Parametric Cost Estimating of Sterile Building Using Artificial Neural Network & Genetic Algorithm Model

By Mohamed Zahran , Hossam Hosny and Abdelmonem Sand

Abstract— **This paper presents a parametric cost estimating model by using artificial neural network and genetic algorithms, for a cretin type of projects in the industrial field. These projects include sterile buildings, such as pharmaceutical and foods industrial projects. An extensive survey has been conducted to identify the most important cost indicators for such type of projects. Historical Data of 14 of previous similar projects have been collected. Based on the derived cost indicators and the collected historical data, the required model have been developed and validated in process of parametric cost estimating for such projects in the early stage of project life cycle.**

Index Terms— **parametric cost estimating, early stage, sterile buildings, cost indicator, neural network, genetic algorithms**

I. Introduction

Conceptual cost estimating is one of the most important and challenging activities during project planning, which occurs at the early stages of a project life where limited information is available and many unknown factors affecting the project costs.
Every project begins its life with a concept proposed by the owner and refined by the designer. Planning decisions in this early stage of any project are vital, as it can have the biggest influence on the subsequent outcome of the project. Conceptual cost estimating is the determination of the project's total costs based only on general early concepts of the project.
While many studies have indicated the importance of accurate conceptual cost estimates, there has been little effort directed at improving the conceptual cost estimate processes, especially for construction projects in the industrial field.
Any industrial construction project is a very complex undertaking, which can be composed of hundreds or thousands of construction work items. These work items are often performed by workers or crews from different crafts, utilizing various materials of many different varieties. Due to these complexities, numerous factors can affect the construction processes and ultimately their costs.
The research focus was directed on industrial construction projects, specifically aimed at project encompassing special hygienic buildings called sterile buildings, such as pharmaceutical, food and dairy industrial projects. The design and construction of those types of buildings require additional considerations to comply with the rules and regulations for both regional and international authorities and markets, while there is absence of researches in this area
The main objective of this research is to develop accurate reliable and practical method of systematic parametric cost estimating that can be used by organizations involved in the planning and execution of industrial construction projects.
The intended modeling methodology will be based on using Artificial Neural Network (ANN) technique to develop the required cost estimating model. Moreover genetic algorithms (GA) will be used to produce solution of the network. The ANN-GA +model will use Excel spread sheets as a data base information modeling and Evolver software as genetic algorithm based program.

II. Literature Review

Parametric cost estimating is a method of evaluating the costs of a project from the parameters characterizing the project but without describing it completely, using historical data from similar projects (Charles, 2006).
Recently, the artificial intelligence applications have been widely used in cost estimation for construction projects. Among all artificial intelligence areas, Artificial Neural Network (ANN) had been proved itself as the most promising technique this may be due to its ability to learn by itself, generalize solutions, and adequately respond to highly correlated, noisy, incomplete, or previously unseen data (El Gafy, 2001). Construction engineering and management has been considered a fertile field for many neural networks applications.

A. Morcous (1997) developed a neural network model for the purpose of estimating the quantities and costs of reinforced concrete bridges over the Nile River

B. Setyawati et al. (2002) developed a neural network for cost estimation. They suggested regression analysis with combined methods based on percentage errors for obtaining the appropriate linear regression which describe the artificial neural network models for cost estimating.

C. Murat et al. (2004) developed a cost estimation model for building based on the structural system for the early design process. They suggested that their model establishes a methodology that can provide an economical and rapid means of cost estimating for the structural system of future building design process. They argued that neural networks are capable to reduce the uncertainties of estimate for a structural system of building, while the accuracy level of the developed model was 93%.

D. Jamshid (2005) also examined cost estimation for highway projects by artificial neural network and argues that neural network approach might cope even with noisy data or imprecise data. They reported that artificial neural network could be an

Manuscript received April 12, 2014.

Mohamed Zahran M.Sc. Student, Building and Construction Dept., Faculty of Engineering, Arab Academy for Science and Technology and Maritime Transport Cairo, Egypt.

Hossam Hosny Professor, Construction Engineering Dept., Faculty of Engineering, Zagazig University Zagazig, Egypt

Abdelmonem Sanad Professor, Building and Construction Dept., Faculty of Engineering, Arab Academy for Science and Technology and Maritime Transport, Cairo, Egypt.

appropriate tool to help in solving problems which comes from a number of uncertainties such as cost estimation at the conceptual phase.

