RESEARCH ARTICLE OPEN ACCESS

Conceptual Cost Estimate of Libyan Highway Projects Using Artificial Neural Network

Emad Elbeltagi[1], Ossama Hosny[2], Refaat Abdel-Razek[3] and Atif El-Fitory[4]
[1]Structural Engineering Department, Faculty of Engineering, Mansoura University
[2]Construction and Architectural Engineering Department, The American University in Cairo
[3]Construction Engineering and Utilities Department, Faculty of Engineering, Zagazig University
[4]Graduate Student, Construction Department, Arab Academy for Science and Technology, Cairo

Abstract
It is well known that decisions at early stages of a construction project have great impact on subsequent project performance. Conceptual cost estimate is a challenging task that is done with limited information at the early stages of a project life where many factors affecting the project costs are still unknown. The objective of this paper is to support decision makers in predicting the conceptual cost of highway construction projects in Libya. Initially, the factors that significantly influence highway construction are identified. Then, an artificial neural network model is developed for predicting the cost. The network is trained and tested with a total of 67 projects historical data. Training of the model is administered via back-propagation algorithm. The model is coded ad implemented using MATLAB® to facilitate its use. An optimization module is also added to the Neural Network model with the objective of minimizing the error of the predicted cost. The model is then validated and the results show better predictions of conceptual cost of highway projects in Libya.
Keywords: Construction, Conceptual cost estimating, Neural Networks, Highway projects, Libya.

I. INTRODUCTION

Conceptual cost estimate is one of the most important activities to be performed during the project planning phase. It includes the determination of the project's total costs based only on general early concepts of the project [Kan 2002]. Like all other planning activities, conceptual cost estimating is a challenging task. This is due to the availability of limited information at the early stages of a project where many factors affecting the project costs are still unknown.

Highways construction cost depends on many factors. The identification and selection of those factors that may be used to describe a project and define/affect its cost is an essential task. Such factors must be measurable for any new highway project that is required to estimate its' conceptual cost. Accordingly, it is necessary to determine the cost estimating relationships in terms of the selected factors. A parametric cost estimate is one that uses Cost Estimating Relationships (CERs) and associated mathematical algorithms (or logic) to establish cost estimates [Hegazy and Ayed 1998]. Traditionally, cost estimating relationships are developed by applying regression analysis to historical project information. Regression analysis is a good method of determining the relationship between the project factors and cost, and determining the appropriate mathematical form for the model [Hegazy and Ayed 1998].

Several efforts have been done to develop models for parametric cost estimation using traditional techniques. Trost and Oberlender [2003] developed a model named the Estimate Score Procedure to enable the project team to make a score for an estimate and predict its accuracy based on that score. Forty Five drivers were selected to measure the accuracy of the cost estimate. Using factor analysis and multivariate regression analysis these factors were grouped into eleven orthogonal factors. A computer model was developed to automate the procedure. Five of the eleven factors were identified by the multivariate regression analysis to be significant. David et al. [2006] developed a linear regression model to predict the construction cost of buildings and concluded that the best regression model gives a mean absolute percentage error of 19.3%. Martin and Thomas [2002] developed a method for calculating the variance of total project cost based on standardized component costs for a set of database for Public School projects.

Artificial Neural Networks (ANN), as one of the artificial intelligence techniques, has been extensively used for cost estimate. ANN presents itself as an approach of computation and decision making that may potentially resolve some of the major drawbacks of traditional estimating techniques. It holds a great promise for rendering the parametric method of cost estimating which is a reliable and reasonably accurate way to prepare cost estimates [Ayed 1997].

Several studies in the literature have used ANNs for parametric cost estimate in Highway construction. Adeli and Wu [1998] formulated a regularization ANN model and presented neural network architecture to estimate highway construction costs. The model was applied on reinforced concrete pavements. It was based on slid mathematical formulation that made the estimate more reliable and predictable. Hegazy and Ayed [1998] also developed an ANN for parametric cost estimation of highway projects. They used spreadsheet program as a media for implementing their developed ANN model. On another research effort, an ANN model was developed to estimate the percentage increase in the cost of a typical highway project from a baseline reference estimate [Al-Tabtabai et al. 1999]. In this study, they considered several factors such as environmental, company specific and other factors that may affect the increase in cost. The model measured the combined effect of those factors and the percentage change on expected cost. Georgy and Barsoum [2005] developed an ANN model for parametric cost estimation of school construction projects in Egypt. They employed both statistical methods and ANNs for estimating construction costs and concluded that a single 3-layer ANN that has number of neurons in the hidden layer equal to two thirds of the number of neurons in the input layer, produce optimum results. Hosny [2006] developed an ANN model for predicting increase in time and cost of construction projects in Egypt based on several influential factors such as: project type, contract type, owner behavior, design completeness, cost/time rate and others. Gaber et al. [1992] developed an ANN for assessing the risk in industrial projects. Several factors were considered and results showed that the potential benefit of using ANN in accessing risks for industrial projects.

The motivation for this study is the limited research in the area of conceptual cost estimating, especially for the highway construction industry in Libya, and the need for a better conceptual cost estimating methodology and tools. The paper aims at developing a neural network model for predicting the conceptual cost for highway projects in Libya. The details of the model development are presented along with its implementation procedures.

