Modern Management based on Big Data III
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220084

Forecasting Construction Cost Using Artificial Neural Network for Road Projects in the Department of Public Works and Highways Region XI

Donna Ville GANTE [a,1], Dante L. SILVA [b], and Meriam P. LEOPOLDO [c]
[a] *School of Graduate Studies, Mapua University, Manila, Philippines*
[b] *School of Civil, Environmental, & Geological Engineering, Mapua University, Manila, Philippines*
[c] *College of Engineering and Architecture, Mapua Malayan Colleges Mindanao, Davao, Philippines*

Abstract. The development of roads has been one of the nation's most essential infrastructural initiatives. It is an essential mode of transportation that plays an important role in our everyday lives. Because of its importance, the government has allotted large budgets in making roads in different parts of the country. The quantity and complexity of road construction projects have substantially expanded in recent years. Numerous novel methods and technology have been developed to facilitate road construction budgeting, planning, and decision-making. Using Artificial Neural Network (ANN), this study constructed a forecasting model to accurately anticipate the future costs of road improvements. Between 2017 and 2020, fifty (50) completed road projects from the Department of Public Works and Highways (DPWH) Regional Office XI were utilized by the researcher. The DPWH RO XI is one of the country's largest implementing offices for constructing public roads catering the entire Davao Region. This research used the project cost as the dependent variable while the independent variables are the construction activities' revised duration and variation order in running the model. Multiple linear regression model performance was compared to the performance of the neural network prediction model. The data included the major construction activities for road projects with its corresponding revised duration, actual and planned cost, and the reason for variation order. It demonstrates that the neural network models outperform to the multiple linear regression (MLR) model in terms of prediction accuracy. This research offers a model to the government agencies and contractors implementing road construction in predicting road construction costs more accurately.

Keywords. Construction Project Management, Project Cost, Construction Schedule, Artificial Neural Network

1. Introduction

A solid construction plan serves as the basis for developing a budget and timetable [1]. From a competitor's standpoint, estimate is a crucial step in the construction process. In

[1] Corresponding Author: Donna Ville L. Gante, School of Graduate Studies, Mapua University, Manila, Philippines, E-mail: dvlgante@mymail.mapua.edu.ph

fact, the true cost of the project was only be known once it has been completed, and the contractor could only make the projected profit if the estimated cost is less than or equal to the project's actual cost [2]. Project cost estimating faces several challenges throughout the process. Cost overruns, schedule delays, scope revisions, contingencies and inflation are all common obstacles [3]. In the road construction sector cost forecasting is one of the most important aspects of planning, budgeting, and decision-making. The greater the project's scope is understood, the more precise estimates may be provided, as more project needs are specified [4]. Prior knowledge of these expenditures is crucial for both the contractor and the owner. Using previous data to forecast future results, trend analysis was implemented which is a statistical technique and was achieved through monitoring cost and schedule variations [5].

Firms creating cost plans can utilize a list of standard elements from a published industry guide or create their own from a library of group and sub-elements. The goal of standardizing data is to guarantee that all functional components are identified. After determining the element descriptions, it is time to assign a target price to each, and one way to do so is to look for any data used to generate the budget as this would have already established a benchmark [6].

Cost and schedule discrepancies might create undesired consequences that would lead to low customer satisfaction. Creating new techniques and adopting innovations are essential to resolve some of the most frequently encountered problems in the construction industry [7]. Artificial Neural Networks are designed to train the input and output of data, they have the potential in showing updated results using the new training examples [8]. It was shown that positive applications of Artificial Neural Networks in terms of cost prediction, scheduling, risk assessment, claims and disputes, resolve outcomes and decision making [9].

This study attempts to estimate the future total cost of road construction projects by employing an artificial neural network and comparing it to a multiple linear regression model that can predict future costs. This research used the completed road projects under the DPWH Regional Office XI in forecasting the road project's actual-to-planned cost. The DPWH Region XI is one of the country's largest implementing offices for constructing public roads catering to the entire region XI.

2. Related Studies

All construction projects have costs and knowing those costs ahead of time is critical for both contractor and the owner. Trend analysis is a mathematical technique for predicting future outcomes based on historical data. This is accomplished by keeping track of cost and schedule variations. The ability to estimate future performance is one of the tangible benefits of comprehensive data collection and analysis [10]. The skeleton framework of a specified budget is constructed by dividing the cost limit into proportional sections or components of a project's scope of work. The need of including all parts must be highlighted, as eliminating one will only reveal the problem during the design process which is undesirable [11]. In the Philippines, the DPWH adopted the Standard Specifications for Highways, Bridges, and Airports where the major work activities are the facilities for engineers, other general requirements, subbase and base courses, surface courses, bridge construction, drainage, and slope protection structures, miscellaneous structures, and material details [12]. In an extensive review of previous research, an upto-date application of Artificial Neural Network were found in the context of cost, duration, risk analysis, productivity, safety, dispute, unit rate, and hybrid models. It authorizes the significance of ANNs in implementing a variation of forecast, sorting, enhancement, and creating framework of connected tasks in construction management [13]. ANN have better pattern recognition and learning abilities to get a reliable result [14].

