

# A Neural Network Model Based on Support Vector Machine for Conceptual Cost Estimation in Construction Projects

Behnam Vahdani<sup>a,\*</sup>, Seyed Meysam Mousavi<sup>b</sup>, Morteza Mousakhani<sup>c</sup>, Mani Sharifi<sup>d</sup>,  
Hassan Hashemi<sup>e</sup>

<sup>a</sup>*Instructor, Industrial engineering research center, Qazvin Branch, Islamic Azad University, Qazvin, Iran*

<sup>b</sup>*Ph.D. Student, Young Researches Club, South Tehran Branch, Islamic Azad University, Tehran, Iran*

<sup>c</sup>*Associate Professor, Department Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran*

<sup>d</sup>*Assistant Professor, Faculty of Industrial and Mechanical Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran*

<sup>e</sup>*M.Sc, Young Researches Club, South Tehran Branch, Islamic Azad University, Tehran, Iran*

Received 12 August, 2011; Revised 19 November, 2011; Accepted 22 December, 2011

---

## Abstract

Estimation of the conceptual costs in construction projects can be regarded as an important issue in feasibility studies. This estimation has a major impact on the success of construction projects. Indeed, this estimation supports the required information that can be employed in cost management and budgeting of these projects. The purpose of this paper is to introduce an intelligent model to improve the conceptual cost accuracy during the early phases of the life cycle of projects in construction industry. A computationally efficient model, namely support vector machine model, is developed to estimate the conceptual costs of construction projects. The proposed neural network model is trained by a cross validation technique in order to produce the reliable estimations. To demonstrate the performance of the proposed model, two powerful intelligent techniques, namely nonlinear regression and back-propagation neural networks (BPNNs), are provided. Their results are compared on the basis of the available dataset from the related literature in construction industry. The computational results illustrate that the presented intelligent model performs better than the other two powerful techniques.

*Keywords:* Construction projects; Conceptual cost estimation; Support vector machine; Cross validation.

---

## 1. Introduction

Conceptual cost estimation has major impacts on planning, design, cost management and budgeting in construction projects. Accurate estimation assists planners and experts to assess the project feasibility and effectively control the costs during its life cycle (Cheng et al., 2010). The conceptual cost estimation is experience-oriented. In the feasibility studies, experts can only predict the construction cost based on preliminary design. Under inappropriate conditions, experts refer to historical cases, and then make a decision for the conceptual cost based on their knowledge and experience. However, the construction cost is determined by considering numerous factors, for instance, geological property and earthquake impact. Analyzing the conceptual cost by using a simple linear form cannot predict it precisely because of such complicated assessment process and nonlinear nature in construction projects (Cheng and Wu, 2005).

In the related literature, there are some studies to estimate the conceptual cost of construction projects. Particularly, in the last decade, neural networks (NNs) as the main techniques are extensively applied in this field (Chen and Zhai, 2002). Williams (1994) utilized the NNs to predict changes in the construction cost index and showed that the back-propagation neural network (BPNN) model could not accurately predict the cost index movement because of the complexities involved. Hsieh (2002) presented the evolutionary fuzzy neural inference model (EFNIM) to provide an evolutionary construction conceptual cost estimation model. In this model, genetic algorithms are applied for optimization and fuzzy logic for representing the uncertainty and approximate reasoning. Cheng and Wu (2005) utilized an artificial intelligent technique, namely support vector machines (SVMs), to perform the prediction of construction conceptual cost. An et al. (2007) presented the SVM

---

\* Corresponding author E-mail address: [b.vahdani@gmail.com](mailto:b.vahdani@gmail.com)

model for evaluating the quality of conceptual cost estimates, and investigated the application of the model in construction areas. Kong et al. (2008) proposed the intelligent approach based on the SVM to forecast the construction project cost. The results illustrated that the prediction accuracy of the least square SVM was better than the NN. Cheng et al. (2010) presented an artificial intelligence approach, the evolutionary fuzzy hybrid neural network (EFHNN), to improve the accuracy of the conceptual cost prediction in construction industry. This approach hybridized the NN and high order neural networks into a hybrid neural network.

