Journal of Science and Transport Technology

University of Transport Technology

Forecasting Construction Price Index using Artificial Intelligence Models: Support Vector Machines and Radial Basis Function Neural Network

Tuan Thanh Nguyen[1], Dam Duc Nguyen[1,*], Son Duc Nguyen[1], Indra Prakash[2], Phong Van Tran[3], Binh Thai Pham[1]

[1]University of Transport Technology, Ha Noi, Viet Nam

[2]Dy. Director General (R), Geological Survey of India, Gandhinagar 82010, India

[3]Institute of Geological Sciences, Vietnam Academy of Sciences and Technology, Ha Noi, Viet Nam

Article info

Type of article:
Original research paper

***Corresponding author:**
E-mail address:
damndl@utt.edu.vn

Received: 30/11/2022
Accepted: 12/12/2022
Published: 27/12/2022

Abstract: Estimation of Construction Price Index (CPI) is important for a market economy and it is a measure to manage construction investment costs. This is a tool to help organizations and individuals to reduce the effort and management of expenses for construction projects by reducing time of procedures for calculating and adjusting the total investment for the estimation and evaluation of contract price. The CPI is an indicator that reflects the level of construction price fluctuations of the type of work over time. In this study, the CPI data of Son La province, Vietnam from January 2016 to March 2022 (75 data) has been used for the modeling. Two Artificial Intelligence (AI) models namely Support Vector Machine (SVM) and Radial Basis Function Neural Network (RBFN) were proposed to predict the CPI based on limited input data. Performance of the models in correctly predicting CPI was evaluated using standard statistical indicators such as Coefficient of Determination (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) based on the historical CPI data. The results show that performance of both the models is good in predicting CPI, but performance of the SVM model (R^2 train = 0.915, R^2 test = 0.811) is the best in comparison to RBFN model (R^2 train = 0.985, R^2 test = 0.733). The proposed AI models can be used to quickly and accurately predict the CPI of an area to help management agencies, investors, construction contractors for assessing cost of construction for the purchase and development of properties/ infrastructures.

Keywords: Construction price index, Artificial Intelligence, SVM, RBFN.

1. Introduction

Construction projects require high investment capital and may take many years to complete, therefore, prior estimation of Construction Price Index (CPI) or Construction Cost Index (CCI) is required to properly manage construction investment of project costs till completion [1]. The CPI is a tool to help organizations and individuals to prepare and manage the cost of construction investment projects in advance to reduce effort, time and procedures for calculating and adjusting the total investment. The use of CPI contributes in removing difficulties and obstacles for investors and contractors to build contracts considering estimated cost of the projects. This is especially effective when the market has fluctuations of construction prices due to domestic and foreign economic factors; and social issues which are affected by many micro and macroeconomic factors such as Prices of fuel, cost of construction materials and labor prices [2]. This volatility can have a great impact on business activities, especially for large and long-term projects [3].

In the process of determining the CPI, the specialized construction agency needs to evaluate the price indicators based on the available information about the price index of the region and the neighborhood to ensure the appropriate indicators considering fluctuations of the regional market, where there is no great difference between localities [4]. The CPI reflects price fluctuations in the construction market in different localities. While determining the CPI we have to choose a category and number of certain representative works for the calculation of this Index [3]. The average CPI in the selected period, exclude some expenses for compensation, support and resettlement, loan interest during the construction time; and the initial working capital for business. The cost structure of the CPI calculation must be consistent with the cost structure depending on the regulations, on the construction management investment costs, which is fixed Until there is a change [5]. The provincial People's Committee is responsible for building, managing and operating the database system to serve the state management of construction investment costs in the fields of industry and local construction. To limit the potential financial risks for managers, there is a need for a CPI to predict expected costs.

In recent years, many forecasting techniques have been developed to handle complicated issues of forecasts. In general, there are two basic methods to predict construction costs, (1) traditional qualitative methods, and (2) quantitative methods. The traditional method must determine the relationship between prediction or dependent variables and independent variables [6]. Previous researchers have predicted the future CPI based on traditional methods [7, 8]. The main disadvantage of the traditional method is to identify all the variables to be predicted for dependency variables. The methods are represented in the time series analysis, which are a range of data points listed at evenly spaced in the order of time [9]. Time series methods try to predict future data values of a series based on the analysis of previous data values by using internal statistics between data. Elfahham (2019) has proposed a multivariable time series to provide to the parties involved in building a reliable tool to expect the price of the upcoming projects, but currently popular time series prediction models do not show promising results, especially in medium and long-term forecasting [1].