II. RESEARCH METHODOLOGY

In order to achieve the objective of this study, a three-step research methodology is adopted: First, identifying the main factors that affect cost estimating of highway projects in Libya. This is achieved through reviewing the literature related to highway projects. Then, developing a questionnaire survey to identify the most important factors on the conceptual cost estimate of highway projects in Libya; second, collecting relevant data corresponding to the important factors identified in the previous step from previously completed highway projects; third, designing an ANN model for estimating the conceptual cost in the early stages of highway projects. The developed model is then tested for obtaining the best-possible network configuration. These steps are presented in the following sections.

III. FACTORS IDENTIFICATION AND DATA COLLECTION

3.1 Questionnaire Survey

Reviewing the literature revealed a list of 18 factors to be considered as the most influential on parametric cost estimate of highway projects [Hegazy and Ayed 1998, Al-Tabtabai et al. 1999, Wilmot and Bing 2005, Jui et al. 2005, and Nassar et al. 2005]. Having identified these factors as shown in Table 1, a questionnaire survey is designed to rank these factors based on their effect on parametric cost estimate of Libyan highway projects. The 18 factors were grouped into three categories: project-specific factors, project participants' factors, and environmental factors.

The designed questionnaire was distributed among 90 Architectural/Engineering (A/E) firms. These A/E firms are authorized from the Libyan Engineers Syndicate (85 private + 5 public) working in the western district of Libya. The questionnaire was also given to the secretarial of housing and utilities, the secretarial of transportations and communications, and the chairman of the savings and real-estate investment bank. The questionnaire Participants were from different parties, including: owners, general managers, contract administrators, project managers, financial managers and cost estimators working on Libyan highway industry. Agencies such as, secretarial of housing and utilities, secretarial of transportations and communications are the main supervisor of the highway construction in Libya. Participants are asked to indicate the importance of each factor, which should be considered during preparing preliminary estimate.

Each participant was asked to give a weight from zero to 100 for each factor based on its influence on the preliminary cost estimate of highway construction. A weight of zero means that the corresponding factor has no effect while a weight of 100 means that the factor extremely affects the cost. Participants were also asked to suggest other factors that are not listed in the questionnaire, and may affect the preliminary cost of the project.

3.2 Questionnaire Analysis

Among the 90-surveys that were sent to participants only 61-survey were returned, giving a response rate of approximately 68%. The participants' responses received were tabulated and analyzed individually. The average experience for all respondents is 15 years, and the average annual work size they were involved in is 5,000,000 Libyan Dinar (LYD). The survey showed that a conceptual estimate is prepared based on experience and there is no tool used to assist estimators performing their job.

No.	Factor description	Importance index	Considered factors
1	Project type (freeway, artery, local, others)	94.26	*
2	Construction of detours (Many detours difficult to build, few & easy to build, normal, none)	91.39	*
3	Project location	81.56	*
4	Year of project construction	78.28	*
5	Project scope (new, rehabilitate, others)	76.23	*
6	Size of project (Length in Km)	73.36	*
7	Project capacity (1-lane, 2-lanes, 2-lanes divided, others)	72.95	*
8	Project duration (/day)	67.21	*
9	Construction season (winter, summer, fall, spring)	61.07	*
10	Soil Type (Rocky, mixed, clay, sand, others)	57.38	*
11	Financial condition (Highly unfavorable, Unfavorable [no regular payments], Favorable [regular payments], Highly favorable (reg. Pay.& good finance)	55.66	*
12	Hauling distance (i.e., transporting material & equipment)	44.67	
13	Pavement thickness	40.98	
14	Water body	40.16	
15	Preservation of utilities	39.75	
16	Cost escalation (i.e., Rate of inflation)	25.82	
17	Type of consultant (Performance, degree of cooperation)	23.77	
18	Contractor performance/Interest	22.13	

Table 1: Factors Affecting Construction Cost of Highway Projects in Libya

As shown in Fig. 1, 39% and 28% of the respondents are senior personnel and managers respectively. Their experience was reflected in the level of completeness, consistency and precision of the information provided, which provides further validity for the survey results.

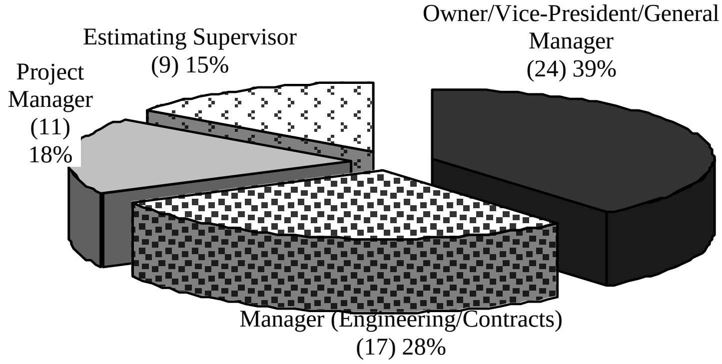

Figure 1: Respondents Classification

The importance index (reflects the effect of a specific factor on cost estimate) is calculated according to the rating scale of each factor which is estimated by each respondent according to their experience as stated in Eq. 1. The importance index for each factor is calculated as [Dutta 2006]:

$$\text{ImportanceIndex} = \left(\sum_i v_i * f_i\right)/n \quad [1]$$

Where: v_i = rating of each factors (0 - 100%).
f_i = frequency of responses.
n = total number of responses.

