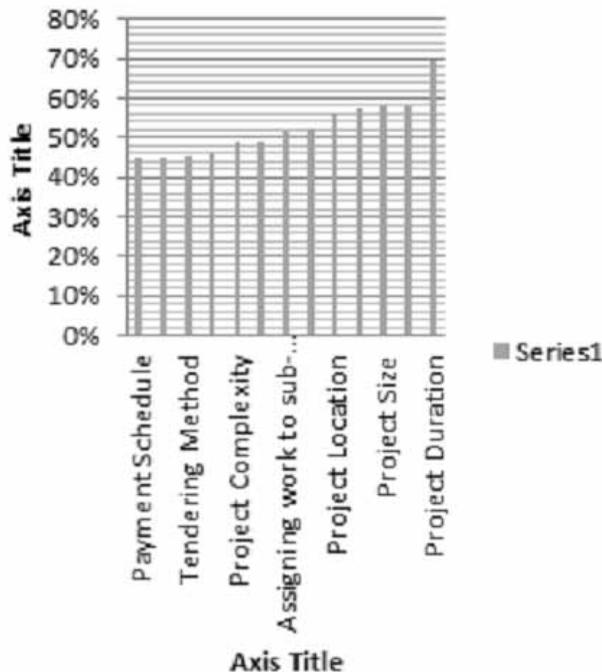


Table 2. Factors affecting commercial site overheads in Egypt

Sr.	Factor
1	Project Location
2	Project Direct Cost
3	Project Duration
4	Contract Type
5	Class of contracting company
6	Ownership type of contracting company

#### 4.1 The Influence of Class of Contracting Company

The analysis of the collected project data according to the class of the contracting company classified the results into four categories. These are the first four categories of contracting companies as enrolled in “The Egyptian Federation for Construction and Building Contractors”. Table 3 represents the statistical analysis of the collected project data according to the rank of the contracting company. Moreover, the analysis of the collected project data according to the class of the contracting company classified the results into four categories (see figure 2).

#### 4.2. The Influence of Direct Cost

The analysis of the collected project data according to the direct cost of the project classified the results into five categories (see figure 3).

Table 3. Projects data statistical analysis according to class of contracting company

Rank of Company	Number of Projects	Average Site Overhead %	Minimum%	Maximum%	Range
Rank 1	27	9.4%	7.98%	10.93%	2.95%
Rank 2	8	9.8%	9.16%	12.20%	3.04%
Rank 3	7	9.3%	8.13%	10.17%	2.04%
Rank 4	13	8.4%	7.42%	10.51%	3.09%

Figure 2. Project data analysis according to class of contracting company

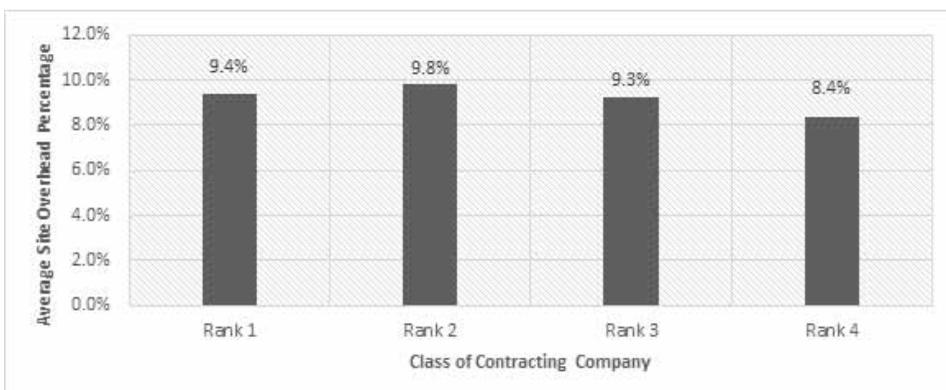
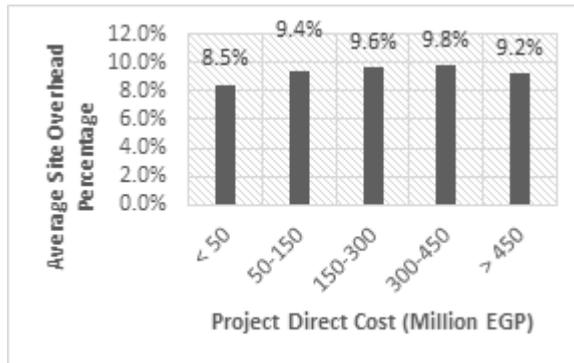


Figure 3. Project data analysis according to project direct cost



### 4.3. The Influence of Project Duration

The analysis of the collected project data according to the duration of the project classified the results into four categories (see figure 4)

The 55 collected projects data were categorized into 10 projects with duration of less than 18 months, 21 projects with duration between 18 and 36 months, 15 projects with duration between 36 and 60 months, and 9 projects with duration of more than 60 months.

### 4.4. The Influence of Project Location

The analysis of the collected project data according to the location of the project classified the results into two categories, inside the capital city and outside capital city (see figure 5).

### 4.5. The Influence of Contract Type

The analysis of the collected project data according to the contract type of the project classified the results into three categories, Cost-plus Contracts, Unit rate Contracts and Lump-sum Contracts (see figure 6)