In this paper, a new intelligent model based on a SVM model is adopted for construction projects in order to improve the decision making in feasibility studies. The proposed model can be successfully utilized for long term prediction of the conceptual costs in construction industry. A cross validation technique is also applied to the training dataset in the SVM model not only to avoid over fitting but also to produce the reliable results. This proposed intelligent model is validated by using the available dataset from the related literature in construction industry for estimating the costs of these projects. In addition, comparisons are made among the proposed model and other famous techniques to illustrate the applicability and suitability.

The rest of the paper is organized as follows. In section 2, the nonlinear regression problem is defined and the proposed intelligent model is introduced. The applicability of the proposed model is examined by using the available dataset provided in the building projects in section 3. In addition, this paper compares the estimation accuracy and required effort of the proposed model with other famous techniques, namely nonlinear regression and BPNN. Finally, in section 4, concluding remarks and future research are provided.

## 2. Proposed Intelligent Model

Prediction of cost time series data produced by using linear techniques cannot often be performed accurately. Real applications in construction industry are not amenable to linear estimation techniques (Hsu et al., 2003; Mousavi and Iranmanesh, 2011; Sapankevych and Sankar, 2009; Tavakkoli-Moghaddam et al., 2011; Vahdani et al., 2011). Indeed, cost data applications in construction projects are complicated and nonlinear in nature. Hence, traditional linear estimation techniques are not applicable; it is necessary to develop advanced approaches such as artificial intelligence for estimating time series data in these projects. To overcome the shortcoming of the commonly-used techniques, this paper presents on the cost estimation of these projects by introducing a new intelligent model based on two powerful techniques as follows.

### 2.1. Support Vector Machine

SVMs were first introduced by Vapnik in the late 1960s on the foundation of SLT. However, since the middle of 1990s, the algorithms employed for SVMs started emerging with the greater availability of computing power, paving the way for numerous real-life applications with considerable results (Burges 1998; Mousavi and Iranmanesh 2011; Tavakkoli-Moghaddam et al. 2011; Vahdani et al. 2011). The basic SVM deals with two-class problems in which the data are separated by a hyper plane defined by a number of support vectors as illustrated in Fig. 1 (Peng et al. 2004). A simple introduction to SVM is presented here for completeness. For more descriptions, readers may refer to the tutorials on SVMs (Burges 1998; Vapnik 1999). The least squares support vector machines (LS-SVM) is a new training technique based on the SVM; however, it requires only the solution of a set of linear equations instead of the long and computationally hard quadratic programming problem involved in the standard SVM. In fact, the LS-SVM works with a least squares cost function (Vapnik 1999).

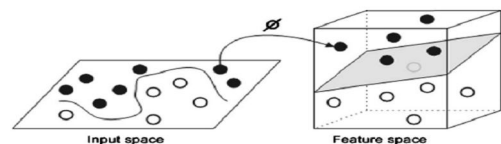


Fig. 1. Nonlinear support vector machine

**Classification:** The purpose of the SVM is to learn a separate function that separates training instances from distinct groups according to their class labels. Based on this viewpoint, SVMs become a class of supervised learning models with major applications in solving problems in classification and regression (Huang, 2012).

Through mapping input vectors  $x$  into a high-dimensional feature space, SVM models built in the new space can express a linear or nonlinear decision boundary in the original space. In the new space, an optimal separation between instances of distinct classes is obtained by the hyper plane which has the maximal distance to the nearest training instances. Thus, SVMs are regarded as an approach that produces the maximum margin hyper plane to present the maximum separation between distinct classes. The maximum margin hyper plane for a provided learning problem is uniquely illustrated by the instances that are closest to it, and these instances are regarded as support vectors. Moreover, the separate function may be linear or nonlinear. In the linearly separable case, the instances may be divided by a linear hyper plane; otherwise, the case is nonlinearly separable (Huang, 2012).