The eighteen factors mentioned above were ranked according to their importance to construction cost of highway projects in Libya as shown in Table 1. Factors with Importance Index smaller than 50% were omitted. Accordingly, eleven factors were considered as the most influential factors for cost estimate of Libyan highway projects. These factors are marked with asterisks as shown in Table 1.

3.3 Data Collection

Having identified the main factors affecting cost estimate in Libyan highway industry, the next step is to collect data relevant to these factors from previous highway projects constructed in Libya. Accordingly, another questionnaire is prepared to collect data relevant to these eleven factors. The questionnaire is distributed to 125 contractors. Personal contact was the major communication tool used to get contractors' organization to participate in this study. The interviewees were mostly construction managers and project administrators. A total of 91 contractors have responded and participated in this research which represents almost a 73% response rate.

Data were collected from 91 projects constructed during the period from year 1989 to year 2006. Figure 2 shows the distribution of the 91 projects. It is noticed that 68% of the projects were constructed during the period from year 2002 to year 2005. The 91 projects are classified as: 13% freeways, 25% rural roads, 34% artery, and 28% local roads. By examining the collected data of the 91 projects, it is found that some projects have some data missed and others have similar data. Accordingly, twenty-four projects were removed, leaving only 67 projects for further use by the ANN. Most of these projects are constructed in three major areas in Libya namely: El-Zawia, Sorman, and Abo-Esa.

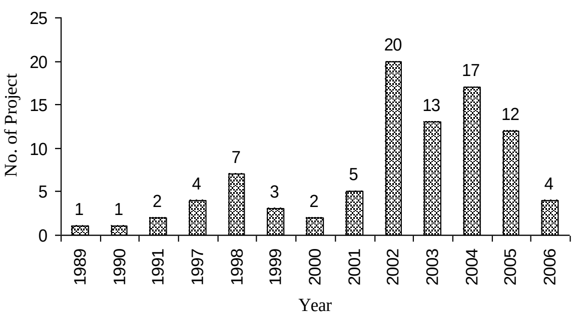

Figure 2: Distribution of the 91 Projects over the 1989-2006 Period

3.4 Time and Location Adjustment of Collected Data

One of the collected data for the 91 projects is the project actual cost. As this cost may be affected by the project location or the construction year, therefore, this cost is adjusted for both the time and location for consistency. Also, when estimating the cost of a new project, it is necessary to adjust the predicted cost according to the location and time for the project.

Adjustment for time represents the relative inflation or deflation of costs with respect to time due to factors such as labor rates, material costs, interest rates, etc. [Peurifoy and Oberlender 2002]. The inflation rate for seven years from year 2001 to year 2007 published by the Libyan National Authority for Information are used to adjust the cost of the projects data and later on the predicted cost of a new project. As 68% of the projects were constructed during the period from year 2002 to year 2005, the average inflation rate were calculated and used to adjust all the cost data collected to the year 2005. The unit cost at any year n is calculated as follows:

$$Unit\ cost\ \left(LYD/m^2\right) = Given\ unit\ cost \times \left(1+i\right)^n \qquad [2]$$

Where: *i* is the average inflation rate for the period (from year 2005 to the current year); *n* is the number of years from 2005 to the current year.

Similarly, an adjustment is needed for the difference in location of the used project. The adjustment represents the relative difference in costs of materials, equipment, and labor with respect to the two locations [Peurifoy and Oberlender 2002]. The Secretarial of Housing and Utilities and Secretarial of Transportations which are the main owners of the highway projects in Libya categorized Libya to four areas. This categorization is based on the political circumstances and construction cost. The two secretariats allow a bonus ratio that differs for each of the four areas. The bonus ratio is the extra amount that can be added to the construction cost. Using this bonus ratio, all the collected cost data are scaled back to El-Zawia zone (42.3 % of the projects were constructed at El-Zawia zone). The highway unit cost for any location is calculated as follows:

$$\begin{aligned} &Unit\ cost\ in\ a\ given\ city\ \left(LYD/m^2\right) = \\ &Actual\ unit\ cost\ for\ that\ city \times \frac{A}{B} \end{aligned} \qquad [3]$$

Where: *A* is the bonus ratio for a given city; *B* is the bonus ratio for El-Zawia (Since it the base city for the model).

IV. NEURAL NETWORK MODEL FOR PARAMETRIC COST STIMATING

Having identified the main factors that affect the cost estimate of highway projects in Libya and collected the relevant data, an ANN model is developed that uses these factors as the inputs and the adjusted cost as the output. The major strength of ANNs is their ability to learn from examples and to generalize that knowledge to novel cases. Using an appropriately configured ANN model and a sufficient set of past completed highway projects, an ANN model would be able to arrive at accurate forecasts of the cost of a new construction highway project. Based on this concept, the development of the ANN model proceeded in four phases: design phase, preliminary implementation phase, refinement phase, and validation phase. The design phase addresses the preliminary planning of the ANN model and its variable representation. The preliminary implementation phase entails the preliminary configuration and training of the ANN model and testing for proper functioning. Then, the refinement phase concerns the changing of the model parameters for improved performance and increased accuracy in cost prediction. Finally, the validation phase addresses the final validation of the developed model.