Figure 4. Project Data Analysis according to Project Duration

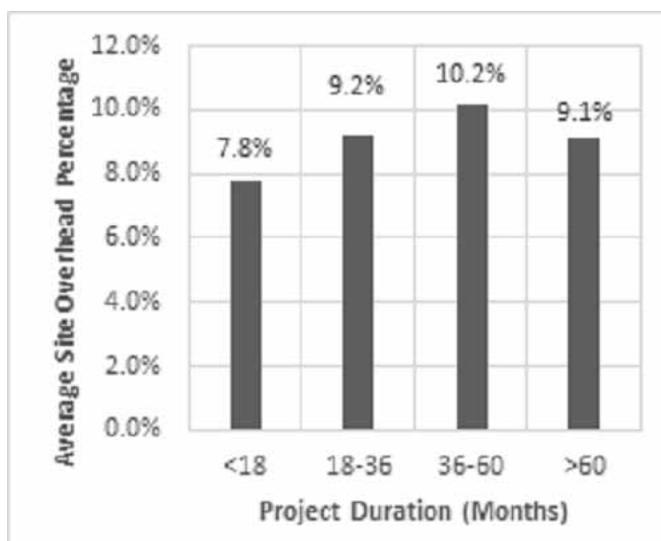


Figure 5. Project data analysis according to project location

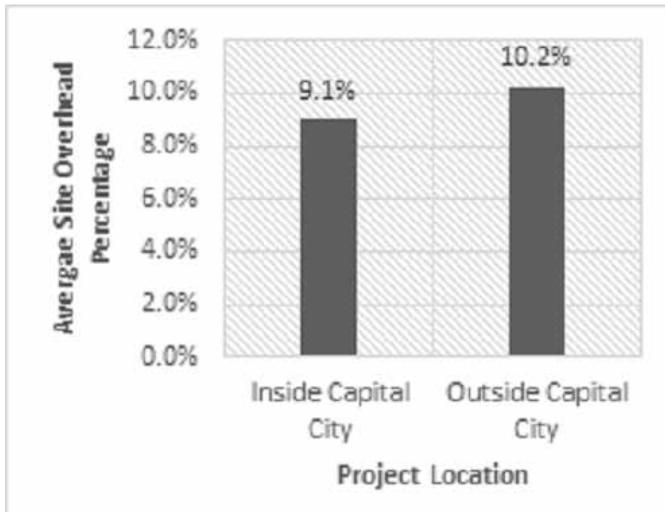
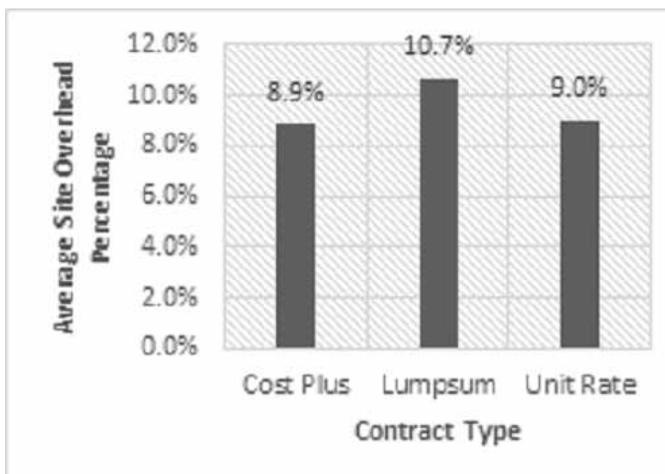


Figure 6. Project data analysis according to contract type



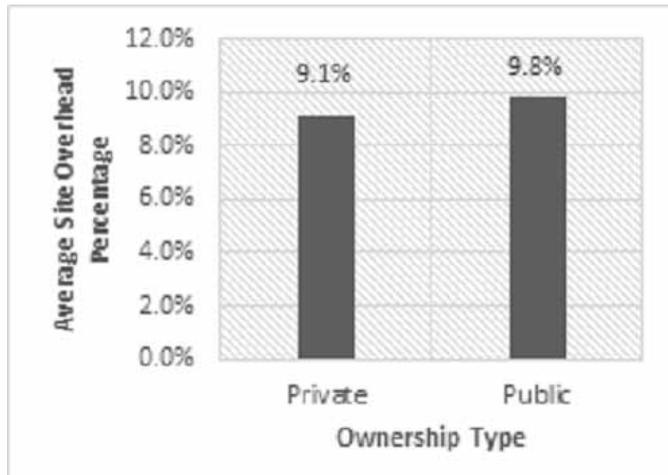
#### 4.6. The Influence of Ownership Type of the Contracting Company

The analysis of the collected project data according to the ownership type of the contracting company classified the results into two categories, Private Sector and Public Sector (see figure 7)

### 5. MAJOR FACTORS ANALYSIS

At this phase of the data collection and analysis, it is important to analyze the major factors collected from the previous sections. This process will enhance the study of the factors affecting site overhead percentage for commercial projects in Egypt. Sensitivity for each factor will be studied to show the effect of changing any variable or factor on the resulted site overhead percentage, to conclude how sensitive the site overhead percentage is to the change of any factor influencing it.

Figure 7. Project data analysis according to ownership type



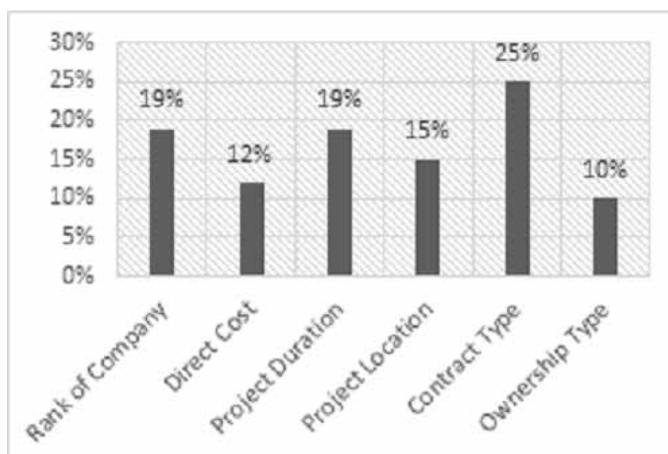
After the analysis of the data collected from experts and comparing it with the previously concluded percentages in the first survey, final weights for the governing factors influencing the site overhead percentage were concluded as shown in Figure 8.

From the study of the results of the survey in Figure 8, it is clear that the contract type is the most significant and critical factor that most influences the site overhead percentage. This is due to the critical role that the type of contract plays in determining the budget and tendering method and pricing including the possible allocated risk on each party for each type of contract. On the other hand, the ownership type of contracting company has the lowest influence on the site overhead percentage.

## 6. DESIGN OF THE NEURAL NETWORK

The main objective of this research is to develop an artificial neural network to predict the site overhead cost as a percentage of the total direct cost for commercial projects in Egypt. This can enhance the accuracy of estimators and decision-makers in proposing an estimate for the bid applying for.

Figure 8. Weights for each major factor



## 6.1. Steps for Designing the Neural Network

The process of designing the neural network for estimating the site overhead percentage for commercial projects in Egypt used in this research went through 8 major steps:

1. Problem definition
2. Data collection
3. Software selection
4. Data set preparation
5. Network architecture design
6. Network strategy and training setting
7. Network evaluation and selection
8. Network testing and validation

These steps of designing the network follow an iterative method of trial and error as illustrated in figure 9 to achieve the most accurate and valid network for estimating the site overhead percentage for commercial projects in Egypt.