Quantitative methods such as Artificial Neural Network (ANN) are one of the Artificial Intelligence (AI) calculation systems that simulate the human brain's learning ability [10, 11]. The neural network is applied to forecast escalation in the cost of high-speed projects with reasonable accuracy [12]. Kim et al (2004) show that ANN is most beneficial to long-term forecasts than other statistical methods based on limited historical data [13]. Some studies have tried to combine more predictable tools in a model. Cheng et al (2013) Building a forecast combination model helps to identify the best forecasting and based on optimizing the various combinations of the project cost using forecasting models [14]. The results of applying the model to the actual project should show high accuracy and minimum risk of major errors. Gwang-Hee Kim (2004) applied three techniques MRA, NN and CBR to estimate the construction cost of Korean residential buildings. These three approaches used data containing 530 historical costs. The results show that the NN machine learning model is more accurate than the CBR or MRA model [13].

As the model development is continuous process, so in this study we have further explored other two good Machines Learning (ML) models namely Support Vector Machines (SVM) and Radial Basis Functions Networks (RBFN) to forecast CPI using avialble construction cost data of Son La Province, Vietnam. The advantage of the SVM algorithm is that it works well for large data samples and often gives results that are superior to other algorithms in supervised learning. On the other hand RBNF model has advantage of easy design, good generalization and strong tolerance to input noise besides online learning ability. The proposed models' algorithms were developed using Weka 3.8.4 software. The Weka is free software available under the GNU General Public License.

The results of this study would be useful in quickly and accurately predicting CPI to the management agencies, investors, construction contractors to pre-plan the construction investment costs. This will also help in suitably adjusting changing construction cost with time.

2. Preparation of database

In most of the world, the construction price indicators are announced by official state agencies [15]. The CPI is usually published on the official state agency website (for example, the Ministry of Construction or the Department of Construction). The quotation may be done on a quarterly, half - year or annual basis. In neighboring countries as well as many other parts of the world, the construction price index is assessed based on the balance of supply and demand of the market. The CPI is distinguished by the type of work, area and time of quotation [16]. In addition, they are expressed by the ratio of construction costs at the time of comparison with the construction cost at the time of reference. Therefore, the CPI shows the increase or decrease of construction costs over time [17].

In Vietnam, data information on construction norms, construction prices, construction investment capital, CPI issued or announced by competent state agencies. Database of construction investment projects, construction contracts collected through investigations, surveys or provided by organizations and individuals under the coordination, information sharing mechanism and information. Periodic statistical reporting regime [18].

The data in this study is of Son La province CPI from January 2016 to March 2022 (75 data), which is providing a clear picture of economic changes each year. This data is based on several input parameters such as material, labor, construction machinery and equipment which are affecting cost of the construction. The data was randomly divided into 70% (53 data from Jan 2016 to Dec 2020) for the training process and 30% (22 data from Jan 2021 to March 2022) for the process of verification (validation) [19]. In order to evaluate the accuracy of performance of the models in correctly predicting CPI, standard statistical indices: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2) were used.

3. Models and Methodology of Validation

3.1. Radial Basis Function Networks (RBFNs)

Radial Basis Function Networks (RBFNs), a popular alternative to the MLP Neural Networks, are defined as a supervised neural networks for solving modeling problems in poly dimensional space [20]. The architecture of this network is designed comprising of three layers namely input layer consisting of 14 neurons, hidden layer (referred to as the RBF units) which takes in a set of inputs and produces outputs through activation function, and an output layer that contains one neuron. The input data is processed by the RBF units using the K-means algorithm to reduce its dimensionality and then transform the data to a new space [21]. The learning procedure of the RBFN is carried out in two phases: (i) the numbers of clusters (hidden neurons) are calculated using the K-means algorithm and (ii) optimal estimation of the kernel parameter. The RBFN is trained to optimize kernel parameters to minimize the error E as follows:

$$E = \sum_{i=1}^{n}(y_i - O(x_i)) + \sum_{j}^{m}\lambda_j w_j^2) \qquad (1)$$

where w_j is the load coefficient and O(x) is the output of the RBF network:

The output value of the output layer is calculated as:

$$O(x) = \sum_{j=1}^{m} w_j h_j \tag{2}$$

$$h_j(x) = \exp(-\frac{\left\| x - c_j \right\|^2}{r^2}) \tag{3}$$

where hj(x) is output value; cj is the center point of the basis function; r is radius of the basis function, and m is the number of clusters.

3.2. Support Vector Machine (SVM)

The SVM is a machine learning algorithm that produces an optimal separating hyperplane to differentiate classes that overlap and are not separable in a linear way. It was originally developed for classification purposes; however, it can also be used for regression problems [22]. In this study, SVM for regression (SVR) was implemented. SVR is a kernel-based learning regression method that was proposed by Cherkassky (2020) [23]. It is based on the computation of a linear regression function in a multidimensional feature space. Hence, modeling a linear regression hyperplane for nonlinear relationships is possible with the feature space. Two forms of SVM regression, namely, “epsilon (ε)-SVR” and “nu (v)-SVR,” are commonly used in the SVM model. The original SVM formulations for regression (SVR) uses parameter cost (c) and epsilon (ε) to apply a penalty to the optimization for points that are incorrectly predicted. Zhang et al. [24] have utilized SVR in environmental monitoring studies to predict SOC. In SVM regression, the Gaussian Radial Basis Function (RBF) kernel was applied. We employed the RBF kernel to obtain an optimal SVM regression model which is important to obtain the best set of penalty parameters C and kernel parameters gamma (γ) for the SOC training data. In the present study, we evaluated the training set and then tested the model performance on the validation set.