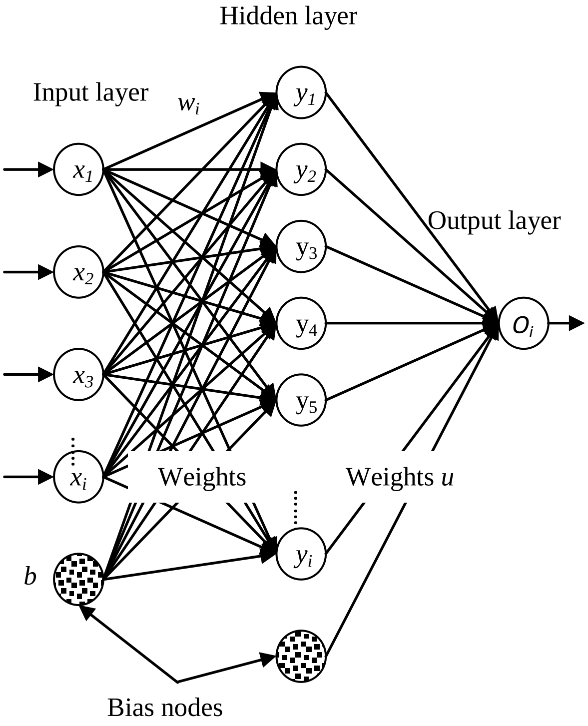

Figure 3: Typical Structure of Multi-Layer ANN

4.1 Design Phase

The first step of designing the ANN model is to identify the data representation scheme for each input and output variable, as shown in Table 1. This is entirely dependent on the nature of each factor. Some factors, such as project size and project duration, have numerical nature and thus represented by their corresponding numerical values. Other factors have linguistic or non-numerical nature, such as project type, project location, and so forth. These factors are represented through equivalent numerical values. For example, the project type has been represented using numerical values as: Highways (1), rural roads (2), artery roads (3) and local roads (4). Configuring the ANN is a complex and dynamic process that requires the determination of the internal structure and rules (i.e., network architecture, learning algorithm, the number of hidden layers and neurons in each layer…etc.). In this study, the multilayer feed-forward back-propagation neural networks are utilized as they are suitable for modeling the nonlinear mapping type of problems [Hegazy and Ayed 1998]. Figure (3) shows a typical architecture of multilayer feed-forward neural networks with an input layer, an output layer, and one hidden layer.

As shown in Fig. (3), the artificial neurons are arranged in layers, and all the neurons in each layer have connections to all the neurons in the next layer. Associated with each connection between these artificial neurons, a weight value (w_i) so that the total of i inputs (x_i) to the single neuron is:

$$input = \sum_i w_i \, x_i + b \qquad [4]$$

Where: b is the connection weight associated with a bias node having input value = 1.

This input passes through an activation function to produce the values of y_i of the hidden layer(s) or O_i of the output layer. The activation function may have many forms. The most familiar and effective form is the sigmoid function [Seleemah 2005], defined as:

$$Output = \frac{1}{1 + exp^{-\alpha(\text{input})}} \qquad [5]$$

Where: α is a constant that typically varies between 0.01 and 1.00.

Signals are received at the input layer, pass through the hidden layers, and reach the output layer, producing the output of the network. The learning process primarily involves the determination of connection weights and bias matrices and the pattern of connections. It is through the presentation of examples, or training cases, and application of the learning rule that the neural network obtains the relationship embedded in the data.

In this study the neural networks were designed to have an input layer that consists of eleven input nodes representing the most important factors that affect the cost estimate of highway projects. These factors are: project type, construction of detours, project location, year of project construction, project scope, project size, project capacity, project duration, construction season, soil type and financial condition. The output layer consists of one node representing the unit cost of a highway project construction.

An important factor that can significantly influence the ability of a network to learn and generalize is the number of patterns in the training and testing sets. Although it takes longer time to train a network, using higher number of training patterns increases the ability of the network to learn and achieve more accurate results. Günaydın and Doğan [2004] state that there are no acceptable generalized rules to determine the size of the training data for suitable training. However, having a total of 67 data patterns; it was decided to use 75% of the data (50 projects) for training, 15% of the data (10 projects) for testing, and 10% (seven projects) for validation. These sets were randomly selected and extracted from the data.

4.2 Input and Output Data Normalization

Data are generally normalized for effective training of the model being developed. The normalization of the data is the scaling of the input and output pairs within the range (-1, 1) or the range (0, 1) depending on the processing function. It is used to allow the squashing of the values to improve the network performance [Hegazy et al. 1994]. Furthermore, the neural networks usually provide improved performance when the data lie within the range (0, 1) [Seleemah 2005]. Because a sigmoid function is used, a slow rate of learning occurs near the end points of the sigmoid function. To avoid this, all input values and associated outputs for this study are transformed to values within the range (-1, 1) by using the (Max Min processing) method. The Max Min processing function used to modify the data as follow [Hegazy and Ayed 1998]:

$$Scaled\ value = \left[\frac{2\times(Original value - Min value)}{(Max.\ value - Min valu)}\right] - 1 \qquad [6]$$

Where, *Max. value* and *Min. value* are the maximum and minimum data values within a specific data set.

4.3 ANN Structure

One important issue of multilayer feed-forward ANNs is to determine the appropriate number of hidden layers and the number of neurons in each layer. This process is done through trial and error. There is no unique solution for representation schemes. Different ANNs can produce similar results with the same set of training data [Seleemah 2005]. Hegazy et al. (1994) heuristically suggest that the number of hidden nodes may be set as one-half of the total input and output nodes. Sodikov and Student (2005), on the other hand, suggest that the minimum number of hidden nodes can be more than or equal to (p-1)/(n+2), where p is the number of the training examples and n is the number of the inputs of the networks. In their study, an iterative process is used of increasing the number of nodes in one and two hidden layers till the network reaches its desired performance.