E. Arafa (2011) developed an Artificial Neural Networks model to estimate the cost of building construction projects at the early stage, for Gaza strip, Palestine.

F. Hosny (2011) developed an Artificial Neural Networks model that can materially help construction planners in the accurate determination of the expected time contingency of any future building projects.

G. Asal (2014) developed an Artificial Neural Networks model that can estimate the cost contingency which must be added to the actual cost of any future building projects.

III. Data Collection & Analysis

The identification of the cost indicators or the most important cost factors affecting projects' total cost is a crucial step to develop a reliable cost estimating model. In this research, the determination of cost factors was reached collectively through two approaches; first one, the factors concluded at the comprehensive literature review from the previous studies, 14 factors were collected, these factors presented in the first section of table (1). Second one, the cost factors derived from interviews conducted with three project managers and five cost experts in the pharmaceutical and food industrial projects, 22 factors were collected, these factors presented in the second section of table (1).

As a result, the total number of determined cost factors is 36.

Table 1. The Determined Cost Factors

Factors collected from the literature review
1.Project location
2.Desired completion time for the project
3.Site topography
4.Accumulative built-up area
5.Other supplementary buildings (W.tank, administration, warehouse,… etc)
6.Desired structural system
7.Consultant fees
8.Desired level of contractor's prequalification
9.Contractor overhead
10. Need for special contractor(s)
11. Reinforcement price
12. Cement price
13. Labor price
14. Inflation
Factors derived from interviews
1.Site accessibility
2.Site Constraints
3.Subsistence of time constrains
4.Owner requirements for bid packaging for multiple Contractors
5.Environmental impact assurance system requirements
6.Applying safety system during construction
7.Buildings closeness (attached, semi-attached or separated)
8.Accumulative Sterile Areas (total area)
9.Structural design loads
10. Geotechnical nature of soil
11. Desired HVAC system
12. Desired Firefighting system
13. stainless Steel price
14. Percentage of imported material
15. Availability of required power
16. Target market (regional or international)
17. Type of products and type of production method
18. Special finishing required for sterile areas
19. Additional requirements for structural system regarding sterile manufacturing
20. Additional requirements for HVAC system regarding sterile manufacturing
21. Industrial safety requirements (firefighting, fire alarm, .. etc)
22. Currency exchange rate
International insurances (if any)

In order to identify the most important cost factors "Cost Indicators" through the previously determined cost factors, a questionnaire survey was conducted for the purpose of gathering experts opinions, the questionnaire was built to obtain experts responses for both "Impact" and " Degree of Existence" for each cost factor.

This study used probability sampling technique for infinite population to calculate the required sample size (respondents). The calculations were based on confidence level 95%, and confidence interval 15% (Godden, 2004). The Sample Size was computed as per the following equation:

$$SS = \frac{Z^2 \times (p) \times (1-p)}{C^2} \qquad (1)$$

Where SS = Sample Size, Z = Z-values (Cumulative Normal Probability), the equivalent Z-value for a 95 percent confidence level is (1.96), P is equal to (20%) according to the number of answers (five answers) for each question, but since 50% is the critical case percentage in the calculation of sample size, the value used for P in the equation is 0.5, and C = Confidence interval, expressed as decimal (0.15 = +/- 15 percentage points)

Therefore:

$$SS = \frac{(1.96)^2 \times (0.5) \times (1-0.5)}{(0.15)^2} = 43$$

The required number of respondents is not less than 40 experts. However, the target number of questionnaire recipients shall consider a percentage of about 40% of no response to the questionnaire, thereafter, the target number of questionnaire recipients will be as follows

43 X (1+40%) = 60 expert. (2)

After the effort done to send the questionnaire and collect the experts opinions, the questionnaire responses were analyzed though two stages statistical analysis, then logical relations study. First stage, statistical analysis, was conducted based on calculating the importance of each cost factor (Importance Index, IMP.I) by multiplying the average weighted impact of each factor (Severity Index, S.I) times its average weighted degree of existence (Frequency Index, F.I), by using the following equations:

$$(IMP.I) = F.I * S.I \qquad (3)$$

$$(F.I) = \Sigma a_F * n / N \quad (4)$$

$$(S.I) = \Sigma a_S * n / N \quad (5)$$

Where: a = constant expressing the weight assigned to each responses (ranges from 1 for very low to 5 for very high), n = frequency of each response, N = total number of responses.