For the linearly separable case, consider a given set  $S$  with  $n$  labeled training instances  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ . Each training instance  $x_i \in R^k$ , for  $i = 1, \dots, n$ , belongs to either of the

two classes versus its label  $y_i \in \{-1, +1\}$ , where  $k$  is the input dimension. The maximum margin hyper plane can be described by the following equation:

$$y = b + \sum w_i y_i x(i) \cdot x, \quad (1)$$

where indicates the dot product; the vector  $x$  indicates a test example and the vectors  $x(i)$ s are the support vectors. In this equation,  $b$  and  $w_i$  are parameters that determine the hyper plane and must be learned by the SVM. To obtain an optimal hyper plane, one solves the following convex quadratic programming (QP) problem:

$$\text{Minimize } \frac{1}{2} \|w\|^2 \quad (2)$$

$$\text{Subject to } y_i (w \cdot x_i + b) \geq 1, \quad i = 1, \dots, n.$$

where the function  $K(x(i), x)$  is defined as the kernel function. There are several kernels for generating the inner products to build SVMs with different types of nonlinear decision surfaces in the input space. The commonly-used kernel functions contain the polynomial kernel  $K(x, y) = (xy + 1)^d$ , and the Gaussian radial basis function  $K(x, y) = \exp(-1/\delta^2(x - y)^2)$ , where  $d$  is the degree of the polynomial kernel, and  $\delta^2$  is the bandwidth of the Gaussian radial basis function (Huang, 2012).

**Regression:** The concept of a maximum margin hyper plane explained above is only used in classification. However, the SVMs have been developed for general estimation and prediction problems, containing a version of SVM for regression (LS-SVM).

The objective of LS-SVM is to find a function that approximates the training instances well by minimizing the prediction error. When minimizing the error, the risk of over-fitting is decreased by concurrently trying to maximize the flatness of the function. To determine an optimal hyper plane, one again solves the following quadratic programming problem:

$$\text{Minimize } \frac{1}{2} \|w\|^2 \quad (3)$$

$$\text{Subject to } \|y_i (w \cdot x_i + b)\| \leq \varepsilon,$$

where  $\varepsilon \geq 0$  indicates the bound for the prediction error.

The above convex optimization problem will be feasible in cases where  $f = \langle w, x \rangle + b$  actually exists and approximates all pairs  $(x_i, y_i)$  with  $\varepsilon$  precision. To permit some errors in the exchange for the technique flexibility, one introduces slack variables  $\xi_i, \xi_i^*$  to tackle otherwise infeasible constraints of the following optimization problem:

$$\begin{aligned} \text{Minimize } & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{Subject to } & \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4)$$

The constant  $C$  calculates the trade-off between the flatness of  $f$  and the amount up to which deviations larger than  $\varepsilon$  are tolerated. By constructing the Lagrangian function, this optimization problem can be constructed as a dual problem:

$$\begin{aligned} L = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \lambda_i \left( \begin{matrix} \varepsilon + \xi_i - y_i + \langle w, x_i \rangle \\ + b \end{matrix} \right) \\ & - \sum_{i=1}^l \lambda_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \\ & - \sum_{i=1}^l \lambda_i (\eta_i \xi_i + \eta_i^* \xi_i^*), \quad \text{and } \lambda_i, \lambda_i^*, \eta_i, \eta_i^* \geq 0. \end{aligned} \quad (5)$$

Solving the Lagrangian, one provides the optimal solutions  $w^*$  and  $b^*$ :

$$\begin{aligned} w^* &= \sum_{i=1}^l (\lambda_i - \lambda_i^*) x_i, \\ b^* &= y_i - \langle w^*, x_i \rangle - \varepsilon, \quad 0 \leq \lambda_i \leq C, \quad i = 1, \dots, l, \\ b^* &= y_i - \langle w^*, x_i \rangle + \varepsilon, \quad 0 \leq \lambda_i^* \leq C, \quad i = 1, \dots, l. \end{aligned} \quad (6)$$