The simple user-friendly interface of the “Neural Designer” guides the steps needed to develop the neural network model. These steps are as follows:

The “Neural Designer” automatically divides the data set to three types of instances:

- 60% training instances
- 20% selection (validation) instances
- 20% testing instances

This classification reflecting on 55 project data sets:

- 33 training instances
- 11 selection (validation) instances
- 11 testing instances

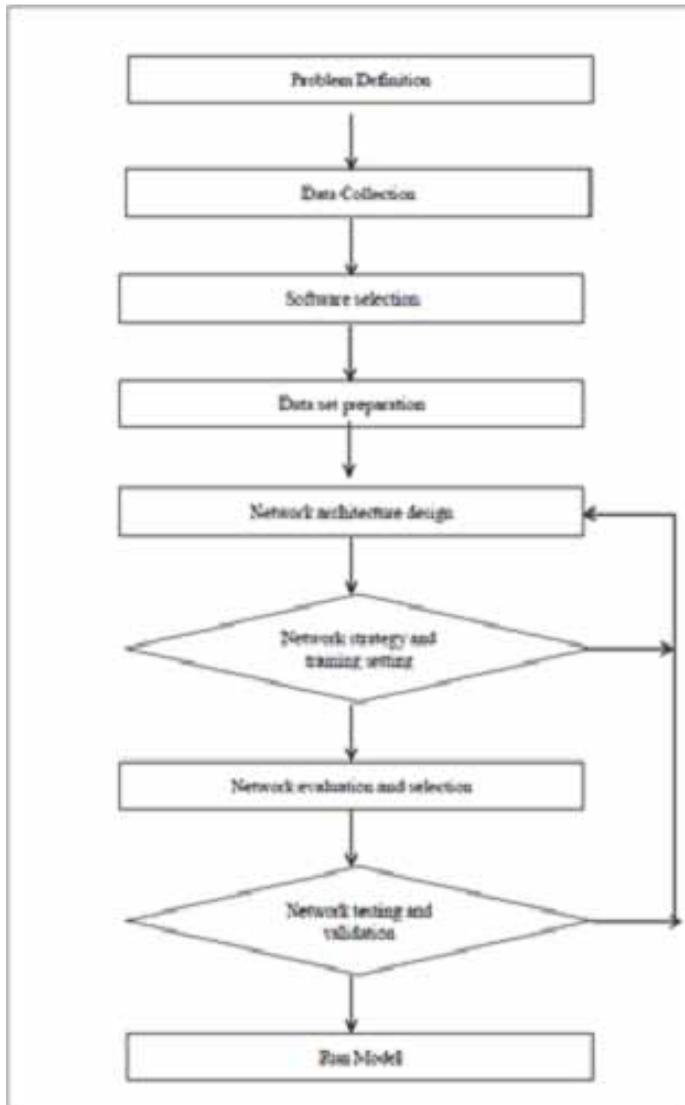
## 6.2. Training Strategy

In this research Gradient descent training strategy is chosen. The back-propagation works on the gradual error reduction between inputs and the target output. It develops the inputs to output leveling by minimizing the root mean squared error and the training process shall be finished when the RMSE becomes constant. The root mean square error is a good overall measure of whether a training run was successful. In developing the neural network in this research, the loss index is set to be the Root Mean Squared Error and the training algorithm is set to be Gradient descent.

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (x_i - E(i))^2}}{n}$$

- Where n is the number of projects to be evaluated in the training phase
- Xi is the model output related to the sample
- E is the target output

Figure 9. Steps for designing the neural network



### 6.2.1 First Package – Models with One Hidden Layer and Hyperbolic Tangent Transfer Function

The first run package of models resulted in 10 models with one hidden layer and hyperbolic tangent transfer function (see figure 10 and 11). Models vary in the number of neurons in the hidden layer starting from 3 neurons to 12 neurons. Table 4 shows the results of the first run package of models.

### 6.2.2 Second Package – Models with One Hidden Layer and Logistic Transfer Function

The second run package of models resulted in 10 models with one hidden layer and logistic transfer function (see figure 12). Models vary in the number of neurons in the hidden layer starting from 3 neurons to 12 neurons. Table 5 shows the results of the second run package of models.

From table 5, model 15 has the lowest RMSE of 0.463 corresponding to a mean percentage relative error of 7.04%, while model 17 has the highest RMSE of 0.657 corresponding to a mean

Table 4. First run package of models

Model	Hidden Layers	No. of Neurons in 1st Layer	No. of Neurons in 2nd Layer	Transfer Function	RMSE	Mean Absolute Error	Mean % Relative Error
Model 01	1	3	0	Hyperbolic Tangent	0.47	0.37	7.70%
Model 02	1	4	0	Hyperbolic Tangent	0.44	0.36	7.43%
Model 03	1	5	0	Hyperbolic Tangent	0.61	0.52	10.84%
Model 04	1	6	0	Hyperbolic Tangent	0.85	0.71	14.93%
Model 05	1	7	0	Hyperbolic Tangent	0.46	0.39	8.25%
Model 06	1	8	0	Hyperbolic Tangent	0.27	0.23	4.86%
Model 07	1	9	0	Hyperbolic Tangent	0.51	0.40	8.42%
Model 08	1	10	0	Hyperbolic Tangent	0.51	0.41	8.56%
Model 09	1	11	0	Hyperbolic Tangent	0.47	0.36	7.50%
Model 10	1	12	0	Hyperbolic Tangent	0.44	0.35	7.26%

Figure 10. Results of first run package of models

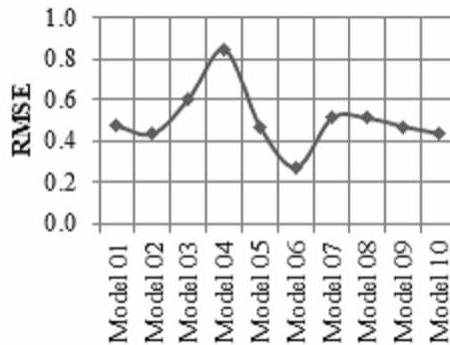
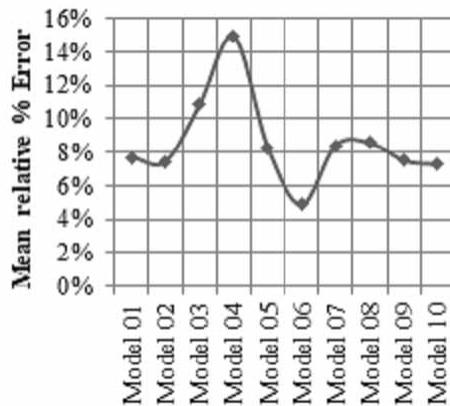


Figure 11. Results of first run package of models



**Table 5. Second run package of models**

Model	Hidden Layers	No. of Neurons in 1st Layer	No. of Neurons in 2nd Layer	Transfer Function	RMSE	Mean Absolute Error	Mean % Relative Error
Model 11	1	3	0	Logistic	0.47	0.36	7.62%
Model 12	1	4	0	Logistic	0.54	0.36	7.59%
Model 13	1	5	0	Logistic	0.47	0.37	7.71%
Model 14	1	6	0	Logistic	0.56	0.41	8.54%
Model 15	1	7	0	Logistic	0.46	0.34	7.04%
Model 16	1	8	0	Logistic	0.51	0.38	7.96%
Model 17	1	9	0	Logistic	0.66	0.50	10.38%
Model 18	1	10	0	Logistic	0.52	0.36	7.47%
Model 19	1	11	0	Logistic	0.56	0.39	8.15%
Model 20	1	12	0	Logistic	0.49	0.37	7.68%

percentage relative error of 10.38%. The results of the second run package of models show that the RMSE changed in a non-linear trend as the number of neurons increases as shown in Figure 12.