3.3. Validation methods

To evaluate and compare the models’ performance standard statistical measures namely coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) were used by matching the measured and estimated values. R^2 is an important criterion in regression analysis. Values of R^2 between the predicted result and the actual outcome, ranges from 0 to 1. A high R^2 value indicates a good correlation between the predicted value and the actual value. For the accuracy assessment, training data was used in the construction of the models, whereas separate testing data was used for the validation of the models [25].

RMSE is an error measurement of the mean squared difference between the model's predicted and actual outputs [26], while MAE measures the mean error between them. Compared with R^2, lower RMSE and MAE values indicate better performance of AI, ML algorithms. The formula for calculating the above three criteria can be found in the documents [27-31].

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \tag{4}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| y_i - \hat{y} \right| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i-1}^{N} (y_i - \hat{y})^2} \tag{6}$$

Where: $\hat{y}$ predicted value of y; $\bar{y}$ mean value of y;

4. Results and Discussion

Analysis of results (Table 1) show that SVM model has lower training efficiency but better verification value than RBFN model. The RMSE error value of the SVM model on training and testing data are 1.338 and 2.009, respectively, whereas for the RBFN model these values are 0.561 and 2.055, respectively. The value of MAE for SVM model on training data is 0.789 and for the testing it is 1.19, whereas for the RBFN model it is

0.269 for training and 1.297 for verification (validation/ testing).

Table 1. RMSE, MAE analysis of the models using data

Parameters	Training		Test	
	RBFN	SVM	RBFN	SVM
R^2	0.985	0.915	0.733	0.811
RMSE	0.561	1.338	2.055	2.009
MAE	0.269	0.789	1.297	1.19

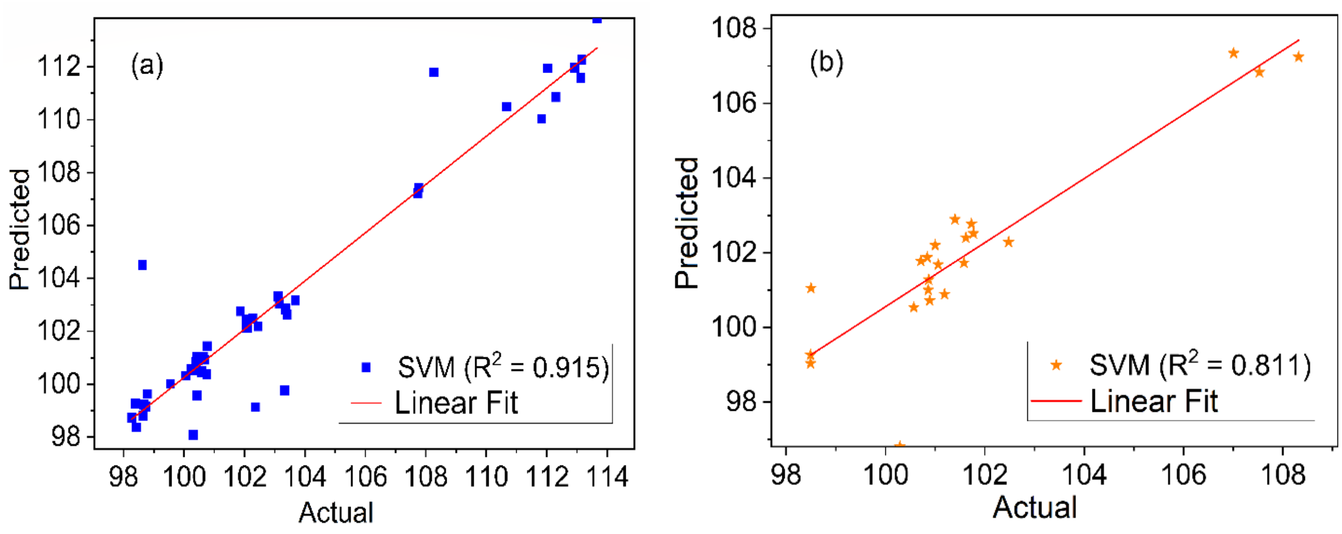
The analysis of R² results (Fig.1) show that both the models have a good value on training data (SVM: 0.915 and RBFN: 0.985), whereas R² value for the SVM model is better (R² = 0.811) than the RBFN model (R² = 0.733) on the testing/ validation data. R² results show that the predictive capability of CPI of both the models is good.