Similar to classification, the inner products may be replaced by proper kernels for nonlinear problems. The tradeoff between minimizing the prediction error and maximizing the flatness of the regression function is monitored by enforcing the upper limit  $C$  on the absolute value of the coefficients  $w_i s$ . The upper limit will restrict the influence off the support vectors on the shape of the regression function and will be a parameter that the user must specify beside  $\varepsilon$ . The larger  $C$  is, the more closely the function may fit the data. In the degenerate case where  $\varepsilon = 0$ , the algorithm simply conducts least-absolute-error regression under the coefficient size constraint and all training instances will be support vectors. Conversely, if  $\varepsilon$  is large enough, the error approaches zero, and the algorithm outputs the flattest that encloses the data irrespective of value of  $C$  (Chen and Wang, 2007; Huang, 2012; Salgado and Alonso, 2007).

## 2.2. Cross Validation Technique

Cross validation is regarded as well-known technique for estimating generalization error (Efron, 1983; Efron and Tibshirani, 1993). Among different types of the cross

validation,  $k$ -Fold is taken into account in this paper.  $K$ -Fold cross validation is one way to enhance the holdout technique. The dataset is separated  $k$  subsets, and the holdout technique is repeated  $k$  times. Each time, one of the  $k$  subsets is employed as the test set and the other  $k-1$  subsets are considered to form a training set. Then, the average error across all  $k$  trials is calculated. The main merit of this method is that it is not important how the data is separated. Every data point becomes in a test set exactly once, and becomes a training set  $k-1$  times. The variance of the resulting estimate is decreased as  $k$  is increased. A variant of this method is to randomly separate the data to a test and training set  $k$  different times. The  $k$ -fold cross validation technique illustrates how an optimal generalization can be made using neural network structures. In the related literature, 3-fold and 10-fold cross validation approaches were suggested to estimate in the real-life applications (Duan et al., 2001; Efron and Tibshirani, 1993).

### 2.3. Steps of Proposed Intelligent Model

The proposed model is on the basis of two powerful techniques, namely LS-SVM and cross validation. In this model, LS-SVM is regarded as a supervised learning tool to handle input–output mapping and concentrates on cost data characteristics in construction projects, and the  $k$ -fold cross validation is employed to train the LS-SVM in order to provide a more realistic evaluation of the accuracy by dividing the total dataset into multiple training and test sets, and to offer reliable results. The proposed intelligent model can be a computationally efficient combination and can be suitable for the cost prediction of construction projects. The steps of the proposed model are provided as follows:

*Step 1. Dividing all data into the training dataset and test dataset:* the training data are utilized to construct the LS-SVM model, and the test data are used to verify the LS-SVM model performance.

*Step 2. Training data:* The model applies sequential data as training data. In this step, the sequential data reflects the identified attributes, and the training data are normalized into the same range (0, 1), which help avoid numerical difficulties. The function utilized to normalize data is illustrated in (7):

$$x_{sca} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

Therefore, after effecting this transformation, all the variables become dimensionless.

*Step 3. Training the LS-SVM:* In this step, LS-SVM is deployed to handle input–output mapping. The Gaussian radial basis function kernel is applied as a reasonable choice (Hsu et al., 2003; Huang, 2012). By using the  $k$ -fold cross validation technique on the training dataset, the LS-SVM training is performed to obtain the prediction model.