### 6.2.3 Third Package – Models with Two Hidden Layer and Hyperbolic Tangent Transfer Function

The third run package of models resulted in 36 models with two hidden layers and hyperbolic tangent transfer function (see figure 13). Models vary in the number of neurons in each hidden layer starting from 3 neurons to 10 neurons. Table 6 shows the results of the third run package of models.

From table 6, model 54 has the lowest RMSE of 0.415 corresponding to a mean percentage relative error of 6.88%, while model 42 has the highest RMSE of 1.068 corresponding to a mean percentage relative error of 14.84%. The results of the third run package of models show that the RMSE changed in a non-linear trend as the number of neurons increases as shown in Figure 13.

### 6.2.4 Fourth Package – Models with Two Hidden Layer and Logistic Transfer Function

The fourth run package of models resulted in 36 models with two hidden layers and logistic transfer function; models vary in the number of neurons in each hidden layer starting from 3 neurons to 10 neurons. Table 7 shows the results of the fourth run package of models.

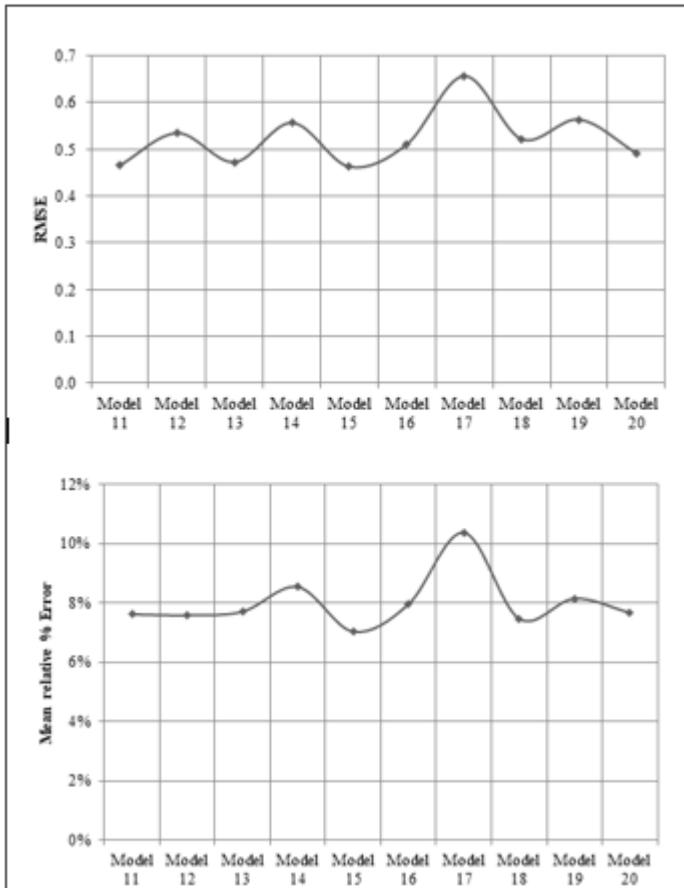
From table 7, model 65 has the lowest RMSE of 0.188 corresponding to a mean percentage relative error of 3.50%, while model 78 has the highest RMSE of 0.647 corresponding to a mean percentage relative error of 9.21%. The results of the fourth run package of models show that the RMSE changed in a non-linear trend as the number of neurons increases as shown in Figure 14.

### 6.2.5. Selecting the Best Model

The run of the trial and error process as discussed before resulted in 92 trial models categorized into four packages that vary in the number of hidden layers, number of neurons in each layer, and the transfer function. The next Figure 15 illustrates the linear regression for the scaled output OHP. The predicted values are plotted versus the actual ones as squares. The solid line indicates the best linear fit. The grey line would indicate a perfect fit. By reviewing the 92 models, it is concluded that model 65 is the optimum choice with the following design considerations stated in table 8 and network architecture as shown in Figure 16.

As shown above, Model 65 designed with the following considerations:

Figure 12. Results of second run package of models



- 6 input neurons (one neuron for each factor choice)
- 2 hidden layers
- 6 neurons in the first hidden layer
- 5 neurons in the second hidden layer
- Logistic transfer function for both layers
- 1 output neuron represents the site overhead percentage (OHP)

## 7. VALIDATION AND TESTING OF THE MODEL

To evaluate the validity and reliability of the developed model, the model has to be tested by predicting the site overhead percentage of new projects not previously introduced to the training of the model.

The prediction accuracy is evaluated by comparing the predicted value with the real-life or actual value and calculates the difference between the two values. The relative error is calculated and evaluated according to the acceptable relative error of the model resulted in the developing process.

### 7.1. Validation of the Model

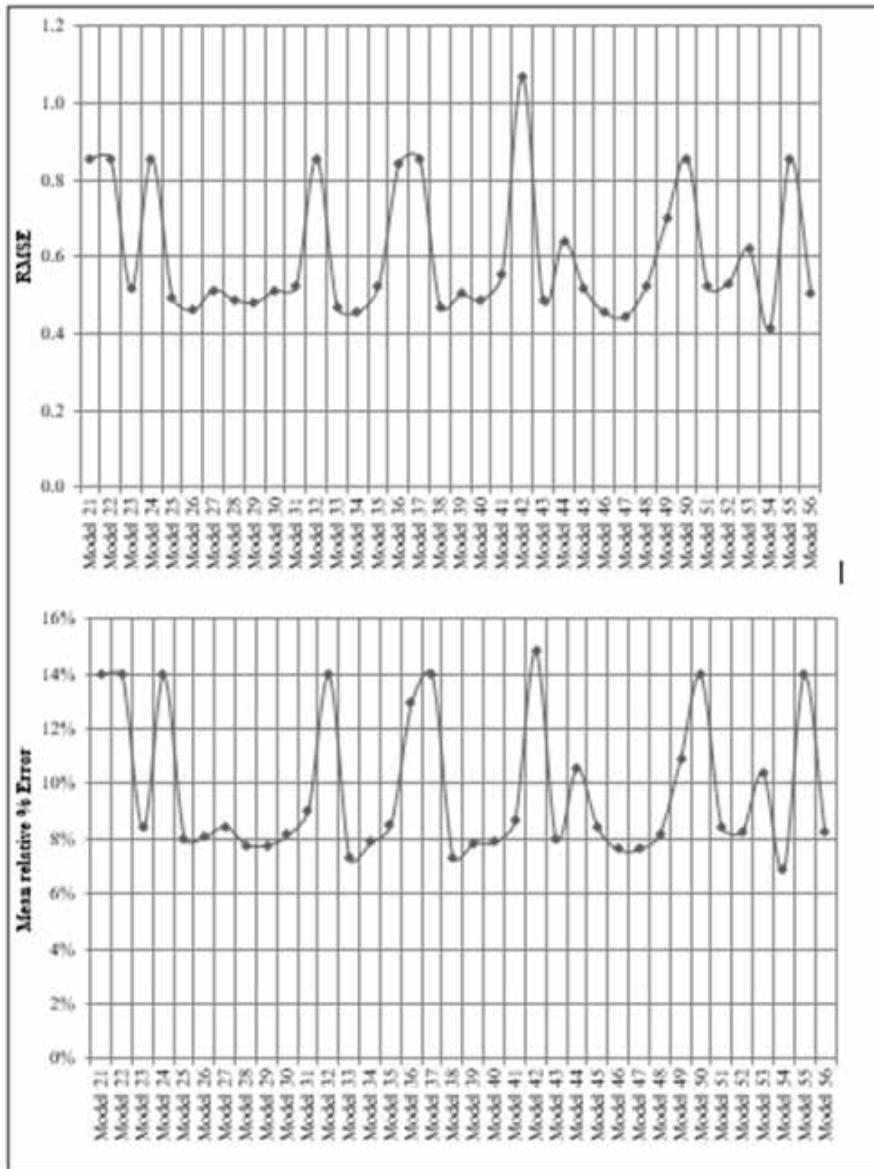
For the validation process, the model was tested on the base cases previously introduced in the learning process of the model and compared to the acceptable relative percentage error of the model. The