The cumulative Frequency-Error chart of SVM and RBFN models for the training phase is presented in Fig.2 (a, c) and for the verification stage in Fig.2 (b, d). Analysis of results show that the error values corresponding to the training data set and the verification data set are small. With SVM model, the percentage of samples with errors between the test value is in the range of [-0.5; 1], similar to the data set of errors in the range [0; 2]. As for the RBFN model, the percentage of samples has an error between the test value in the range of [-0.5; 1], similar to the data test set within the range of [-0.5; 2].

Figure 3 shows comparative results of the forecast values and the actual calculated values of the CPI on the training data and test data. The results show that the values predicted and actual in this case are close to each other, which indicate that the model suitability for the accurate estimation of CPI with the input data.

The actual value of price from January 2016 to January 2017 that is the CPI tends to decrease sharply. From the beginning of January 2017 to September 2017, the CPI was in a stable state with negligible fluctuations, the last three months of 2017 tended to increase. In 2018, the CPI in the first 3 months was stable but decreased in April and kept stable until the end of August, then increased in September and also fluctuated slightly in the last months of the year. In 2019, the CPI remained at a stable level. In 2020, CPI fluctuated slightly, as it increased in the first months of the year and the end of the year. From 2021 to the first quarter of 2022, the price index (CPI) tends to skyrocket as the economy gradually stabilized and recovered.

Figure 4 results show, the forecast value of the two models in also different, the SVM model (Figure 4.a) the forecast value decreased slightly in April and increased in May. In RBFN (Figure 4.b) the forecast for April and May tends to decrease. Check the current contruction index in Son La shows an uptrend.



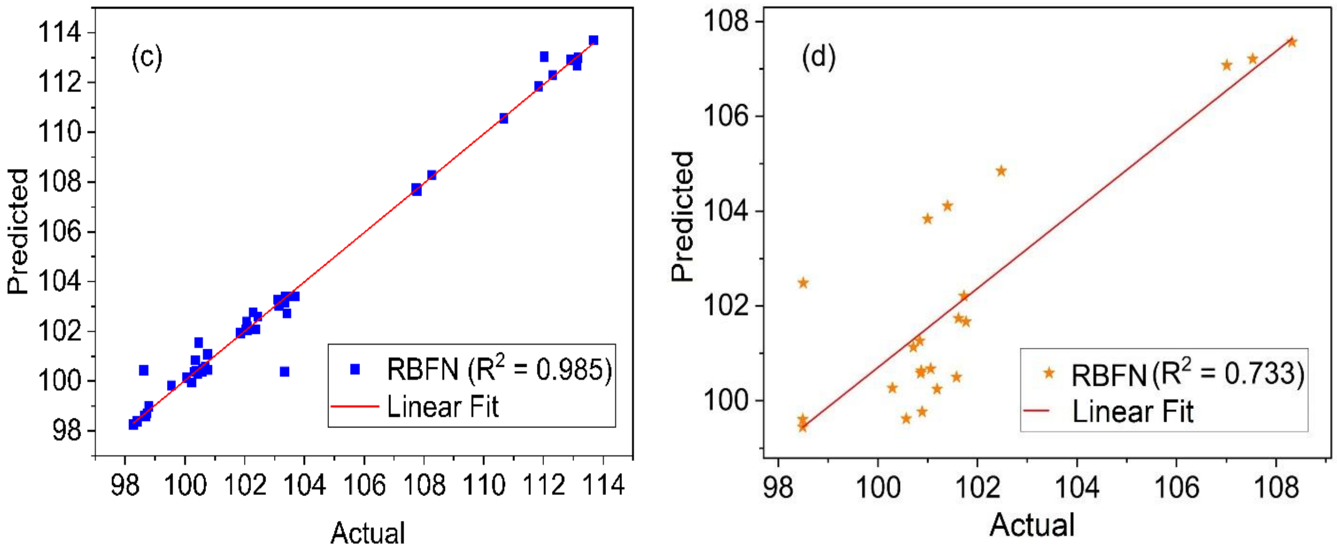


Fig 1. Coefficient of determination R² with SVM model: (a) training data; (b) test data; RBFN model: (c) training data; (d) test data