As a result, all factors were ranked in a descending order according to their Importance Index. Afterwards, in order to determine the most important factors from the ranked factors, a datum of 60% for the relative importance was set to distinguish between the most important cost factors and lowest important cost factors. The relative importance of each cost factor was calculated as percentage reference to the highest importance index which has 100% relative importance.

The deduced factors from this stage "The Most important Cost Factor" have approximate accumulative ratio equal to 80% of the total importance index "the summation of all importance indices". Table (2) represents the statistical analysis done in this stage.

Table 2. The Statistical Analysis

No.	Cost Factors	Importance Index	%	Ratio (%)	Accum-ulative Ratio
1	Currency exchange rate	19.04	100	3.93	78.45%
2	Desired HVAC system	18.13	95.2	3.74	
3	Inflation	17.33	91.1	3.58	
4	Accumulative built-up area	17.17	90.2	3.54	
5	Special finishing required for sterile areas	16.77	88.1	3.46	
6	Desired Firefighting system	16.46	86.5	3.40	
7	Availability of required power	16.42	86.2	3.39	
8	Additional requirements for HVAC system regarding sterile manufacturing	16.30	85.6	3.36	
9	Desired completion time for the project	16.07	84.4	3.32	
10	Accumulative Sterile Areas (total area)	15.61	82	3.22	
11	Reinforcement price	15.15	79.6	3.13	
12	% of imported material	14.84	77.9	3.06	
13	Additional requirements for structural system regarding sterile manufacturing	14.50	76.1	2.99	
14	Desired level of contractor's prequalification	14.48	76.1	2.99	
15	Target market (regional or international)	13.94	73.2	2.88	
16	Contractor overhead	13.85	72.7	2.86	
17	Subsistence of time constrains	13.66	71.8	2.82	
18	International insurances (if any)	13.28	69.7	2.74	
19	Other supplementary buildings (W.tank, administration, warehouse,… etc)	13.06	68.6	2.69	
20	Need for special contractor(s)	12.69	66.6	2.62	
21	Desired structural system	12.42	65.2	2.56	
22	Cement price	12.12	63.6	2.50	
23	Project location	11.83	62.1	2.44	
24	Buildings closeness (attached, semi-attached or separated)	11.78	61.9	2.43	
25	Site topography	11.70	61.4	2.41	
26	Labor price	11.69	61.4	2.41	
27	Industrial safety requirements (fire fighting, fire alarm, .. etc)	11.35	59.6	2.34%	21.55%
28	Geotechnical nature of soil	10.88	57.1	2.24	
29	Structural design loads	10.81	56.8	2.23	
30	Applying safety system during construction	10.56	55.5	2.18	
31	stainless Steel price	9.95	52.3	2.05	
32	Type of products and type of production method	9.95	52.2	2.05	
33	Consultant fees	9.06	47.6	1.87	
34	Site accessibility	8.54	44.9	1.76	
35	Environmental impact assurance system requirements	8.51	44.7	1.76	
36	Owner requirements for bid packaging for multiple contractors	8.42	44.2	1.74	
37	Site Constraints	6.44	33.8	1.33	
Total		484.78	-	100	100%

Where:

$$\text{Percentage } (\%) = \frac{IMP.I_n}{\text{Maximam } IMP.I} \times 100 \quad (6)$$

$$\text{Ratio} = \frac{IMP.I_n}{\Sigma\, IMP.I} \times 100 \quad (7)$$

$$\text{Acc. Ratio} = \Sigma_n^1 \text{ Ratio} \quad (8)$$

Table (2) shows that the first 26 factors are the most important factors. These factors were shaded by gray color and will proceed to the next analysis stage to select the cost indicators that will be used to develop the ANN model.

Through the second stage of questionnaire analysis "the logical relation study", an extensive study was conducted on the highest (most) important cost factors, which were concluded from the first Stage, to group factors having logical correlation and to select one prominent factor from each group as an indicator, to produce the final list of Cost Indicators. The main purpose of this process was to eliminate redundancy and simplify the input data required to the creation of the Neural Model.

The study concluded that, the Consumer price index can be used as an indicator for reinforcement price and other material prices, labor prices and inflation, due to the proportional relation between these factors, also the accumulative sterile areas directly influence the amount of special finishing required for sterile areas, additional requirements for HVAC system regarding sterile manufacturing and additional requirements for structural system regarding sterile manufacturing, due to the same relationship between these cost factors and the selected cost indicator.