Table 6. Third run package of models

Model	Hidden Layers	No. of Neurons in 1st Layer	No. of Neurons in 2nd Layer	Transfer Function	RMSE	Mean Absolute Error	Mean % Relative Error
Model 21	2	3	3	Hyperbolic Tangent	0.86	0.67	13.97%
Model 22	2	4	3	Hyperbolic Tangent	0.86	0.67	13.97%
Model 23	2	4	4	Hyperbolic Tangent	0.52	0.40	8.39%
Model 24	2	5	3	Hyperbolic Tangent	0.86	0.67	13.97%
Model 25	2	5	4	Hyperbolic Tangent	0.50	0.38	8.04%
Model 26	2	5	5	Hyperbolic Tangent	0.46	0.39	8.09%
Model 27	2	6	3	Hyperbolic Tangent	0.51	0.40	8.46%
Model 28	2	6	4	Hyperbolic Tangent	0.49	0.37	7.81%
Model 29	2	6	5	Hyperbolic Tangent	0.48	0.37	7.78%
Model 30	2	6	6	Hyperbolic Tangent	0.51	0.39	8.14%
Model 31	2	7	3	Hyperbolic Tangent	0.52	0.43	9.08%
Model 32	2	7	4	Hyperbolic Tangent	0.86	0.67	13.97%
Model 33	2	7	5	Hyperbolic Tangent	0.47	0.35	7.29%
Model 34	2	7	6	Hyperbolic Tangent	0.46	0.38	7.87%
Model 35	2	7	7	Hyperbolic Tangent	0.52	0.41	8.53%
Model 36	2	8	3	Hyperbolic Tangent	0.84	0.62	12.96%
Model 37	2	8	4	Hyperbolic Tangent	0.86	0.67	13.97%
Model 38	2	8	5	Hyperbolic Tangent	0.47	0.35	7.35%
Model 39	2	8	6	Hyperbolic Tangent	0.50	0.38	7.86%
Model 40	2	8	7	Hyperbolic Tangent	0.49	0.38	7.87%
Model 41	2	8	8	Hyperbolic Tangent	0.55	0.42	8.69%
Model 42	2	9	3	Hyperbolic Tangent	1.07	0.71	14.84%
Model 43	2	9	4	Hyperbolic Tangent	0.49	0.38	8.02%
Model 44	2	9	5	Hyperbolic Tangent	0.64	0.51	10.60%
Model 45	2	9	6	Hyperbolic Tangent	0.52	0.40	8.44%
Model 46	2	9	7	Hyperbolic Tangent	0.46	0.36	7.62%
Model 47	2	9	8	Hyperbolic Tangent	0.45	0.36	7.62%
Model 48	2	9	9	Hyperbolic Tangent	0.52	0.39	8.18%
Model 49	2	10	3	Hyperbolic Tangent	0.70	0.52	10.87%
Model 50	2	10	4	Hyperbolic Tangent	0.86	0.67	13.96%
Model 51	2	10	5	Hyperbolic Tangent	0.52	0.40	8.46%
Model 52	2	10	6	Hyperbolic Tangent	0.53	0.40	8.27%
Model 53	2	10	7	Hyperbolic Tangent	0.62	0.50	10.44%
Model 54	2	10	8	Hyperbolic Tangent	0.42	0.33	6.88%
Model 55	2	10	9	Hyperbolic Tangent	0.86	0.69	13.98%
Model 56	2	10	10	Hyperbolic Tangent	0.51	0.40	8.28%

model was validated by running the model on the 38 base cases resulted from the 55 training project data and the model prediction was compared to the real-life value as shown in Figure 17, and then

Figure 13. Results of third run package of models



relative percentage error was calculated. The relative percentage error should be within the acceptable range of the model to be accepted, if not, the result is considered not accepted.

Through the analysis of the validation process, it was concluded that the model predicted 33 base cases within the acceptable range of the model, and 5 base cases out of the acceptable range. These results represent 90% accuracy for the model selected.

### 7.2. Testing of the Model

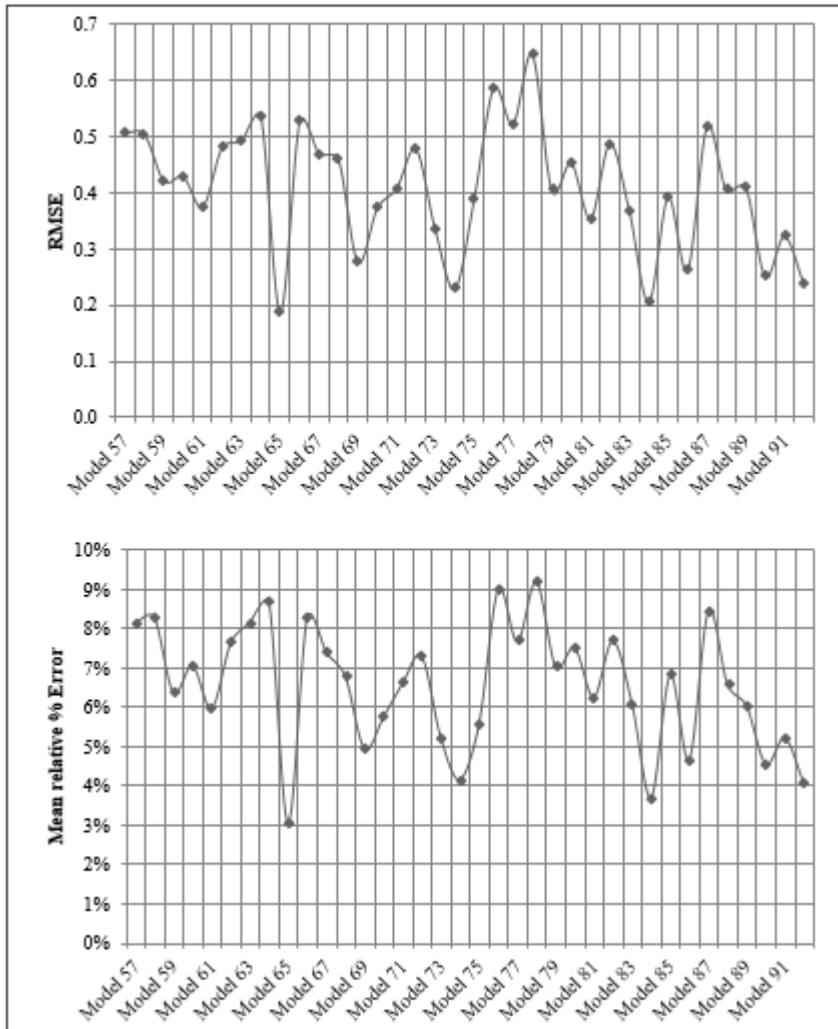
A number of 6 projects listed in table 9 were selected randomly aside from the 55 projects used for the learning process of the model; these 6 projects were not introduced to the model in the training