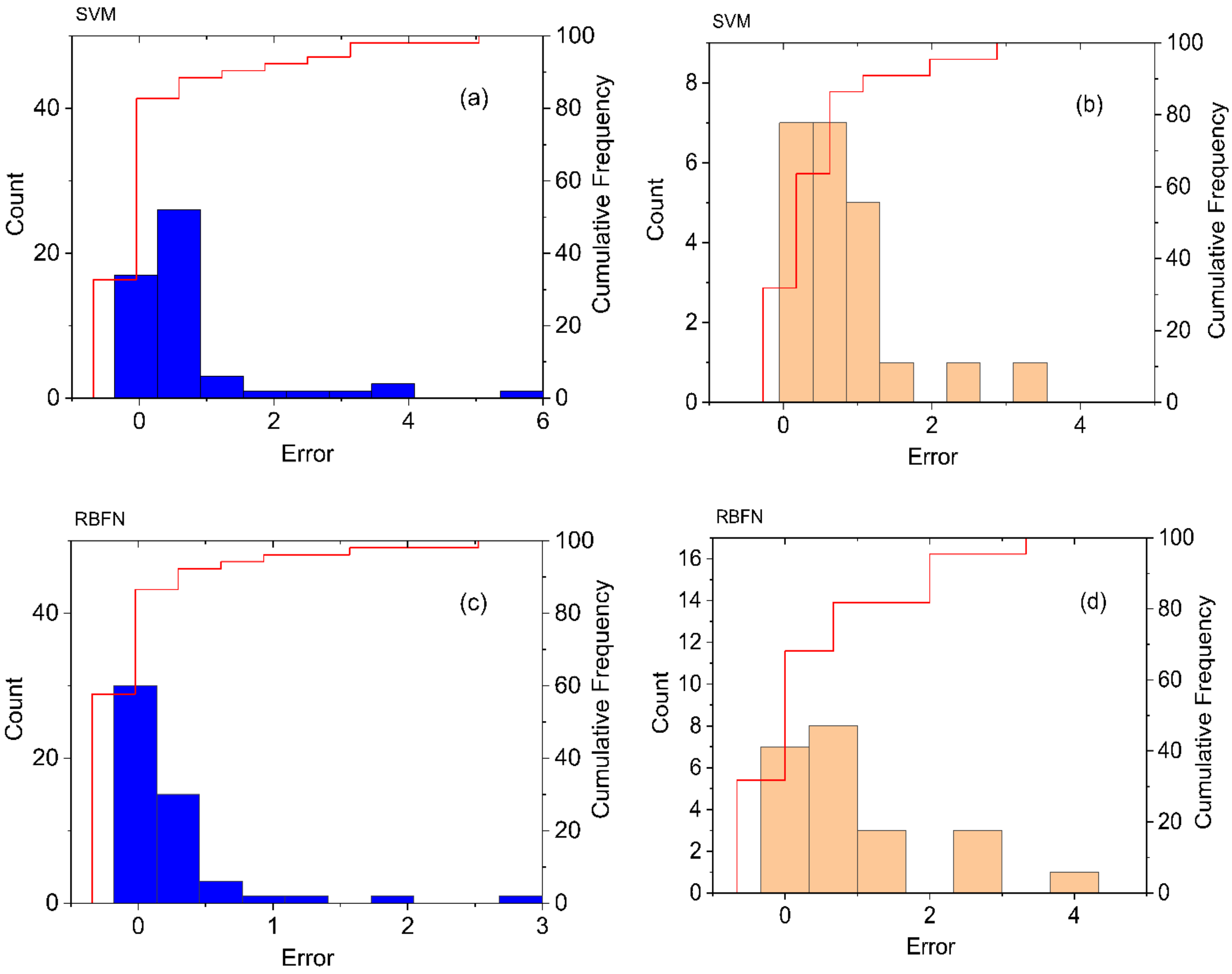


Fig 2. Cumulative Frequency-Error chart with SVM model: (a) training data; (b) test data; RBFN model: (c) training data; (d) test data