*Fitness definition:* According to (Chen and Wang, 2007), the fitness of the training dataset can be obtained easily, but is prone to over-fitting. To handle this problem, the  $k$ -fold cross validation technique is utilized. In this technique, after dividing the training dataset into  $k$  subsets randomly, the regression function is built with a given set of parameters ( $C_i, \delta_i$ ) by using  $k-1$  subsets as the training set. The last subset is considered the validation. The above procedure is repeated  $k$  times. Consequently, the fitness function is defined as the  $MAPE_{CV}$  of the  $k$ -fold cross validation technique on the training dataset:

$$\text{Fitness} = \min f = MAPE_{cv}, \quad (8)$$

$$MAPE_{cv} = \frac{1}{l} \sum_{j=1}^l \left| \frac{y_j - \hat{y}_j}{y_j} \right| \times 100\% \quad (9)$$

where  $y_j$  is the actual value;  $\hat{y}_j$  is the validation value and  $l$  is the number of subsets. The solution with a smaller  $MAPE_{CV}$  of the training dataset has a smaller fitness value.

*Step 4. Determining parameters:* In this step, the LS-SVM parameters are provided to perform the estimations.

*Step 5. Adopt the optimal parameter combination to build the LS-SVM model:* Substituting the test dataset into LS-SVM model obtaining the estimation values. By considering performance criteria to compute the error between actual and estimation values, through the testing performance can verify the LS-SVM model estimated capability.

Fig. 2 depicts the framework of the proposed intelligent model. This model is applied to seek a better combination of the parameters in the LS-SVM so that a smaller  $MAPE_{CV}$  is obtained during prediction iterations.

### 3. Model Validation and Comparison Results

In this section, in order to test the effectiveness of the proposed model based on the LS-SVM with cross validation technique, it is applied to the available dataset on the basis of real data presented in Hsieh (2002). In this dataset, 10 factors affecting the building project cost in the conceptual phase are considered. These factors can be categorized into two main groups: owner's preliminary requirements and site investigations. Table 1 illustrates 80 input patterns and 20 validation cases simulated. These input patterns are real collective projects in Taiwan from 1997 to 2001 (Hsieh, 2002).

The descriptions of these patterns are as follows:

- (1) Site area (in square meters);
- (2) Geology property;
- (3) Influencing householder number;

- (4) Earthquake impact;
- (5) Planning householder number;
- (6) Total floor area (in square meters);
- (7) Floor over ground (in stories);
- (8) Floor underground (in stories);
- (9) Decoration class;
- (10) Facility class; and
- (11) Normalized cost of the building project.

The cost of building projects were in dollars per square meters. In Table 1, factors 1, 3, 5, 6, 7, and 8 are

quantitative, whereas factors 2, 4, 9, and 10 are qualitative. The qualitative factors are described in Table 2.

According to the dataset, 80 rows of data are selected to be the training data and 20 rows for testing the proposed intelligent model as well as two famous techniques, including nonlinear regression and BPNN. Then the model uses the sequential data as training data. To solve the difficulties from the small training dataset, a cross validation technique is utilized in the training phase.