In addition, project location was found to be a realistic cost indicator representing site topography and availability of required power, also the target market strongly affects the selection of desired HVAC system, percentage of imported materials and the desired firefighting system.

On the other hand, contractor overhead and need for special contractor can be indicated by desired Level of Contractor's prequalification according to the inevitable logical relation between the contractor overhead and the contractor prequalification, finally desired completion time for the project was selected to represent the subsistence of time constrain. Table (3) represents the selected Cost Indicators for all important cost factors.

In addition to the concluded cost indicators, a new cost factor (Project Status) emerged as a result of multiple suggestions from respondents; this cost factor was found to be of great value to the Neural Model, as it indicates the status of the project as a new project, extension or renovation, this factor presented in table (3) indicator No. 13.

Table 3. The Selected Cost Indicators

No.	Cost Indicator	Grouped Factors (Factors having a logical correlation)
1	Currency exchange rate	1. Currency exchange rate
2	Consumer price index	2. Reinforcement price
		3. Cement price
		4. Labor price
		5. Inflation
3	Accumulative built-up area	6. Accumulative built-up area
4	Accumulative Sterile Areas (total area)	7. Accumulative Sterile Areas (total area)
		8. Special finishing required for sterile areas
		9. Additional requirements for HVAC system regarding sterile manufacturing
		10. Additional requirements for structural system regarding sterile manufacturing
5	Project location	11. Project location
		12. Site topography
		13. Availability of required power
6	Target market (regional or international)	14. Target market (regional or international)
		15. Desired HVAC system
		16. Desired Fire fighting system
		17. % of imported material
7	International insurances (if any)	18. International insurances (if any)
8	Desired level of contractor's prequalification	19. Desired level of contractor's prequalification
		20. Contractor overhead
		21. Need for special contractor(s)
9	Desired completion time for the project	22. Desired completion time for the project
		23. Subsistence of time constrains
10	Other supplementary buildings (W. tank, gate house,… etc)	24. Other supplementary buildings (W. tank, administration, warehouse,… etc)
11	Desired structural system	25. Desired structural system
12	Buildings closeness (attached, semi-attached or separated)	26. Buildings closeness (attached, semi-attached or separated)
13	Project Status	

These cost indicators were used in two crucial actions; first one is the process historical data collection; the collected projects must full fill those indicators, otherwise, the project was neglected. Second on is developing the proposed ANN model; these indicators will be the input data for the model.

IV. Model Development

Design of the artificial neural network model requires important course of actions; (1) selection of the used Software(s) for both modeling and simulation (optimizing), (2) determination of the network architecture and type, (3) historical data collection and categorization, (4) model execution, (5) model implementation, (6) trial and error practices and (7) model validation.

A. *Selection of Used Software(s)*

For the modeling, Microsoft Excel 2010 was the selected software to be used in this research as the data base software to develop the neural network model. This software is running under Microsoft windows 7 operating system. Microsoft Excel is designed to be user friendly; allowing its users to simply construct a neural network model without having extensive programming knowledge.

For the optimization process, Evolver add-in version 5.1.1 was added to Microsoft Excel, this software produced by Palisade Corporation. Evolver uses Genetic Algorithms as a technique to search for the near optimal solution throughout the optimization process.

One of the main advantages of this software is simplicity of usage for its user. It easily allows the user to design and apply any required constrains with any number of constrains. It also allows users to customize the values of some important features such as the permutation and crossover. Finally the installation compatibility of this software with all Microsoft Excel versions is a vital attribute.

B. Determination of the ANN architecture and type

The structure of the neural network was designed to consist of three parts; the input layer, one or two hidden layers and the output layer. Each layer encloses a certain number of nodes, each node in the input layer and output layer were linked to all nodes enclosed in the hidden layers by a different weight.

The Input Layer, the first one, enclosing 13 nodes, each node represents one of the selected cost indicators, as concluded in the previous chapter, each cost indicator will be determined by a value; for quantitative indicators, the exact value will be used, however, for qualitative indicators, a preselected value will be used to indicate each case. Table (4) represents the determination values for each cost indicator.