3. Methodology

3.1. Data Collection, Analysis, and Interpretation

This phase involves the collection, organization, and analysis of data pertaining to construction scheduling to meet the study's primary purpose. The method of the data collection in this phase was through the collection of historical data from the DWPH Regional Office XI in Davao City specifically, road projects. After gathering the data that was discussed in the former process, identifying, and classifying each factor in a structured and detailed way.

3.2. Development of an ANN Model

Using the ANN model, the researcher conducted an effective assessment instrument of the established factors from the assessed and evaluated data obtained in the former process.

An ANN model for predicting the actual to planned cost ratio was developed using the following core features of the model: (a) Levenberg-Marquardt as the training algorithm (TA) [15], (b) hyperbolic tangent sigmoid as the activation function (AF) [16], and (c) "n" number of hidden neurons (HN), where "n" is the governing value. In identifying the final prediction model, the Pearson's correlation coefficient (R), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE) was used as the performance markers of the final model [17].

3.3. Relative importance (RI) using Garson's Algorithm (GA)

The researcher determined the significance of input parameters to the output using GA. From the output from the previous phase, the researcher analyzed the regression plots of the ANN Model in determining the RI of each factor [18]. This showed which is the most significant factor and least significant factor to the actual-to-planned cost ratio.

4. Result and Discussions

4.1. Collection and Organization of Factors

The identification of the factor of road construction actual and planned cost came from the inferences of numerous literature reviews historical data. The researcher gathered historical data of the fifty (50) completed projects of DPWH Regional Office XI from 2017 to 2020. The data included the major construction activities for road projects with its corresponding revised duration, actual and planned cost, and the reason for variation order. The major construction of activities on road construction based on the DPWH technical specifications are the facilities for the engineers (Factor 1), other general requirement (Factor 2), earthwork (Factor 3), sub-base and base course (Factor 4), surface courses (Factor 5), drainage (Factor 6), and slope protection (Factor 7).

4.2. Descriptive Statistics of the Dataset

The descriptive statistics of the dataset used in this study are presented in Table 1. It was shown that the highest mean was observed in Factor 2 (other general equipment), while the least was observed in Factor 3 (earthworks). The skewness of the data for all factors was observed to be greater than 1, which suggests that the dataset is highly skewed [19]. Moreover, the kurtosis values observed for all datasets were all greater than 1, which implies that the data distribution is too peaked [20].

Table 1. Artificial Neural Network Model simulations

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
N	50	50	50	50	50	50	50
Mean	148.96	156.70	105.80	106.58	122.50	125.00	146.70
Std. Dev.	179.25	182.22	178.32	162.06	165.70	166.41	179.20
Skewness	1.503	1.398	2.351	1.669	1.590	1.729	1.528
Kurtosis	1.998	1.707	6.387	2.514	2.453	3.445	2.353
Range	720.00	720.00	821.00	660.00	660.00	702.00	765.00

4.3. Development of an Artificial Neural Network Model

Using the core features of the model described in Section 2.2 including the TA, AF, and HN as well as the performance criteria such as R, MSE, and MAPE, the results of the model development for each of the hidden neurons simulated in this study was obtained and was shown Table 2 and Figure 1.

Table 2. Artificial Neural Network Model simulations

HN	R (Training)	R Validation)	R (Testing)	R (All)	MSE	MAPE
1	0.89429	0.95602	0.99337	0.89651	5.412e-05	1.143%
2	0.91425	0.97218	0.95265	0.90717	4.7813e-05	1.030%
3	0.90760	0.97879	0.98647	0.90823	4.227e-05	0.996%
4	0.90899	0.95471	0.95201	0.91031	4.046e-05	0.995%
5	0.91224	0.95754	0.91906	0.91196	3.740e-05	0.978%
6	0.91447	0.90985	0.93166	0.91373	3.634e-05	0.897%
7	0.91684	0.93372	0.96832	0.91462	3.5209e-05	0.867%
8	0.92529	0.96259	0.96180	0.92533	3.0649e-05	0.863%
9	0.93314	0.97943	0.98641	0.93435	2.168e-05	0.834%
10	0.93385	0.98443	0.99095	0.93447	1.5642e-05	0.826%
11	0.93790	0.95139	0.99047	0.94782	1.5126e-05	0.677%
12	0.94469	0.98199	0.98765	0.95710	1.4495e-05	0.667%
13	0.95086	0.99909	0.98153	0.96215	1.4351e-05	0.624%
14	0.95826	0.98475	0.99897	0.98158	1.4003e-05	0.607%
15	0.96931	0.99936	0.99943	0.99216	8.4078e-06	0.541%