Also, it is important to determine which training procedure to adopt. There are many alternative paradigms to choose from. The back propagation algorithm which belongs to the realm of supervised learning rule is the most widely used training technique for problems similar to the current study [Günaydın and Doğan 2004, Seleemah 2005 and Hegazy and Ayed 1998]. Accordingly, the back-propagation learning algorithm (supervised training) is used to perform the training requirements and to construct the current model.

4.4 ANN Training and Refinement

While an initial ANN model for parametric cost estimate of highway projects has been developed, this model needs refinement to reach the optimum performance. For a systematic implementation of the refinement phase, a parametric analysis of several ANN parameters is conducted. This include parameters such as training function, learning function, number of hidden layers, and number of nodes in hidden layer. The optimum performance presumably provides minimum error for the resulting output. This is calculated through the mean absolute percentage error (MAPE) as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Actual_i - Predicted_i\right|}{Predicted_i}\times 100 \qquad [7]$$

Where: i is the project number; n: is the total number of training or testing data set; $Actual_i$: is the actual unit cost. $Predicted_i$: is the unit cost obtained from the ANN.

The MAPE of each group of training cases is called "Training", and the MAPE for testing cases is called "Testing". Then, the average MAPE for training and testing is called "Average".

Different transfer functions (Tansig, Logsig, Purelin), different learning function (TRAINBFG, TRAINCGB, TRAINLM, TRAINCGF, TRAINCGP, TRAINGDM, TRAINGDA, ….etc.) and different network structures (ANN 11-5-1, ANN 11-7-1, ANN 11-5-6-1, … etc.) have been experimented with. A network labeled ANN 11-5-1 means that the network has an input layer of 11 neurons, one hidden layer with 5 neurons and an output layer of one neuron. Also, an ANN 11-5-6-1 means that the network has an input layer of 11 neurons, two hidden layers containing 5 and 6 neurons in layers 1 and 2, respectively and an output layer of one neuron. The results of these experiments are illustrated in Figs. 4, 5 and 6.