Table 4. The determination value for each Cost Indicators

No.	Cost Indicator	Determination Value
I_1	Currency exchange rate	Exact Value
I_2	Consumer price index	Exact Value
I_3	Desired completion time for the project	Exact Value
I_4	Accumulative built-up area	Exact Value
I_5	Accumulative Sterile Areas (total area)	Exact Value
I_6	Other supplementary buildings (W. tank, gate house,… etc) (total area)	Exact Value
I_7	Desired structural system	(1) For mixed (2) For concrete (3) For Steel
I_8	Buildings closeness (attached, semi-attached or separated)	(1) For attached (2) For semi-attached (3) For separated
I_9	Project Status	(1) For renovation (2) For extension (3) For new building in existing project (4) For entirely new project
I_{10}	Project location	(1) For inside Cairo (2) For areas outside Cairo and till 10th of Ramadan city (3) For areas farther than 10th of Ramadan city
I_{11}	Target market (regional or international)	(1) For regional market (2) For international market
I_{12}	International insurances (if any)	(1) No (2) Yes
I_{13}	Desired level of contractor's prequalification	(1) Normal (2) Moderate (3) High

The Hidden Layer(s), the second one, in this layer(s) the number on nodes (hidden nodes) ware calculated by considering one guidance that the number of hidden nodes must be not less than half the summation of the number of nodes in the input and output layers (Hosny, 2011).

Accordingly, 8 hidden nodes ware used in this layer, also an activation function will be used to activate data derived into these 8 hidden nodes. In the trial and error practices, another hidden layer will be added to a new model to be used in a deferent set of trials, the number of hidden nodes in this layer will not be strict to the above mentioned guidance; only 4 hidden nodes will be used in this layer.

The Output Layer, the third layer, this layer encloses only one output neuron representing "Predicted Cost", considering the scaling done in the input layer, data in this layer will be scaled back.

This research tends to use feed forward type of artificial neural network and back propagation learning algorithm. A supervised learning technique will be used; where Inputs were fed to the proposed network model and the outputs then calculated. The differences between the calculated outputs and the actual outputs were then evaluated until *the learning rule is attained*. Learning rule is a procedure for modifying the weights and biases of a network to produce a desirable state.

C. Historical Data Collection Categorization

Historical data for 18 individual projects were collected. These projects have the same type and nature of projects targeted by this research. These historical data were sorted into two categories, training and testing data. First category, the training set of projects' data, represents about 75% - 80% of the collected facts. Accordingly, 14 projects' data were randomly selected to be the training set of data. These set of data will be used to train the model by reducing the difference between the actual cost and the predicted cost by calculating then minimizing the Root of Mean Square Error (RMSE).

Consequently, the second category, the testing set of projects' data, will be 4 projects' data representing 20% - 25% of collected data. This set will be used to monitor the results produced during training the model. Also the average percentage error for this group will be calculated to validate the victor model.

D. Model Execution

The model execution was done by applying the following steps:

Step (1) Data Input: All projects' data were inserted into Microsoft Office Excel in a table consisting of fourteen horizontal input fields for each project; one for the actual cost and thirteen fields representing the cost indicators.

Step (2) Data Scaling: All input data ware scaled to a range from (-1 to 1) using maximum and minimum values of each input filed to suit Neural Networks processing. This was done by using linear equation for scaling:

$$\text{Scaled Value} = \frac{2*(\text{Original value} - \text{Min.value})}{(\text{Max.value} - \text{Min.value})} - 1 \quad (9)$$

Step (3) Weighted Input Sum: Each input was connected to all hidden nodes by a weight. In the model with two hidden layers, each node in the first layer was also connected to all nodes in the second layer using the same previous concept.

The summation of weighted inputs was then calculated using the following equation:

$$X_h = \sum_{i=1}^{13}(I_i * W_{hi}) \qquad (10)$$

Step (4) Activation of Weighted Input Sum: The previously calculated input summations were activated three times separately regarding the trial and error practices. The used activation functions were 'Tanh', 'Sigmoid' and 'Non'. The following equation was used in activation process:

$$H_{h\,activated} = f(X_h) \qquad (11)$$

Thereafter, the model results for each trial were saved in order to select the best fit activation function. The outputs from this step "activated summation" are considered inputs for the next step; either for the second hidden layer or the output layer. In case of second hidden layer, the activated summations of this layer are the inputs for the Output layer.