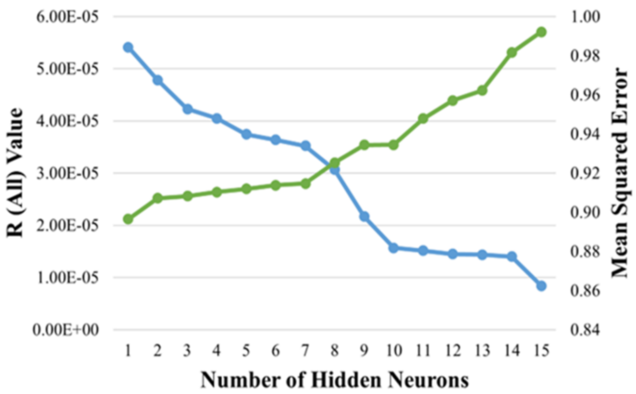

Figure 1. Impact of the Hidden Neurons on the R Value and MSE

Based on the simulations for the final prediction model, it was observed that the best model was seen in the results of the simulation of 15 hidden neurons, giving the largest R value and the least MSE and MAPE. This is similar to the suggestion made in the study of Gunduz and Sahin, wherein the number of HN is recommended to be 2m+1 [21]. The final model structure is 7-15-1 (input-hidden-output). The regression plots of the ANN model development phases are exhibited in Figure 2. The results show a very high R value almost equal to 1 [22].

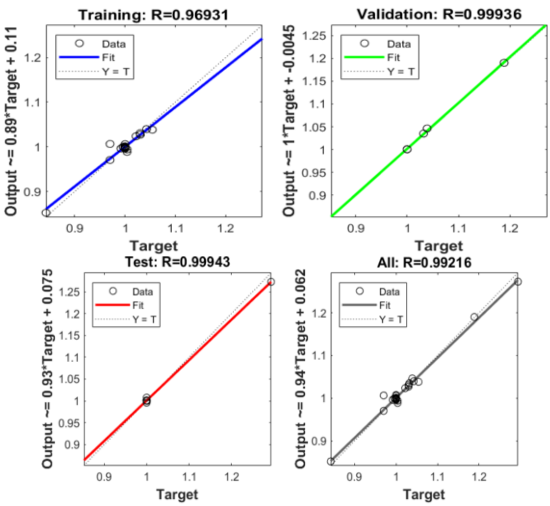

Figure 2. Regression Plots of Neural Network Model Development Phases

4.4. RI using GA

The RI of the input parameters was analyzed through their connection weights using GA. The calculation shows that the most significant parameter to the Actual to Planned Cost Ratio is Factor 4 (Sub-base and Base Course), while the least important factor is Factor 1 (Facilities for the Engineer). The relative importance of each factor is shown in Figure 3.

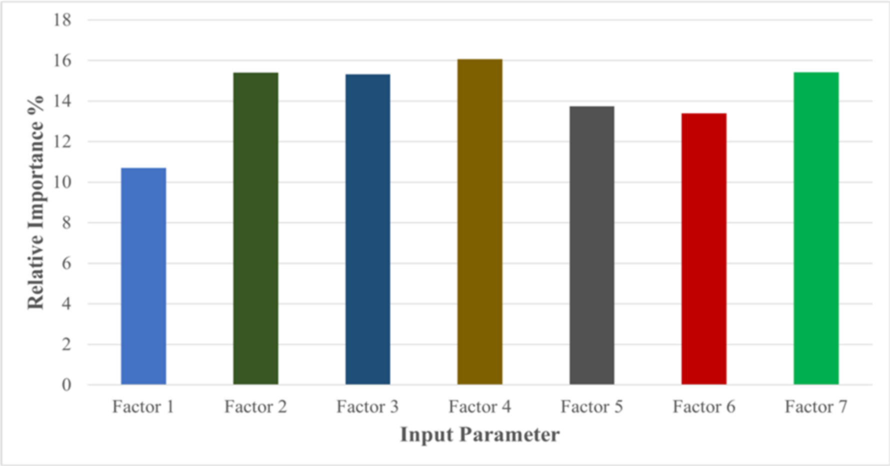

Figure 3. Regression Plots of Neural Network Model Development Phases

4.5. ANN and MLR Model Comparison

The performance of the created neural network prediction model was contrasted to the performance of a MLR model. It demonstrates that the neural network model is greater in terms of prediction accuracy as compared to that of the MLR model. The MAPE for the final neural network model is 0.541% while that of the multiple linear regression model is 2.480%. Figure 4 shows the prediction performance of the final neural network model and the multiple linear regression model [23].