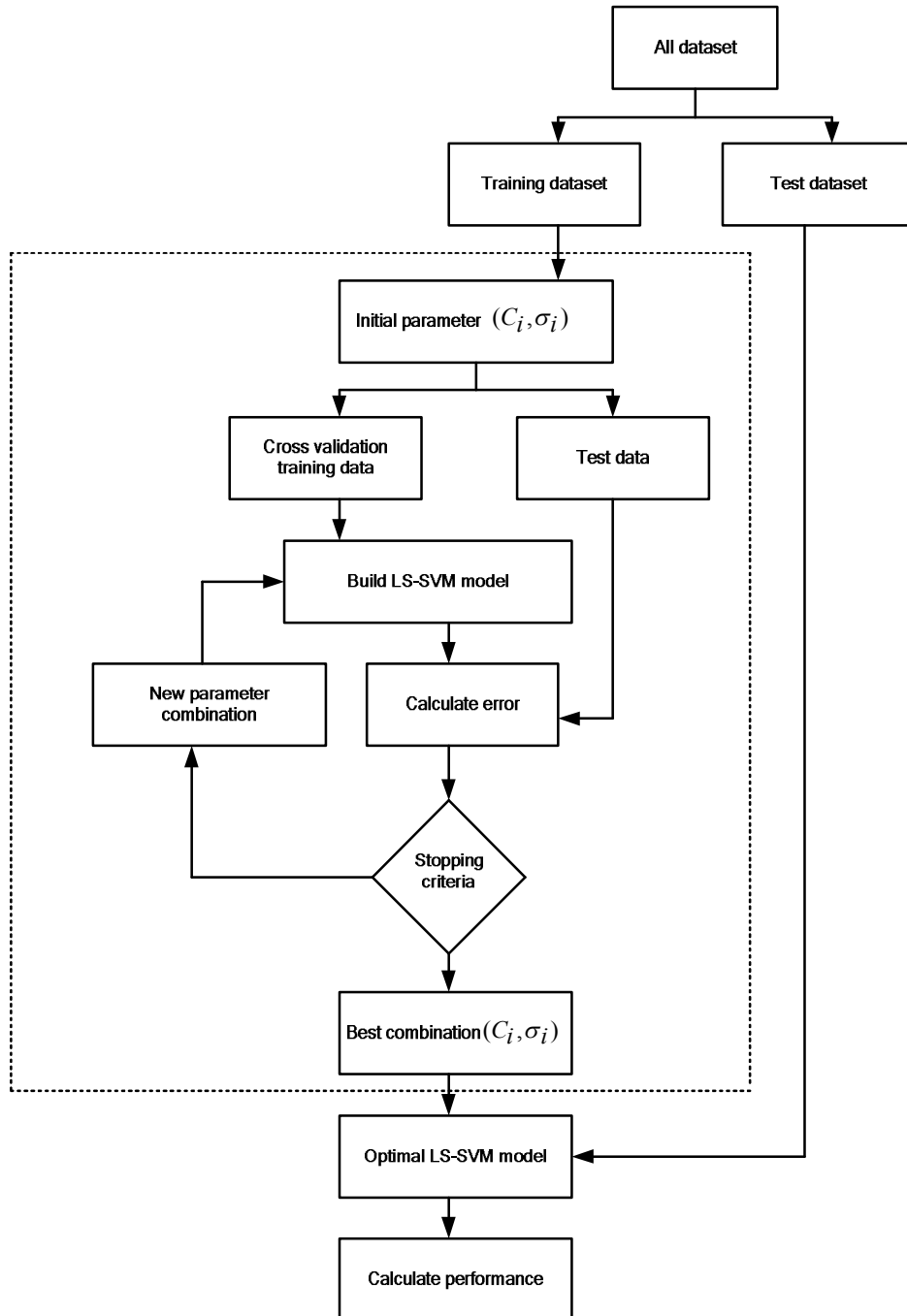


Fig. 2. The proposed LS-SVM model

Table 1  
Patterns for conceptual predicting of building project cost

No.	Input patterns										Output
	Input										
	1	2	3	4	5	6	7	8	9	10	
<i>Training data</i>											
1	873	1	67	1	50	6423.93	10	3	2	2	0.4395
2	1647	1	0	1	24	6163.21	11	2	2	3	0.5234
3	2169	1	0	1	13	6240.83	14	2	2	3	0.9176
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
78	11119	2	161	1	203	17838	15	3	1	2	0.8502
79	16204	2	122	1	255	13412	8	2	3	2	0.9487
80	19816	2	32	1	296	9128	14	2	1	1	0.9532
<i>Test data</i>											
81	1359	2	29	1	80	8146.24	9	2	1	1	0.4847
82	1749	1	36	1	86	13736.74	12	2	3	3	0.5059
83	2316	1	147	1	83	12034.57	11	2	1	3	0.6812
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	3837	2	24	1	45	10642	8	2	1	2	0.7903
99	1573	1	84	1	180	40297	6	2	3	2	0.4377
100	3574	3	164	1	131	25586	19	2	1	2	0.8015

Table 2  
Description of qualitative factors for conceptual cost estimation

Influencing factor	Qualitative option	Value
Geology property	Soft	1
	Medium	2
	Hard	3
Earthquake impact	Low	1
	High	2
Decoration class	Basic type	1
	Normal type	2
	Luxurious type	3
Facility class	Basic type	1
	Normal type	2
	Luxurious type	3

The sequential data reflects identified attributes, and training data are normalized into a (0, 1) range, which helps avoid attributes with greater numeric ranges dominating those with smaller numeric ranges, and also helps to avoid numerical difficulties. The function (7) is used to normalize data.