Table 7. Fourth run package of models

Model	Hidden Layers	No. of Neurons in 1st Layer	No. of Neurons in 2nd Layer	Transfer Function	RMSE	Mean Absolute Error	Mean % Relative Error
Model 57	2	3	3	Logistic	0.51	0.39	8.16%
Model 58	2	4	3	Logistic	0.51	0.40	8.29%
Model 59	2	4	4	Logistic	0.42	0.30	6.38%
Model 60	2	5	3	Logistic	0.43	0.34	7.07%
Model 61	2	5	4	Logistic	0.38	0.29	5.97%
Model 62	2	5	5	Logistic	0.48	0.37	7.64%
Model 63	2	6	3	Logistic	0.49	0.39	8.16%
Model 64	2	6	4	Logistic	0.54	0.41	8.67%
Model 65	2	6	5	Logistic	0.19	0.17	3.05%
Model 66	2	6	6	Logistic	0.53	0.40	8.28%
Model 67	2	7	3	Logistic	0.47	0.35	7.40%
Model 68	2	7	4	Logistic	0.46	0.32	6.78%
Model 69	2	7	5	Logistic	0.28	0.24	4.95%
Model 70	2	7	6	Logistic	0.37	0.28	5.76%
Model 71	2	7	7	Logistic	0.41	0.32	6.62%
Model 72	2	8	3	Logistic	0.48	0.35	7.30%
Model 73	2	8	4	Logistic	0.33	0.25	5.21%
Model 74	2	8	5	Logistic	0.23	0.20	4.15%
Model 75	2	8	6	Logistic	0.39	0.27	5.59%
Model 76	2	8	7	Logistic	0.59	0.43	9.01%
Model 77	2	8	8	Logistic	0.52	0.37	7.68%
Model 78	2	9	3	Logistic	0.65	0.44	9.21%
Model 79	2	9	4	Logistic	0.41	0.34	7.04%
Model 80	2	9	5	Logistic	0.46	0.36	7.52%
Model 81	2	9	6	Logistic	0.36	0.30	6.26%
Model 82	2	9	7	Logistic	0.49	0.37	7.70%
Model 83	2	9	8	Logistic	0.37	0.29	6.10%
Model 84	2	9	9	Logistic	0.21	0.18	3.69%
Model 85	2	10	3	Logistic	0.39	0.33	6.83%
Model 86	2	10	4	Logistic	0.26	0.22	4.67%
Model 87	2	10	5	Logistic	0.52	0.40	8.46%
Model 88	2	10	6	Logistic	0.41	0.31	6.58%
Model 89	2	10	7	Logistic	0.41	0.29	6.03%
Model 90	2	10	8	Logistic	0.25	0.22	4.55%
Model 91	2	10	9	Logistic	0.32	0.25	5.22%
Model 92	2	10	10	Logistic	0.24	0.19	4.05%

process for the aim of the testing process. These projects would represent new cases for the model to work on, that should test the accuracy of the developed model whether it is within the acceptable range of relative percentage error or not. If the value of the relative percentage error exceeds the acceptable

Figure 14. Results of fourth run package of models



range of the model, the prediction would be considered a wrong prediction. And if the value of the relative percentage error lies within the acceptable range, the prediction would be considered right.

The developed model has an acceptable relative percentage error of  $\pm 3.05\%$ . By considering this error margin and after running the program on the 6 test projects, the model predicted 5 out of 6 projects within the acceptable range. That represents an accuracy of approximately 84%. The accuracy of the developed model is considered to be very high for predicting the site overhead percentage of commercial projects in Egypt (see figure 18).

## 8. CONCLUSION

The research main objective was to identify the factors affecting site overhead costs for commercial projects in Egypt, and the application of neural networks to accurately estimate this percentage. Various studies on neural networks and their application in construction cost estimates have been reviewed to build a prevailing knowledge on the subject. These studies offered neural network applications for

Figure 15. Run of linear regression

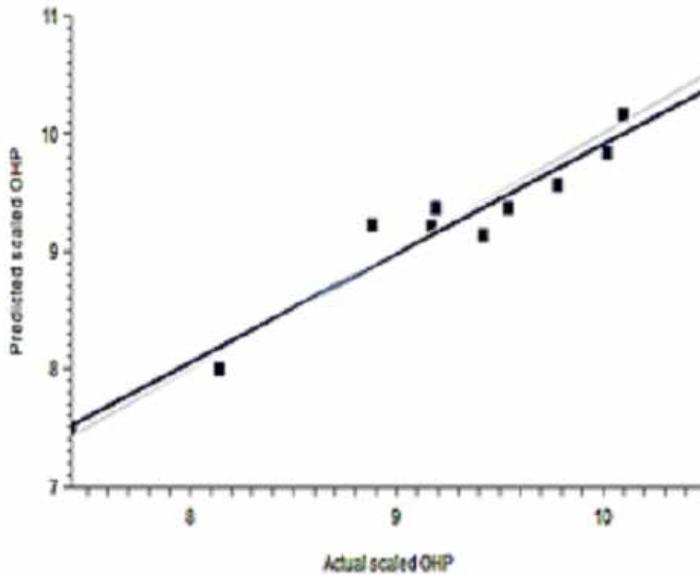


Table 8. Model 65 design considerations

No. of input neurons	No. of hidden Layers	No. of neurons in hidden layer 1	No. of neurons in hidden layer 2	Transfer function	RMSE	Relative Error %
6	2	6	5	Logistic	0.188	3.05%