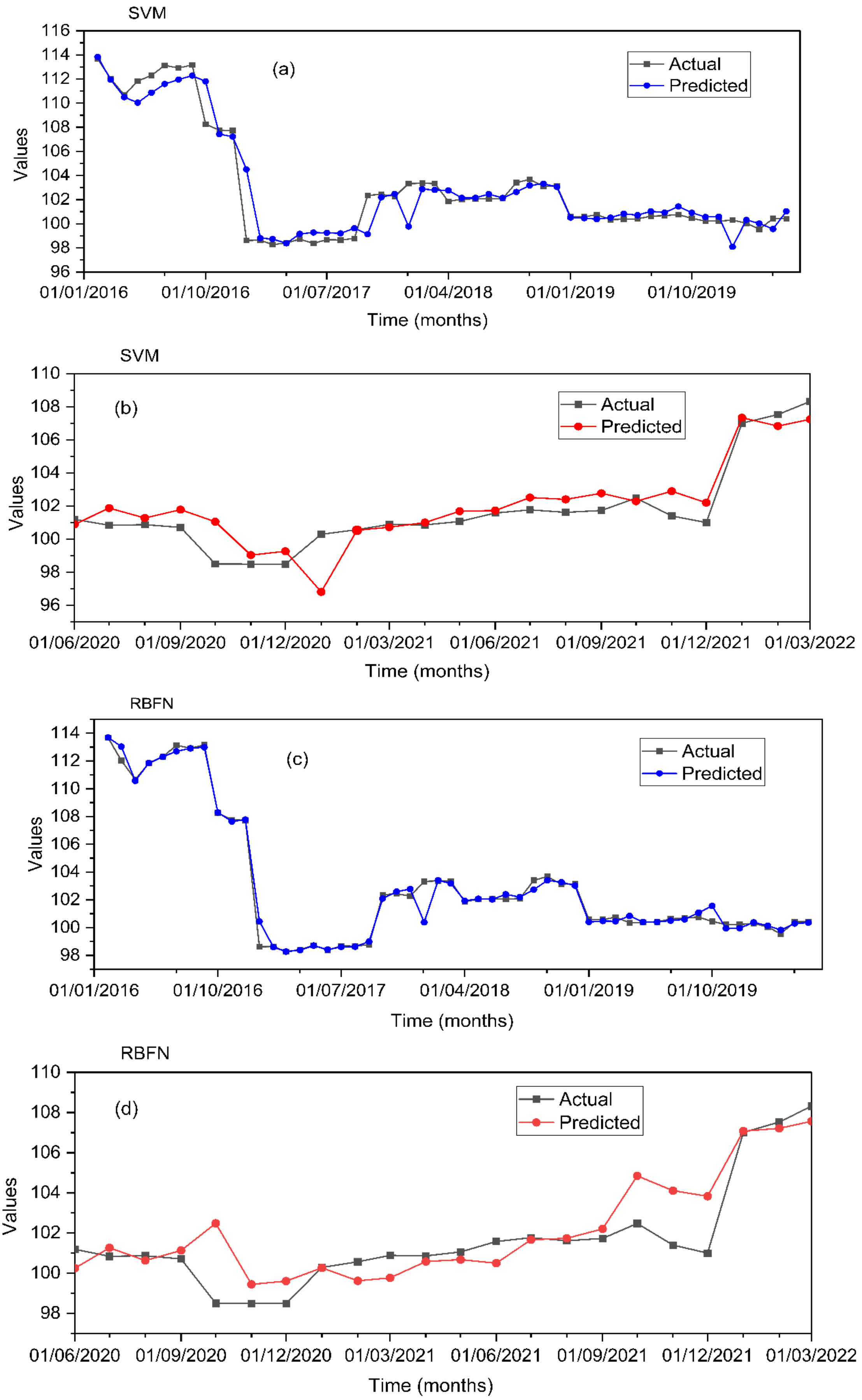


Fig 3. Results of two models SVM and RBFN: (a, c) training data; (b, d) test data.