Step (5) Weighted Output Sum & Activation: The hidden nodes, in the first or second hidden layer, where then connected to the Output node by weights. The weighted summation was then calculated using the following equation:

$$X_o = \sum_{h=1}^{N}(H_h * W_{ho}) \qquad (12)$$

Thereafter an exit activation function was used to activate the weighted summations, as demonstrated in the previous step (4), the weighted summation where activated three times using the same activation functions, and the following equation was used:

$$O = f(X_o) \qquad (13)$$

The output from this step is the scaled predicted cost

Step (6) Output: The previously calculated outputs "Scaled Predicted Cost" were then interpreted by scaling back calculation, using below equation, to deduce the predicted cost.

$$\text{Scaled Back Value} = \frac{(\text{output value}+1)*(\text{Max.value}-\text{Min.value})}{2} + \text{Min. value} \qquad (14)$$

Step (7) Calculating Error**:** The difference between the predicted cost and the actual cost (Error) was calculated for the training set, thereafter the Root of Mean square Error (RMSE) was calculated by the applying the following equation:

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(Pi-Ai)}{n}} \qquad (15)$$

Where: (n) is the number of training samples to be evaluated in the training phase, (Ai) is the actual output of the training sample, and (Pi) is the predicted output for the same training sample.

Moreover, the average percentage of difference (%Error) between the Predicted cost (Ai) and the Actual cost (Pi) was calculated for training set for monitoring process and also calculated for testing set for validating the model, according to the following equation:

$$\%\,Error = \frac{(Pi-Ai)}{Ai}x100 \qquad (16)$$

Step (8) Learning Role For Optimization Goal: The final step in the model execution is defining the learning role for the optimization purpose; this role was assigned, in the Evolver, to minimize the summation of the previously calculated RMSE for the training set.

E. *Model Implementation*

After finishing of the Model Execution stage, it's quite important to focus on the settings adjusted in the Evolver before starting simulation. Firstly, all cells designated for input connecting weights were attributed in the Evolver as adjustable cells with decimal fraction values to allow Evolver to search for the best weight values. Similarly all cells designated for output connecting weights were also attributed in the Evolver by the same way. Settings of the adjustable cells, such as mutation and crossover were kept in default value (0.1 & 0.5) respectively.

By the end of all previous steps and settings assignment, it is time to run the model and start optimization and gaining results.

F. *Trial And Error Practices*

To verify this research work, trial and error practices were carried out to conclude the best model. Thirty six trials were applied for model training. These trials were performed in two different groups as shown in Table (5).The table represents a complete summary of description for both group of trials, It is divided into seven fields; trials group number, description of the group, number of hidden layers, number of nodes in each layer, activation function for each hidden layer, exit activation function for output layer and the total number of trials in each group.

Table 5. Trials and error practices Summary

Group No.	Description	number of nodes in each Hidden Layer	Activation Function for Each Hidden Layer(s)	Exit Activation Function	Total No. of Trials
(1)	One hidden layer	• 8 nodes	• Tanh • Sigmoid • Non	• Tanh • Sigmoid • Non	9
(2)	Two hidden layers	• 8 nodes for first layer • 4 nodes for 2nd layer	• Tanh • Sigmoid • Non	• Tanh • Sigmoid • Non	27
Total					36

Table (6) shows the detailed log of all trials. It contains the number of hidden layers in each trial, the number of hidden nodes in each layer, activation function of each hidden layer, exit activation function in the output layer, root mean square error RMSE and the training Error (%Error).

Table 6. Detailed Log for trials and error practices

Trial No.	No. of hidden Layers	No. of hidden Nodes	1st Layer Activation Function	2nd Layer Activation Function	Exit Function	RMSE	% Error (training)
1	1	8	Tanh	N/A	Tanh	2510157.91	19.0%
2	1	8	Tanh	N/A	Sig	31417157.87	49.3%
3	1	8	Tanh	N/A	Non	1151929.69	11.1%
4	1	8	Sig	N/A	Tanh	3415072.69	27.9%
5	1	8	Sig	N/A	Sig	31417182.01	49.4%
6	1	8	Sig	N/A	Non	476374.08	13.1%
7	1	8	Non	N/A	Tanh	4707821.18	41.1%
8	1	8	Non	N/A	Sig	31417157.87	49.5%
9	1	8	Non	N/A	Non	31417157.87	49.4%