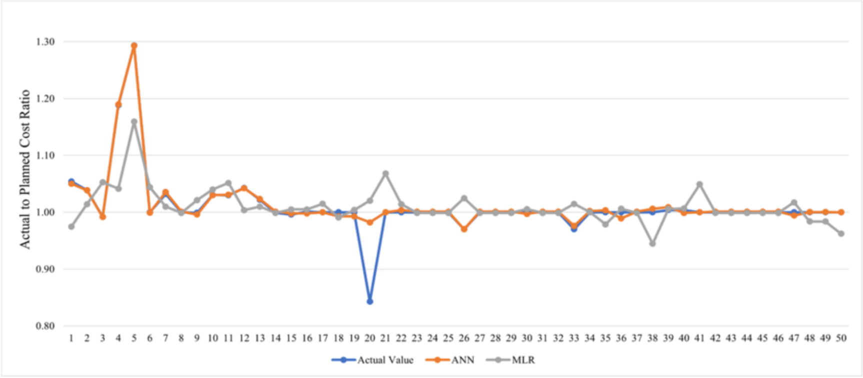

Figure 4. Regression Plots of Neural Network Model Development Phases

5. Conclusion

The purpose of this study is to develop an accurate predictive model for estimating future road project costs based on the project cost and the cost-influencing elements. According to the literature review and from previous research, prediction of construction costs is inaccurate for practical application due to various unanticipated and disruptive phenomena. However, if properly developed they can attain a degree of precision that is advantageous to a variety of applications. Therefore, this research presents a more accurate tool in predicting actual-to-planned cost ratio for road projects. This research provided a mathematical tool in actual-to-planned cost ratio for government agencies and contractors in their preparation for future road projects.

This research has gone through an array of processes to achieve the results of the study. First, the study was able to identify the factors needed through the fifty (50) completed road projects from the DPWH regional office XI. The extracted data indicates that road projects followed the same construction activities, but had various project durations and reasons for variances, which influenced the variance between the projected and actual project costs. Based on the extracted data, the researcher used the Revised Duration and Reasons for Variation Order for each construction activity as factors in applying artificial neural network. This led the researcher to be able to make a record of descriptive statistics of the dataset. It follows that the factors in predicting actual-to-planned cost ratio are statistically significant in predicting actual-to-plan cost ratio. The extracted factors were used in creating the ANN model. After completing the ANN model, the RI of the factor was analyzed using GA. It was then found that the most significant factor to actual-to-planned cost ratio was the sub-base and base course, and the least important factor was the facilities for the engineer.

Lastly, the ANN model was compared to Multiple Linear Regression to test its accuracy in terms of predicting actual-to-plan cost. It showed that the ANN model is superior in terms of predication accuracy with a MAPE of 0.541%. This research concluded that the out of the fifty (50) completed projects of DPWH and seven (7) construction activities, the artificial neural network model can predict its actual-planned-cost ratio with high accuracy.

The researcher only used existing data and not assumptions in formulating the model. The existence of these data is of utmost importance to the development of this model. The study concentrated solely on local data from the DPWH Regional Office XI. Future research studies could opt to use more road projects as samples to improve accuracy and reduce inaccuracies. In addition, to increase the accuracy of the model, a future researcher could use more cost-influencing parameters when estimating the future cost of road improvements. The more detailed a prediction model is, the more it is advantageous to a variety of applications.

References

[1] Wahab, A., & Wang, J. (2021). Factors-driven comparison between BIM-based and traditional 2D quantity takeoff in construction cost estimation. Engineering, Construction and Architectural Management.

[2] Kumar, L., Jindal, A., & Velaga, N. R. (2018). Financial risk assessment and modelling of PPP based Indian highway infrastructure projects. Transport Policy, 62, 2-11.

[3] Johnson, R. M., & Babu, R. I. I. (2020). Time and cost overruns in the UAE construction industry: a critical analysis. International Journal of Construction Management, 20(5), 402-411.