Figure 16. Model 65 design architecture

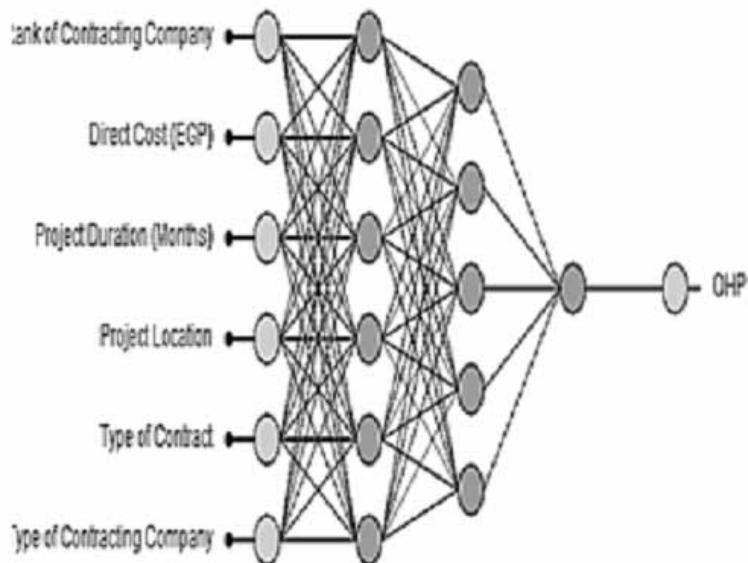


Figure 17. Base cases prediction vs. actual

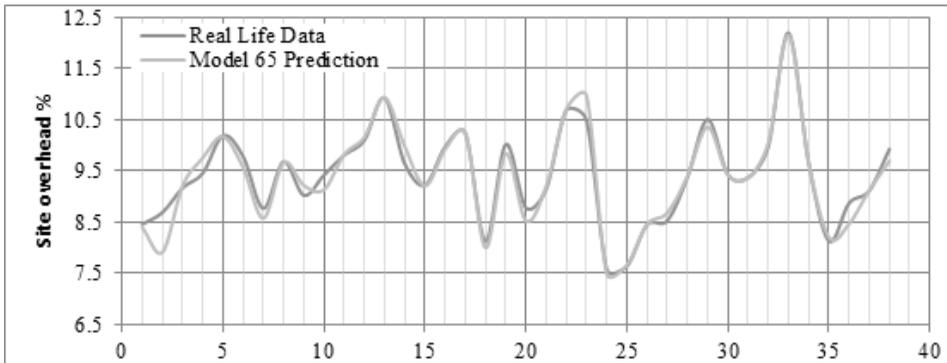
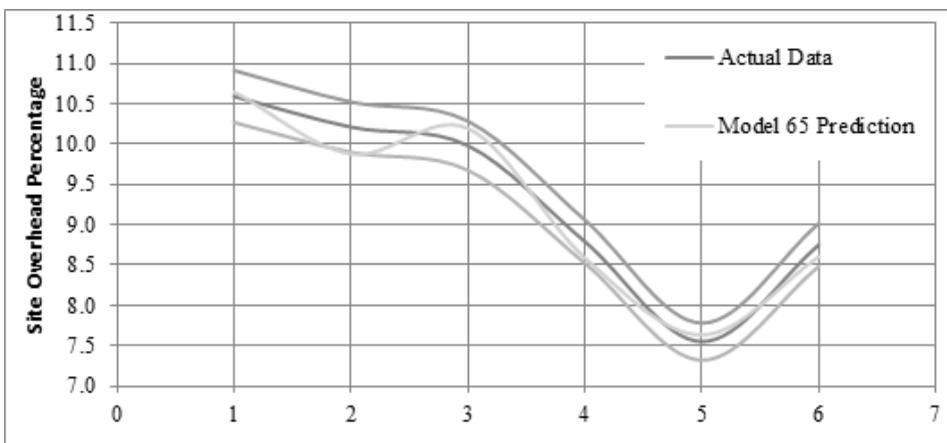


Table 9. Test results

Test Number	Model Prediction	Real life Value	Relative Error	Assessment
Test 1	10.648%	10.590%	-0.548%	Accepted
Test 2	9.878%	10.210%	3.252%	Wrong
Test 3	10.193%	9.980%	-2.134%	Accepted
Test 4	8.588%	8.790%	2.298%	Accepted
Test 5	7.633%	7.550%	-1.099%	Accepted
Test 6	8.604%	8.750%	1.669%	Accepted

Figure 18. Model 65 test assessment



cost engineering fields like structural elements costs, total construction costs, cost per square meter, cost overruns, and overhead costs. It was concluded that thirteen factors may influence site overhead costs of any project. These factors are project duration, contract type, project size, class of company, project location, inflation rate, subcontracted work, need of work, project complexity, cash availability, tendering method, number of competitors, and payment schedule. These factors were used as the

basis for a survey to reflect them on the Egyptian commercial construction industry. Participants in this survey were chosen to be experts in the construction industry especially in commercial projects and with experience of more than ten years in the commercial market. The weights from respondents were calculated on a weighted average base according to each scale weight, and the results of the survey showed five significant factors in addition to one more factor that was not listed and added by the respondents. And a final list of six factors was concluded was project location, project direct cost, project duration, contract type, class of company, and company ownership type.

Data from fifty-five commercial projects were collected according to the factors concluded from this research; these projects were commercial projects that were executed in Egypt in the past seven years. Then a comparative analysis was conducted to study the influence of each factor on the site overhead percentage, this analysis was useful in understanding how each factor choice may influence the overhead percentage of a project. Also, this data would be used as training data for the neural network during the training phase. The collected data and its analysis set the ground base for developing the neural network for estimating the site overhead percentage for commercial projects in Egypt. The “Neural Designer” software was used to develop the neural network model and the process of designing the neural networks went through different steps: coding the data set and importing it to the “Neural designer” for the training process, determining the training algorithm and strategy, determining the network architecture, validation and testing of the model, and developing a GUI for the model

Ninety-two models were developed to reach the optimum network structure, the selected model was developed using the traditional trial and error process, using the back-propagation training algorithm. The developing of the model was based on the loss index of the Root Mean Squared Error (RMSE), and the selection was based on the model with the lowest RMSE. The trial model number 65 was the best choice with a network architecture of 2 hidden layers with 6 neurons in the first layer and 5 neurons in the second layer, with a logistic transfer function for each layer, RMSE of 0.188 and corresponding to a relative percentage error of 3.05%. Then the model was tested on 6 projects which were not introduced before to the network to determine the accuracy of the developed model. The test result was predicting 5 out of 6 projects within the acceptable range and only 1 out of the acceptable range of the model, this results in an accuracy of 84% that is considered to be a very accurate prediction. For the aim of implementing the model into a user-friendly interface, the model function was exported to a Python script and was used to develop a GUI for the model to be used by anyone who cannot deal with the “Neural Designer” software.