Trial No.	No. of hidden Layers	No. of hidden Nodes	1st Layer Activation Function	2nd Layer Activation Function	Exit Function	RMSE	% Error (training)
10	2	8 & 4	Tanh	Tanh	Tanh	2072249.63	16.1%
11	2	8 & 4	Tanh	Tanh	Sig	31417157.87	49.6%
12	2	8 & 4	Tanh	Tanh	Non	527869.96	12.6%
13	2	8 & 4	Tanh	Sig	Tanh	2480541.99	23.0%
14	2	8 & 4	Tanh	Sig	Sig	31417157.94	49.8%
15	2	8 & 4	Tanh	Sig	Non	3181757.65	27.9%
16	2	8 & 4	Tanh	Non	Tanh	4521030.52	35.5%
17	2	8 & 4	Tanh	Non	Sig	31417157.87	49.9%
18	2	8 & 4	Tanh	Non	Non	543916.28	12.2%
19	2	8 & 4	Sig	Tanh	Tanh	1381981.08	22.3%
20	2	8 & 4	Sig	Tanh	Sig	31417157.87	51.5%
21	2	8 & 4	Sig	Tanh	Non	3577819.94	31.7%
22	2	8 & 4	Sig	Sig	Tanh	1181066.30	31.5%
23	2	8 & 4	Sig	Sig	Sig	31739998.80	51.6%
24	2	8 & 4	Sig	Sig	Non	1260278.83	34.4%
25	2	8 & 4	Sig	Non	Tanh	5043942.99	46.2%
26	2	8 & 4	Sig	Non	Sig	31417157.87	49.6%
27	2	8 & 4	Sig	Non	Non	1011171.79	29.5%
28	2	8 & 4	Non	Tanh	Tanh	4504308.19	37.7%
29	2	8 & 4	Non	Tanh	Sig	31417379.69	50.2%
30	2	8 & 4	Non	Tanh	Non	1873278.30	41.2%
31	2	8 & 4	Non	Sig	Tanh	1124561.11	22.9%
32	2	8 & 4	Non	Sig	Sig	31417799.29	50.6%
33	2	8 & 4	Non	Sig	Non	3873127.00	45.1%
34	2	8 & 4	Non	Non	Tanh	5833422.80	47.7%
35	2	8 & 4	Non	Non	Sig	31417157.87	49.6%
36	2	8 & 4	Non	Non	Non	6581432.48	47.9%

As shown in table (6), the minimum RMS was concluded in trial number (6), this trial was clouded by gray in the table. Therefore, it is the recommended structure which should be tested. This structure consists of one hidden layers with activation function Sigmoid for summation of weighted Inputs, where number of hidden nodes were 8, whilst the exit function for output node was "Non". In addition, the average absolute percentage of training error (%Error) was (13.1%).

G. *Model Validation*

Validating the developed neural network is essential to prove that there are no corrections or modification are required after the training process. If the results are good, the network will be ready to use. If not, this needs more or better data or to redesign the network. A part of the facts around 20%, i.e. four facts are set aside randomly from training facts. These facts are used to test the ability of network to predict a new output. The model predicts the expected project total construction cost.

Table (7) presents the actual cost and predicted cost for testing facts which are calculated using the developed model. It shows that the percentage of absolute difference of predicted cost (%Error) ranges from 0.7% to 2.3% with average value of 1.8% which is less than the previously mentioned average absolute percentage of error for the training facts (13.1%). Consequently, the model testing was successfully passed and it is valid to be used in cost estimating processes for such type of projects that are containing sterile buildings.

Table 7. Testing results of developed model

Project No.	Actual Cost	Predicted Cost	Absolute Difference	% Error (Testing)
1	35000000	35776871	776871	2.2%
2	11000000	10745603	254396	2.3%
3	5000000	5102213	102213	2.0%
4	80000000	80590843	590843	0.7%

V. Research Summary & Conclusion

This study focus was directed to develop a reliable parametric cost estimating model which can be used in the early stage of the project life cycle. The study effort was concentrated only on pharmaceutical and food projects that enclose such type of sterile buildings, in Egypt.

In order to develop a reliable cost estimating model, the most important cost factors were determined by applying a set of statistical and logical analysis on the collected factors. These analysis were deduced that 13 factor out of 36 are considered the most important cost factors (cost indicators), these factors were; (1) currency exchange rate, (2) consumer price index, (3) desired completion time for the project, (4) accumulative built-up area, (5) accumulative sterile areas, (6) total area of other supplementary buildings (W. tank, gate house,… etc), (7) desired structural system, (8) buildings closeness, (9) project status, (10) project location, (11) target market, (12) international insurances if any, and (13) Desired level of contractors' qualification.