Three indices are used to evaluate the performance of the proposed model: mean absolute average error (MAPE), mean square error (MSE) and R-squared ( $R^2$ ) given by (10) to (12):

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (10)$$

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (12)$$

where,  $y_i$  and  $\hat{y}_i$  represent the actual and estimated values of the  $i$ -th data, respectively.  $\bar{y}$  is the average of actual data and  $N$  is the number of data.

The overall comparative results based on the MAPE, MSE and  $R^2$  indices are illustrated for the proposed model and two famous techniques in Table 3. The prediction based on the proposed model gives satisfactory results as depicted in Fig. 3. The final prediction results of the proposed model have the minimal three indices, which indicate the high performance and applicability to the conceptual project cost in construction industry. By using the proposed model, project managers can make appropriate decisions to evaluate construction projects, particularly in feasibility studies. It also increases the chance of the success of construction projects.

Table 3  
Overall comparative results

Intelligent techniques	MAPE	MSE	$R^2$
Proposed model	12.336	0.1854	0.8145
Nonlinear regression	21.229	0.5773	0.4227
BPNN	13.845	0.3258	0.6742

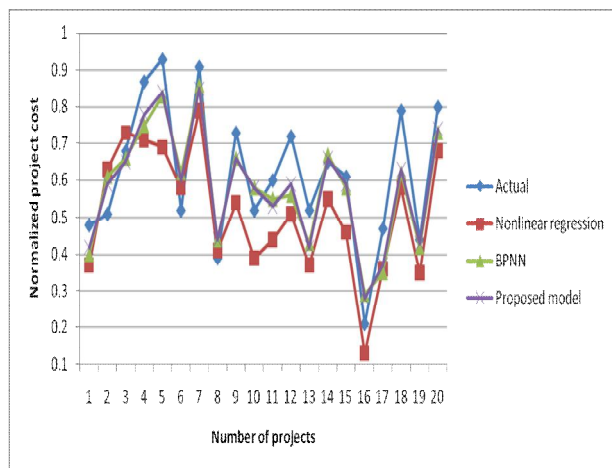


Fig. 3. Pictorial representations of the proposed model and other two famous techniques

#### 4. Conclusion

In this paper, a new intelligent model based on support vector machine and cross validation techniques was proposed to estimate the conceptual cost of construction projects. This model considered the principle of structural risk minimization, searching to minimize an upper bound

of the generalization error rather than minimize the estimation error on the training set. The presented model provided the LS-SVM with a greater potential to generalize the input-output relationship learnt during its training phase for making suitable estimations for new input data. In contrast to traditional testing and assessing techniques, this model had further merits with characteristics of high reliability and avoiding over-fitting. Without drawbacks of canonical neural networks including existing local minimal, spending long time in training period and uneasily converging to fixed values, the proposed model with cross validation technique for the training process was also superior to the neural networks in generalization and accurate ratio of estimation. Cross validation technique was well suited for the reliable estimation. The proposed model for conceptual cost estimation of the construction projects was compared with the two famous techniques, namely nonlinear regression and BPNN. Through comparison, it was concluded that the proposed model had better generalization performance and yielded lower estimation error by using the available dataset in the building projects. Performances of the proposed intelligent model for the conceptual cost estimation can be affected by values of the input parameters. As further research, optimizing the proposed model is offered by using recent meta-heuristic algorithms to improve yields and eliminate some of the difficulties.

#### 5. Acknowledgement

This paper is partly based on the research project supported by Industrial Engineering Research Center, Islamic Azad University, Qazvin Branch.