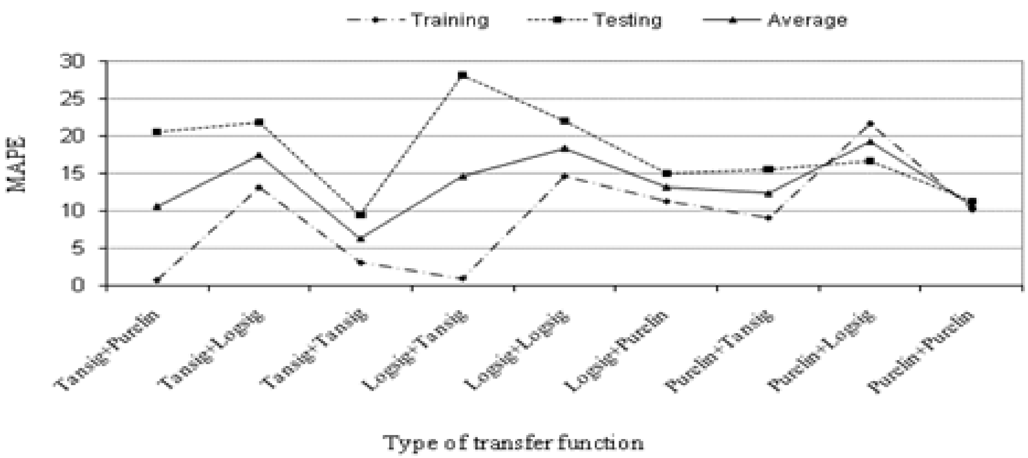

Figure 4: The MAPE Using Different Transfer Function

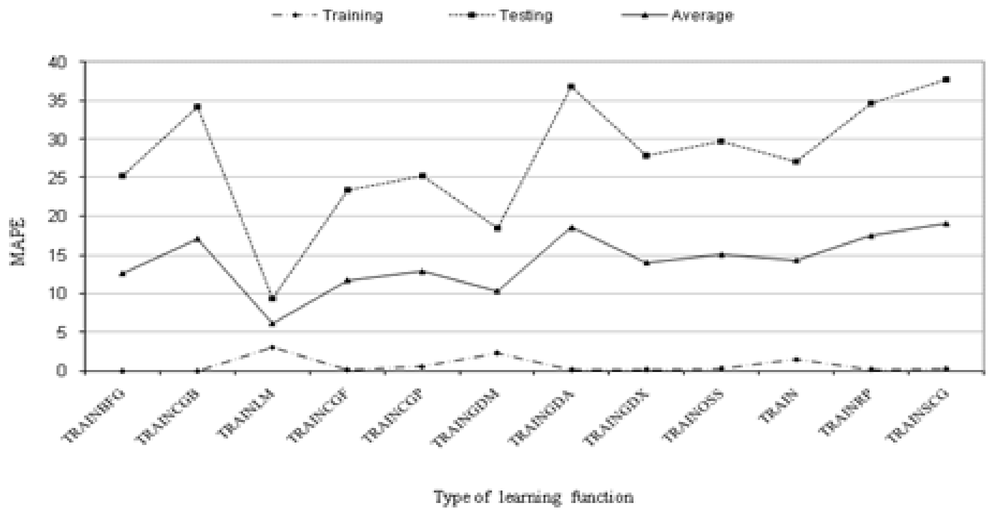

Figure 5: The MAPE Using Different Learning Function

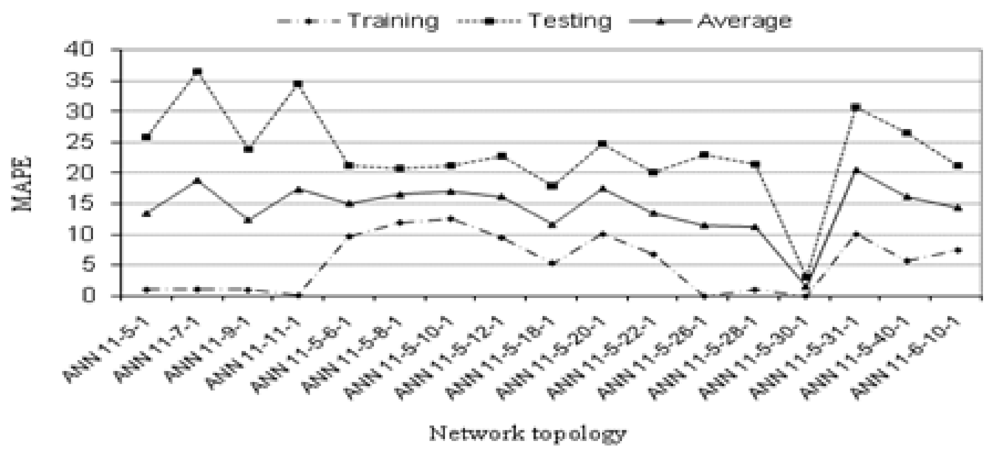

Figure 6: The MAPE Using Different Network Topology

Based on the parametric analysis, the following set of conclusions is found regarding the performance of the ANN model:

- Based on the calculated MAPE for several transfer functions such as: TANSIG, LOGSIG, and PURELIN as shows in Fig. 4, It is found that the "TANSIG" transfer function for both hidden and output layers produced a minimum MAPE for training phase of 3.1%, for testing phase of 9.32%, and a total average error of 6.21%. Accordingly, the "TANSIG" transfer function is used during this study.

- Based on the calculated MAPE for several training functions (such as: TRAINBFG, TRAINCGB, TRAINGDM, and Levenberg-Marquardt (TRAINLM) as shows in Fig. 5), the "TRAINLM" function produced minimum MAPE of 6.21%. Therefore the "TRAINLM" algorithm is selected as the training function of the ANN model in the current study.
- The best performance is achieved with a network consisting of: one input layer with 11 neurons; two hidden layers, the first with 5 neurons and the second with 30 neurons and an output layer with one neuron, as shows in Fig. 6.

Accordingly, the ANN is trained using 50 data sets (projects), and then tested using 10 data sets (projects). The calculated average MAPE is 0.034% for the training phase, 3.058% for testing phase and 1.55% total average error, which represents an acceptable error during the preliminary estimating phase.

V. IMPLEMENTATION AND VALIDATION

To facilitate the implementation process and the use of the developed ANN model, a MATLAB® program of multilayer neural network with one output node is coded in a MATLAB® File. For further system usability, a user-friendly interface for the Parametric Cost Estimating (PCE) program (Fig. 7) is developed using MATLAB® graphical user interface (GUI). This user interface is the way that the program accepts instruction from user and presents results. One important aspect of a practical estimating system is to adapt it to new project situations. This enables it to adjust its nature to become more suited to the user's own work environment. It also enables the buildup of experience and incorporates new encounters into the model. Initially, the user's historical project data is entered into the PCE by clicking the "Add" button. The user will be prompted to enter new project data using the screen shown in Fig. (8). A module is developed to transform the project data into numerical data. When all data are entered, the number of each project data can be seen in the top-left pane of Fig. 7, where the user can select any project to view, modify or delete. The user may view or modify the data that was already entered before by right clicking on any project number; then the "Project Information" screen will pop up (Fig. 8) and can be used to modify the data for the selected project [El-Fitory 2008].

The "Re-Train" button becomes active after modifying, adding or deleting any project. By clicking the "Re-Train" button, the weight matrices parameters are adjusted automatically according to the new situation. The results are presented to the user upon the completion of the training process on the top-right pane as shown in Fig. (7). Once the training process ends, the user can test the model using the specified data by pressing the “Test” button. Afterwards, the developed system could be used to predict the cost estimate for a new project. By selecting the "Project Estimation" button, the "Project Information" screen (Fig. 8) pops-up for entering the data of the new project. Once the data for the new project is entered, the user will be prompted to enter the inflation rate from the year 2005 to the current year and to enter the bonus ration for the new location. The inflation rate for time and the bonus ratio for location adjustments are also integrated within the developed program to adjust the predicted unit cost for both time and location as described earlier. The future unit cost estimate for future time after n years from the year 2005 and for location A, other than El-Zawia location, B, is calculated as follows (i is the average inflation rate from year 2005 to the current year) [El-Fitory 2008]:

$$Future\,unit\,cost\left(LYD/m^2\right) = Predicted\,unit\,cost \times (1+i)^n \times \frac{A}{B} \quad [8]$$

5.1 ANN Optimization

One important feature of the current development is its ability to optimize the results to get the minimum MAPE. The user has the ability to shuffle the data by changing its arrangement in order to minimize the MAPE. The user may click the "Shuffle" button to shuffle all rows of data placement, and then the “Re-train” button is used and observes the value of MAPE for training and testing. This step must be repeated many times to get the optimum row arrangement for the projects. This process may be very lengthy; therefore, the optimization option is added. By clicking the "Optimize" button (Fig. 7), the program starts automatically doing multiple series of shuffling, training, and testing in order to produce better results. The user can manually stop this process when he/she observes the improvement of the MAPE values shown at the top-right pane of Fig. 7. This process stops automatically when there is no more improvement in the MAPE.