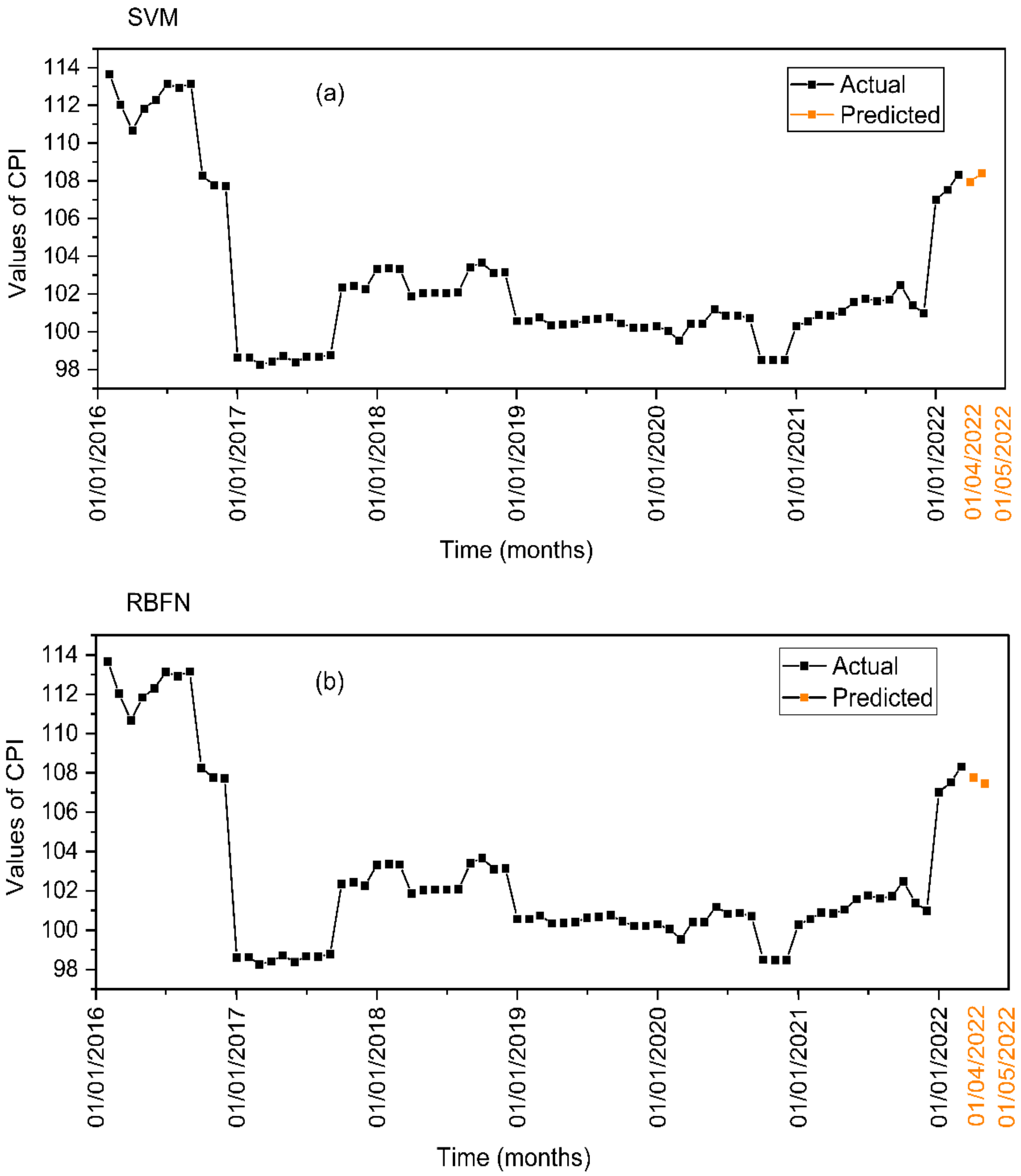


Fig 4. Predicting the value of CPI by (a): SVM model and (b): RBFN model

The forecast results show that the selection of models with good forecast capacity is very important and it directly affects the forecast ability [2]. The higher the value of R^2, better the forecast capacity. Moreover, with highly volatile data in a short time, the forecast ability should be further verified. For the effective forecasting models with high volatile data. We need to analyze parameters affecting the CPI, which will provide a clear insight on the construction price index reaction for the changes of the impact parameters which will be very helpful for the price control [3].

5. Conclusions

Construction Price Index (CPI) is very important for the market economy. It also helps contractors to identify trends, movements and directions in the construction market. In addition, it is also a tool to determine and adjust total investment, estimate of construction costs; and help in the adjustment of contract value and in contract settlement; and also in conversion of investment capital, etc. Most construction investment projects take a long time to complete and are affected by price fluctuations. To limit risks right from the project formulation, it is necessary to have proper tools and methods to forecast the correct CPI in the future.

As the model development is continuous process, so in this study we have further explored AI/ML models namely SMV and RBFN to estimate CPI in comparison to other ML and conventional models. The present study revealed that SVM and RBFN models are very effective forecasting tools to predict correct future CPI for proper planning and management of construction in any area. Out of the two models studied (SVM and RBFN), performance of the SVM model is the best in the prediction of CPI (SVM: $R^2 = 0.811$, RMSE = 2.009, MAE = 1.19; RBFN: $R^2 = 0.733$, RMSE = 2.055, MAE = 1.297). Thus this work is expected to contribute to the construction engineering and management community by helping cost engineers and capital planners in the preparation of bids, estimation of costs and in the preparation of budgets for timely project implementation... Further, it is planned to incorporate more input factors in the models for refining the CPI estimation considering fluctuating costs of the raw material, transport and labour etc. depending on the local and global factors.

References

[1] Y. Elfahham. (2019). Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alexandria Engineering Journal*, 58(2), 499-506.

[2] A. JoukarandI. Nahmens. (2016). Volatility forecast of construction cost index using general autoregressive conditional heteroskedastic method. *Journal of construction engineering and management*, 142(1), 04015051.

[3] M.-T. Cao, M.-Y. ChengandY.-W. Wu. (2015). Hybrid computational model for forecasting Taiwan construction cost index. *Journal of Construction Engineering and Management*, 141(4), 04014089.

[4] A.H. To, D.T.-T. Ha, H.M. NguyenandD.H. Vo. (2019). The impact of foreign direct investment on environment degradation: evidence from emerging markets in Asia. *International journal of environmental research and public health*, 16(9), 1636.

[5] H.L. Chen. (2007). Developing cost response models for company-level cost flow forecasting of project-based corporations. *Journal of Management in Engineering*, 23(4), 171-181.

[6] J.-w. XuandS. Moon. (2013). Stochastic forecast of construction cost index using a cointegrated vector autoregression model. *Journal of Management in Engineering*, 29(1), 10-18.

[7] S.M. TrostandG.D. Oberlender. (2003). Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *Journal of construction Engineering and Management*, 129(2), 198-204.

[8] P. NguyenandV. Likhitruangsilp. (2017). Identification risk factors affecting concession period length for public-private partnership infrastructure projects. *International Journal of Civil Engineering and Technology*, 8(6), 342-348.