Moreover, the best structure of the model was achieved through trial and error practices.

Finally, the testing process of the developed model clearly shows that the percentage of absolute difference of predicted cost (%Error) ranges from 0.7% to 2.3% with average value of 1.8%. Accordingly, the developed artificial neural network model has been proved itself as reliable management tool for estimating the total construction cost at the early stage of both food and pharmaceutical projects in Egypt.

VI. Recommendations

For future researches, the following potential areas of studies and attempts, if explored, would provide increased validity to the findings of this research:

A. It is recommended that a standard database system for storing information about all completed projects should be developed and applied by the construction companies in Egypt, this attempts will enrich the process of developing any future ANN model.

B. The model should be augmented to take into consideration the other different types of construction projects. For example: the medical, Commercial and administrative construction projects.

References

[1] Arafa, M. and Alqedra, M. (2011). "Early Stage Cost Estimation of Buildings Construction Projects Using Artificial Neural Networks." Journal of Artificial Inelegance, 4(1): PP 63 - 75.

[2] Asal, E. (2014). "Contingency Assessment in Cost, using Artificial Neural Network." M.Sc. Thesis. College Of Engineering and Technology, Arab Academy For Science, Technology And Maritime Transport, Cairo, Egypt.

[3] Attal, A. (2010). "Development of Neural Network Models for Prediction of Highway Construction Cost and Project Duration." M.Sc. Thesis. College of Engineering and Technology, Ohio University.

[4] Godden, B. (2004). "Sample Size and Confidence Interval Tutorial." http://williamgodden.com/samplesizeformula.pdf

[5] Charles, A. T. (2006). "A bottoms-up approach to cost estimation using parametric inputs." M.Sc. Thesis. College Of Engineering and Technology, Ohio University, United states of America.

[6] El-Gafy, M. A. (2001). "Neural Network Model for Parametric Cost Estimation of Sewer Projects." M.Sc. Thesis. Faculty of Engineering, Cairo University, Giza, Egypt.

[7] Hosny, H. (2011). "Contingency Assessment in Time Schedule, using Artificial Neural Network." M.Sc. Thesis. College Of Engineering and Technology, Arab Academy For Science, Technology And Maritime Transport, Cairo, Egypt.

[8] Jamshid, S. (2005). "Cost Estimation of Highway Projects in Developing Countries: Artificial Neural Network Approach." Journal of the Eastern Asia Society for Transportation Studies, 6: PP. 1036 – 1047.

[9] Lekan, A. M. (2011). "Neural Network-Based Cost Predictive Model For Building Works." Ph.D. Thesis. School Of Environmental Sciences, College Of Science And Technology, Covenant University, Ota, Ogun State.

[10] Morcous, G. (1997). "Estimating of Quantities and Cost of P.C. Bridges Over The Nile in Egypt: Conventional and NeuralNetworks Approaches. " M.Sc. Thesis. Faculty of Engineering, Cairo University, Giza, Egypt.

[11] Murat, H. G. and Dogan, S. Z. (2004). "A Neural Network Approach for Early Cost Estimation of Structural Systems of Buildings." International Journal of Project Management, 22(7): PP 595 - 602.

[12] Phaobunjong, K. (2002). "PARAMETRIC COST ESTIMATING MODEL FOR CONCEPTUAL COST ESTIMATING OF BUILDING CONSTRUCTION PROJECTS.". Ph.D. Thesis. Faculty of the Graduate School, University of Texas, Austin.

[13] Setyawati, B. R., Sahirman, S. and Creese, R. C. (2002). "Neural Networks for Cost Estimation." ASCE International Transactions, EST.13: PP 1 - 9.

Mohamed Zahran M.Sc. Student, Building and Construction Dept., Faculty of Engineering, Arab Academy for Science and Technology and Maritime Transport Cairo, Egypt.

Hossam Hosny Professor, Construction Engineering Dept., Faculty of Engineering, Zagazig University Zagazig, Egypt

Abdelmonem Sanad Professor, Building and Construction Dept., Faculty of Engineering, Arab Academy for Science and Technology and Maritime Transport, Cairo, Egypt.