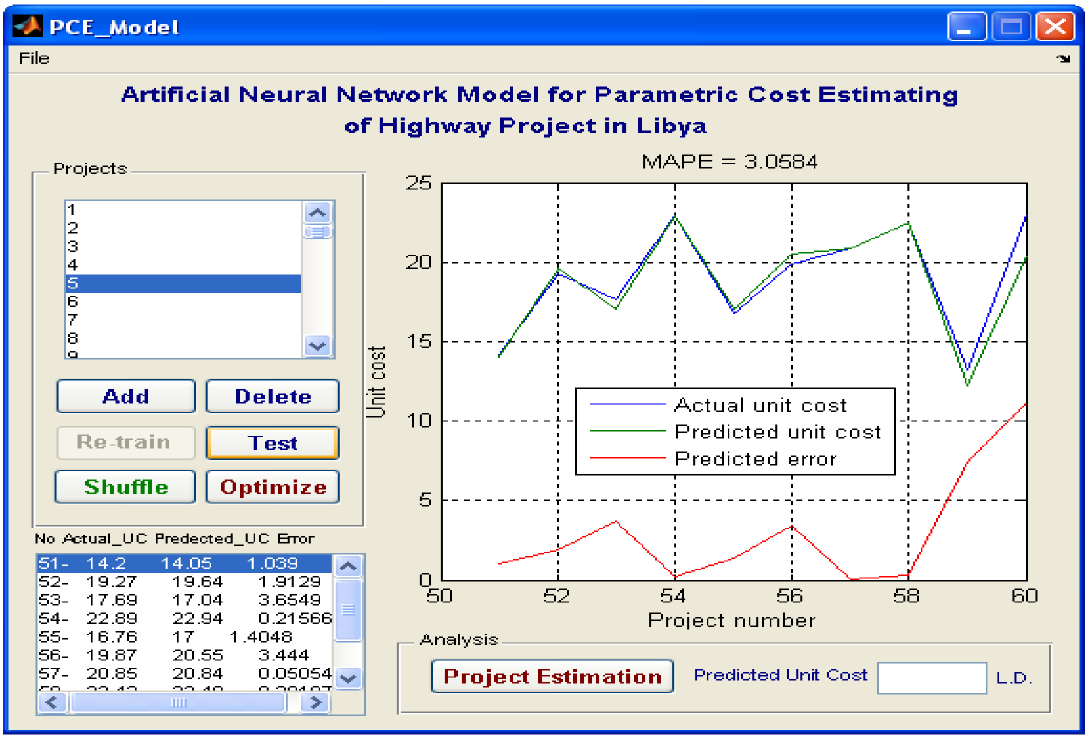

Figure7: Parametric Cost Estimating Program User Interface

Table 2: Percentage Error between Actual Unit Cost and Predicted Unit Cost

Project number	Actual unit cost (LYD/m^2)	Pred. unit cots (LYD/m^2)	Percentage of error
1	19.27	19.65	1.91
2	14.02	14.31	2.02
3	22.10	22.80	3.05
4	18.67	18.62	0.29
5	21.87	24.04	9.02
6	24.17	24.20	0.11
7	18.94	19.65	3.62
	Average error		2.86%

5.2 ANN Model Validation

One of the most important steps in developing a parametric cost model is to verify its accuracy and validity [Dysert 2001]. The validation data should not be used in the training and testing of the model. For this purpose, seven projects are extracted from the 67 projects for validation. Accordingly, the ANN model is used to forecast the unit cost of these seven highway project based on predictions of the model output given the input validation data. The predicted unit costs computed are then compared to the actual unit costs recorded by calculating the estimated error to measure the performance of the network as follows:

$$Estimated\,error(\%) = \frac{(Actual\,unit\,cost - Predicted\,unit\,cost)}{Predicted\,unit\,cost} \times 100 \qquad [9]$$

The results show (Table 2) a good performance of the developed ANN model where the calculated MAPE is 2.86 %, which is less than 20% the typical expected percent error for conceptual cost estimate [Peurifoy and Oberlender 2002]. This validation result proves the validity of the proposed model and its ability to predict the unit cost of highway project in Libya.

VI. CONCLUSIONS

The work presented in this paper aimed to develop an accurate and practical method for conceptual cost estimating that can be used by organizations involved in the planning and execution of highway construction projects in Libya. The research identified eleven factors that significantly influence the cost of constructing highway projects.