[9] H.H. Elmousalami. (2020). Artificial intelligence and parametric construction cost estimate modeling: State-of-the-art review. *Journal of Construction Engineering and Management*, 146(1), 03119008.

[10] M. Parsajoo, D.J. Armaghani, A.S. Mohammed, M. KhariandS. Jahandari. (2021). Tensile strength prediction of rock material using non-destructive tests: A comparative intelligent study. *Transportation Geotechnics*, 31, 100652.

[11] B. Indraratna, D.J. Armaghani, A.G. Correia, H. HuntandT. Ngo. (2022). Prediction of resilient modulus of ballast under cyclic loading using machine learning techniques. *Transportation Geotechnics*, 100895.

[12] C.G. WilmotandB. Mei. (2005). Neural network

modeling of highway construction costs. *Journal of construction engineering and management*, 131(7), 765-771.

[13] G.-H. Kim, S.-H. AnandK.-I. Kang. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and environment*, 39(10), 1235-1242.

[14] M.-Y. Cheng, N.-D. HoangandY.-W. Wu. (2013). Hybrid intelligence approach based on LS-SVM and differential evolution for construction cost index estimation: a Taiwan case study. *Automation in Construction*, 35, 306-313.

[15] J. WangandB. Ashuri. (2017). Predicting ENR construction cost index using machine-learning algorithms. *International Journal of Construction Education and Research*, 13(1), 47-63.

[16] T.P. Williams. (1994). Predicting changes in construction cost indexes using neural networks. *Journal of construction engineering and management*, 120(2), 306-320.

[17] F.Y.Y. LingandV.T.P. Hoang. (2010). Political, economic, and legal risks faced in international projects: Case study of Vietnam. *Journal of professional issues in engineering education and practice*, 136(3), 156-164.

[18] P.T. NGUYENandQ.L.H.T.T. NGUYEN. (2020). Critical factors affecting construction price index: An integrated fuzzy logic and analytical hierarchy process. *The Journal of Asian Finance, Economics and Business*, 7(8), 197-204.

[19] Q.H. Nguyen, H.-B. Ly, L.S. Ho, N. Al-Ansari, H.V. Le, V.Q. Tran, I. PrakashandB.T. Pham. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021.

[20] S. HaykinandN. Network. (2004). A comprehensive foundation. *Neural networks*, 2(2004), 41.

[21] D. GilandM. Johnsson. (2010). In Supervised SOM based architecture versus multilayer perceptron and RBF networks. *The Swedish AI Society Workshop May 20-21; 2010; Uppsala University*, Linköping University Electronic Press: pp 15-24.

[22] V. Vapnik. (1999). The nature of statistical learning theory. *Springer science & business media*.

[23] V. CherkasskyandF.M. Mulier. (2007). Learning from data: concepts, theory, and methods. *John Wiley & Sons*.

[24] Y. Zhang, B. Sui, H. ShenandL. Ouyang. (2019). Mapping stocks of soil total nitrogen using remote sensing data: A comparison of random forest models with different predictors. *Computers and Electronics in Agriculture*, 160, 23-30.

[25] G.J. de Bondt, E. HahnandZ. Zekaite. (2021). ALICE: Composite leading indicators for euro area inflation cycles. *International Journal of Forecasting*, 37(2), 687-707.

[26] A.R. Ghanizadeh, A. Ghanizadeh, P.G. Asteris, P. FakharianandD.J. Armaghani. (2022). Developing Bearing Capacity Model for Geogrid-Reinforced Stone Columns Improved Soft Clay utilizing MARS-EBS Hybrid Method. *Transportation Geotechnics*, 100906.

[27] D.J. Armaghani, H. Harandizadeh, E. Momeni, H. MaizirandJ. Zhou. (2022). An optimized system of GMDH-ANFIS predictive model by ICA for estimating pile bearing capacity. *Artificial Intelligence Review*, 55(3), 2313-2350.

[28] F. Shan, X. He, D.J. Armaghani, P. ZhangandD. Sheng. (2022). Success and challenges in predicting TBM penetration rate using recurrent neural networks. *Tunnelling and Underground Space Technology*, 130, 104728.

[29] H.-V.T. Mai, T.-A. Nguyen, H.-B. LyandV.Q.J.A.i.C.E. Tran. (2021). Prediction compressive strength of concrete containing

GGBFS using random forest model. 2021.

[30] B.T. Pham, T.-A. Hoang, D.-M. Nguyen and D.T.J.C. Bui. (2018). Prediction of shear strength of soft soil using machine learning methods. 166, 181-191.

[31] B.T. Pham, C. Qi, L.S. Ho, T. Nguyen-Thoi, N. Al-Ansari, M.D. Nguyen, H.D. Nguyen, H.-B. Ly, H.V. LeandI.J.S. Prakash. (2020). A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil. 12(6), 2218.