#### 6. References

- [1] An, S. H., Park, U. Y., Kang, K.I., Cho, M.Y., and Cho, H.H., (2007). Application of Support Vector Machines in Assessing Conceptual Cost Estimates. *Journal of Computing in Civil Engineering*, 21, 259–264.
- [2] Burges, C. J. C., (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 121–167.
- [3] Chen, K. Y. and Wang, C.-H., (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management*, 28, 215–226.
- [4] Chen, Q. and Zhai, M., (2002). Study on the Model of Cost Estimation of Construction Based on the BP Neural Network. *Optimization of Capital Construction*, 33-34.
- [5] Cheng, M. Y., and Wu, Y. W., (2005). Construction conceptual cost estimates using support vector machine," 22nd Int. Symp. on Automation and Robotics in Construction (ISARC 2005), Ferrara (Italy).
- [6] Cheng, M. Y., Tsai, H. C. and Sudjono, E., (2010). Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Systems with Applications*, 37, 4224–4231.

- [7] Duan, K., Keerthi, S. and Poo, A., (2001). Evaluation of simple performance measures for tuning SVM hyper parameters. Technical report, Singapore: National University of Singapore, Department of Mechanical Engineering.
- [8] Efron, B., (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78, 316–331.
- [9] Efron, B. and Tibshirani, R. J., (1993). *An Introduction to the Bootstrap*, London: Chapman & Hall.
- [10] Hsieh, W. S., (2002). Construction conceptual cost estimates using evolutionary fuzzy neural inference model. M. S. thesis, Dept. of Construction Engineering, National Taiwan University Of Science and Technology, Taiwan.
- [11] Hsu, C. W., Chang, C. C. and Lin, C. J., (2003). A practical guide to support vector classification, Technical Report, Department of Computer Science, National Taiwan University, Taipei, Taiwan.
- [12] Huang, C.-F., (2012). A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing*, 12, 807–818.
- [13] Kong, F., Wu, X. J., and Cai, L. Y., (2008). A novel approach based on support vector machine to forecasting the construction project cost. *Proc. of the 2008 International Symp. on Computational Intelligence and Design (ISCID 2008)*, 1, 21-24.
- [14] Mousavi, S. M. and Iranmanesh, S. H., (2011). Least squares support vector machines with genetic algorithm for estimating costs in NPD projects. *The IEEE International Conference on Industrial and Intelligent Information (ICII 2011)*, Indonesia, 127–131.
- [15] Peng, K. L., Wu, C. H. and Goo, Y. J., (2004). The development of a new statistical technique for relating financial information to stock market returns. *International Journal of Management*, 21(4), 492–505.
- [16] Salgado, D.R. and Alonso, F.J., (2007). An approach based on current and sound signals for in-process tool wear monitoring. *International Journal of Machine Tools and Manufacture*, 47, 2140–2152.
- [17] Sapankevych, N. I. and Sankar, R., (2009). Time series prediction using support vector machine: A survey. *IEEE Computational Intelligence Magazine*, 4 (2), 24–38.
- [18] Tavakkoli-Moghaddam, R., Mousavi, S. M., Hashemi, H., and Ghodrathnama, A. (2011). Predicting the conceptual cost of construction projects: a locally linear neuro-fuzzy model. *The First International Conference on Data Engineering and Internet Technology (DEIT)*, 398-401.
- [19] Vahdani, B., Iranmanesh, S.H., Mousavi, S.M. and Abdollahzade, M., (2011). A locally linear neuro-fuzzy model for supplier selection in cosmetics industry. *Applied Mathematical Modelling*, Article in press, DOI: 10.1016/j.apm.12.006.
- [20] Vapnik, V. N., (1999). *The Nature of Statistical Learning Theory*. Springer, London.
- [21] Williams, T. P., (1994). Predicting changes in construction cost indexes using neural networks. *Journal of Construction Engineering and Management*, 120, 306–320.