[4] Tayefeh Hashemi, S., Ebadati, O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. SN Applied Sciences, 2(10), 1-27.

[5] Debnath, K. B., & Mourshed, M. (2018). Forecasting methods in energy planning models. Renewable and Sustainable Energy Reviews, 88, 297-325.

[6] Lu, C., Liu, J., Liu, Y., & Liu, Y. (2019). Intelligent construction technology of railway engineering in China. Frontiers of Engineering Management, 6(4), 503-516.

[7] Cabuñas, J. T., & Silva, D. L. (2019). Exploratory Factor-Item Analytic Approach for Construction Project Cost Overrun using Oblique Promax Rotation for Predictors Determination. International Journal of Innovative Technology and Exploring Engineering, 8(6s3), 47-54.
[8] Manahan Malasan, C., S. Villaverde, B., L. Silva, D., & M. de Jesus, K. L. (2021, December). Artificial Neural Network with Sensitivity Analysis: Predicting the Flexural Strength of Concrete Pavement using Locally Sourced Dilapidated Concrete as Partial Replacement. In 2021 5th International Conference on Computer Science and Artificial Intelligence (pp. 408-414).
[9] Waziri, B. S., Bala, K., & Bustani, S. A. (2017). Artificial neural networks in construction engineering and management. International Journal of Architecture, Engineering and Construction, 6(1), 50-60.
[10] Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical systems and signal processing, 20(7), 1483-1510.
[11] Towey, D. (2013). Cost management of construction projects. John Wiley & Sons.
[12] Standard Specifications for Highways, Bridges, and Airports (Vol. 2). (2012). DPWH.
[13] Kulkarni, P., Londhe, S., & Deo, M. (2017). Artificial neural networks for construction management: a review. Journal of Soft Computing in Civil Engineering, 1(2), 70-88.
[14] Alaloul, W. S., Liew, M. S., Wan Zawawi, N. A., Mohammed, B. S., & Adamu, M. (2018). An Artificial neural networks (ANN) model for evaluating construction project performance based on coordination factors. Cogent Engineering, 5(1), 1507657.
[15] LAROZA SILVA, D. A. N. T. E., & MARCELO DE JESUS, K. L. (2020, August). Backpropagation neural network with feature sensitivity analysis: pothole prediction model for flexible pavements using traffic and climate associated factors. In 2020 the 3rd international conference on computing and big data (pp. 60-67).
[16] Monjardin, C. E. F., de Jesus, K. L. M., Claro, K. S. E., Paz, D. A. M., & Aguilar, K. L. (2020, December). Projection of water demand and sensitivity analysis of predictors affecting household usage in urban areas using artificial neural network. In 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) (pp. 1-6). IEEE.
[17] Silva, D. L., de Jesus, K. L. M., Adina, E. M., Mangrobang, D. V., Escalante, M. D., & Susi, N. A. M. (2021, March). Prediction of Tensile Strength and Erosional Effectiveness of Natural Geotextiles Using Artificial Neural Network. In 2021 13th International Conference on Computer and Automation Engineering (ICCAE) (pp. 121-127). IEEE.
[18] Maulion Garduce, C., Laroza Silva, D., & Marcelo de Jesus, K. L. (2021, November). Prediction and Sensitivity Analysis of Shear Strength of Reinforced Concrete Beams with Deformed Hook Steel Fiber using Backpropagation Neural Network coupled with Garson's Algorithm. In 2021 The 5th International Conference on Advances in Artificial Intelligence (ICAAI) (pp. 17-22).
[19] Joh, H., & Malaiya, Y. K. (2014). Modeling skewness in vulnerability discovery. Quality and Reliability Engineering International, 30(8), 1445-1459.
[20] Rikhotso, P. M., & Simo-Kengne, B. D. (2022). Dependence structures between Sovereign credit default swaps and global risk factors in BRICS countries. Journal of Risk and Financial Management, 15(3), 109.
[21] Gunduz, M., & Sahin, H. B. (2015). An early cost estimation model for hydroelectric power plant projects using neural networks and multiple regression analysis. Journal of Civil Engineering and Management, 21(4), 470-477.
[22] Ahmed, A., Ali, A., Elkatatny, S., & Abdulraheem, A. (2019). New artificial neural networks model for predicting rate of penetration in deep shale formation. Sustainability, 11(22), 6527.
[23] Marashi, M., Torkashvand, A. M., Ahmadi, A., & Esfandyari, M. (2017). Estimation of soil aggregate stability indices using artificial neural network and multiple linear regression models. Spanish Journal of Soil Science: SJSS, 7(2), 122-132.