The data used for the model development and validation were based on historical project data collected from 67 completed highway projects, constructed from year 2001 to year 2005 in Libya. The ANN model was designed with eleven neurons in the input layer. The output layer consists of one neuron representing the unit cost of highway construction project per square meter. The results obtained from the ANN model with two layers containing 5 neurons, and 30 neurons on the first and second layers, respectively, were consistent and gave values of predicted unit cost very close to the actual unit cost. The MAPE for the training phase is 0.034%, for the testing phase is 3.058% with total average error is 1.55%. The developed system also can be used to optimize the prediction of the ANN model by shuffling the input data and monitoring the improvements in the MAPE values. The model was validated and the results signify that it has captured the relations embedded in the trained data and in turn indicated better cost prediction of highway projects with a MAPE of 2.86%. In addition, the model provides a methodology to account for inflation and location adjustments.

Fig. 8: Adding New Project Data Screen

REFERENCES

[1] Adeli, H., and Wu, M. (1998), "Regularization neural network for construction cost estimation," Journal of Construction Engineering and Management, Vol. 124, No. 1, PP. 18-24.

[2] Al-Tabtabai, H., Alex, P., and Tantash, M. (1999), "Preliminary Cost Estimation of Highway Construction Using Neural Networks. Journal of Cost Engineering, Vol. 41, No. 3, PP. 19-24.

[3] Ayed, A.S. (1997), *Parametric Cost Estimating of Highway Projects using Neural Networks*, M. Sc. Thesis, Faculty of engineering and Applied Sciences, Newfoundland, Canada.

[4] David, J.L., Margaret, W.E. and Anthony, H. (2006), "Predicting Construction Cost Using Multiple Regression Techniques," Journal of Construction Engineering and Management, Vol. 132, No. 7, PP. 750-758.

[5] Dutta, D. (2006), *Statistics and Operations Research*, Second Edition, Laxmi Publications (P) LTD, Daryaganj, New Delhi.

[6] Dysert, L.R. (2001), "Developing a Parametric Model for Estimating Process Control Costs," International Journal of Cost Engineering, Vol. 43, No. 2, PP. 31-34.

[7] El-Fitory, A. (2008), *Predicting Conceptual cost of Libyan Highway Projects using Artificial Neural Networks*, M.Sc. thesis, Faculty of Engineering, Arab Academy for Science, Technology and Maritime Transport, Cairo Egypt.

[8] Gaber, M., Rabelo, L. and Hosny, O. (1999), "An Integrated Artificial Neural Networks Model for Industrial Projects' Risk Assessment," IEEE and Engineering Management Conference, Virginia, October 1992.

[9] Georgy, M.E., and Barsoum, H.S. (2005), "Neural Networks System for Budget Estimating of School Construction Projects," Eleventh International Colloquium of Structural and Geotechnical Engineering, (ICSGE), May 17-19, Cairo, PP. MG 14-1-MG 14-10.

[10] Günaydın, H.M., and Doğan, S.Z. (2004), "A Neural Network Approach for Early Cost Estimation of Structural Systems of Buildings," International Journal of Project Management, Vol. 22, PP. 595–602.

[11] Hegazy, T., Moselhi O., and Fazio, P. (1994), "Developing Practical Neural Network Application Using Backpropagation," Journal of Microcomputers in Civil Engineering, Vol. 9, No. 2, PP. 145-159.

[12] Hegazy, T. and Ayed, A. (1998), "Neural Network Model for Parametric Cost Estimation of Highway Projects", Journal of Construction Engineering and Management, ASCE, Vol. 124, No. 3, PP. 210-218.

[13] Hosny, O. (2006). "An Artificial Neural Network Approach for Predicting the Effect

of Claims on Project Time and Cost," Engineering Research Journal, Helwan University, Vol. 106, PP. C87-C106.

[14] Jui, S.C., Linlin, W., Wai, K.C., and James, T.O. (2005), "Preliminary Cost Estimates Using Probabilistic Simulation for Highway Bridge Replacement Projects," Proceedings Construction Research Congresss, ASCE, PP. 1-10.

[15] Kan, P. (2002), *Parametric Cost Estimating Model for Conceptual Cost Estimating of Building Construction Projects*, Ph.D. Thesis. University of Texas, Austin, USA.

[16] Martin, S. and Thomas, S. (2002), "Analytical and Approximate Variance of Total Project Cost," Journal of Construction Engineering and Management, Vol. 128, No. 5, PP. 456-460.

[17] Nassar, K.M., Nassar, W.M., and Hegab, M.Y. (2005), "Evaluating Cost Overruns of Asphalt Paving Project Using Statistical Process Control Methods," Journal of Construction Engineering and Management, Vol. 131, No. 11, PP. 1173-1178.

[18] Peurifoy, R.L. and Oberlender, G.D. (2002), *Estimating Construction Costs*, 5th Edition, McGraw-Hill, New York.

[19] Seleemah, A.A. (2005), "A Neural Network Model for Predicting Maximum Shear Capacity of Concrete Beams without Transverse Reinforcement," Canadian Journal of Civil Engineering, Vol. 32, PP. 644–657.

[20] Sodikov J. and Student, G. (2005), "Cost Estimation of Highway Projects in Developing Countries: Artificial Neural Network Approach," Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, PP. 1036 - 1047.

[21] Trost, M. and Oberlender, G. (2003), "Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Regression", Journal of Construction Engineering and Management, Vol. 129, No. 2, PP. 198-204.

[22] Wilmot C.G., and Bing, M. (2005), "Neural Network Modeling of Highway Construction Costs," Journal of Construction Engineering and Management, Vol. 131, No. 7, PP. 765-771.