It was concluded that establishing a database for the finished projects and projects on-hand is recommended, this is for the aim of using it in developing any decision-making model and to get the full benefit from the lessons learned. Also, researchers in the future may focus on a specific aspect in the construction industry and gather all the possible factors affecting it to figure out the governing relation between them. This focus will enhance the accuracy of the models as data would be specific and in direct relation with the target output.

As a summary, the study showed that site overhead costs for commercial construction projects in Egypt are affected by six major factors which are the class of contracting company, project location, project direct cost, project duration, contract type, and ownership type of contracting company. The study also demonstrates the benefits of using artificial intelligence and neural networks in the estimating process of the site overhead costs, and the undeniable advantages of the proposed model to predict the site overhead percentage for commercial construction projects in Egypt are its simplicity, calculation speed, and accuracy. Results of testing the proposed model proved to be very accurate with an accuracy of 84%. The accuracy of the model depends on the amount of the learning data and the degree of accuracy of these data, these data should cover many cases with different situations for the model to build a relation between the different inputs and their effect on the targeted output. Also determining the suitable learning algorithm and activation function for the model plays an important role in developing an accurate calculation model. The testing and validation process of the model is

an essential process to ensure the prediction accuracy of the model, this process should be done using pre-known data inputs and outputs, but not previously introduced to the training model.

The model is much more accurate and simpler to use and minimizes the estimation duration of such items. That will enhance the contractor's ability to accurately predict their costs, and so maximize their profit and their competitive advantage. It is recommended for the construction companies to establish a database for the finished projects and projects on-hand to be used as historical data in developing any decision-making model and to get the full benefit from the lesson learned. A graphical user interface should be developed for any further neural network application for the estimator to deal with it without any previous knowledge of the neural networks or how it may perform.

## REFERENCES

- A, E., & A, A. (2008). Investigating the overhead costs in construction projects in Palestine. *Journal of Financial Management of Property and Construction*, 13(1), 35-47.
- C, C., & C, P. (2002). Estimation of project overheads: A contractor's perspective. *18th Annual ARCOM Conference*, (pp. 53-63).
- Car-Pusic, D., Petruseva, S., Pancovska, V. Z., & Zafirovski, Z. (2020). Neural Network-Based Model for Predicting Preliminary Construction Cost as Part of Cost Predicting System. *Advances in Civil Engineering*, 2020, 2020. doi:10.1155/2020/8886170
- Chan, C. T. W.C.T. (2012). The Principal factors affecting construction project overhead expenses: An exploratory factor analysis approach. *Construction Management and Economics Journal*, 30(10), 903-914. doi:10.1080/01446193.2012.717706
- Chao, L.-C.L.C. (2010). Estimating project overheads rate in bidding: DSS approach using neural networks. *Construction Management and Economics Journal*, 28(3), 287-299. doi:10.1080/01446190903473782
- Chen, H., Yu, Z., Yang, Q., & Shao, J. (2022). Community detection in subspace of attribute. *Information Sciences*, 602, 220-235. doi:10.1016/j.ins.2022.04.047
- Cheng, M., Tsai, H., & Sudjono, E. (2010). Conceptual cost estimates using evolutionary fuzzy hybrid neural network. *Expert Systems with application*.
- Elhegazy, H., Badra, N., Aboul Haggag, S., & Abdel Rashid, I. (2022). implementation of the neural networks for improving the projects' performance of steel structure projects. *Journal Of Industrial Integration And Management*, 7(10).
- ElSawy, I., Hosny, H., & AbdElRazik, M. (2011, May). A Neural Network Model for Construction Projects Overhead Cost Estimating in Egypt. *International Journal of Computer Science Issues*, 8(3).
- Emsley, W. (2002). *Data modeling and the application of a neural network approach to the prediction of total construction costs*. Construction Management and Economics Journal.
- Georgy, M. B. S. (2005, December). Artificial Neural Networks Model for Parametric Estimating of Construction Project Costs. *Journal of Engineering and Applied Science*, 52(6), 1050-1066.
- Hassouna, D. H., Khedr, A. E., Idrees, A. M., & ElSeddawy, A. I. (2020). Intelligent Personalized System for Enhancing the Quality of Learning. *Journal of Theoretical and Applied Information Technology*, 98(13), 2199-2213.
- Hastak, M. (2015). *Skills and Knowledge of Cost Engineering* (6th ed.). AACE International.
- Hosny, O. (2006). An Artificial Neural Network Approach for Predicting the Effect of Claims on Project Time and Cost. *Engineering Research Journal*, 106, C87-C106.
- Idrees, A. M., Alsheref, F. K., & ElSeddawy, A. I. (2019). A Proposed Model for Detecting Facebook News' Credibility. [IJACSA]. *International Journal of Advanced Computer Science and Applications*, 10(7), 311-316. doi:10.14569/IJACSA.2019.0100743
- Idrees, A. M., El Seddawy, A. I., & Moaaz, E. L. (2019). A Proposed Mining Based Business Continuity Information System for Educational Institutes. *Journal of Computational Science*, 15(8), 1133-1149. doi:10.3844/jcssp.2019.1133.1149
- Idrees, A. M., ElSeddawy, A. I., & Zeidan, M. O. (2019). Knowledge Discovery based Framework for Enhancing the House of Quality. [IJACSA]. *International Journal of Advanced Computer Science and Applications*, 10(7), 324-331. doi:10.14569/IJACSA.2019.0100745
- Khedr, A. E., El Seddawy, A. I., & Idrees, A. M. (2014). Performance Tuning of K-Mean Clustering Algorithm a Step towards Efficient DSS. [IJRCST]. *International Journal of Innovative Research in Computer Science & Technology*, 2(6), 111-118.
- Khedr, A. E., Idrees, A. M., & El Seddawy, A. I. (2016). Enhancing Iterative Dichotomiser 3 algorithm for classification decision tree. *WIREs Data Mining Knowledge Discovery*, 6(2), 70-79. doi:10.1002/widm.1177

Mostofi, F., Toğgan, V., Ayözen, Y. E., & Tokdemir, O. (2022). Predicting the Impact of Construction Rework Cost Using an Ensemble Classifier. *Sustainability, 14*(22), 14. doi:10.3390/su142214800

Roxas, L., C., & Maximino, J. (2014). An Artificial Neural Network Approach to Structural Cost Estimation of Building Projects in the Philippines. *DLSU Research Congress*. Manila.

S, A., A, A., & M, A. (2001). The management of construction company overhead costs. *International Journal of Project Management, 19*(5), 295-303.

S, A., A, B., & M, A. (1999). Project overhead costs in Saudi Arabia. *Cost Engineering-ANN Arbor then Morgantown, 41*, 33-37.

Wei, B. (2022). Application of Neural Network Based on Multisource Information Fusion in Production Cost Prediction. *Wireless Communications and Mobile Computing, 2022*.