ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 2, pp. 325-335
DOI: 10.33168/JSMS.2022.0216

Predictive Performance of Building Construction Estimation: An Analysis based on ANN Model

SeokHeon Yun

Depart of Architectural Engineering, GyeongSang National University, South Korea

gfyun@gnu.ac.kr

Abstract: Recently, various technologies capable of implementing artificial intelligence have been developed along with the expansion of interest in machine learning. Efforts are also being made to introduce various machine learning technologies in the construction field. The most important thing in the construction field is construction costs, and technologies that can be used practically in predicting construction costs have not been developed. The learning model of construction cost prediction can be composed of various elements, and depending on these factors, the prediction performance is greatly affected. In this study, the influencing factors affecting the prediction of construction costs are derived, and based on these, the change in construction cost prediction performance according to the depth of the Artificial Neural Network (ANN) model and the configuration of the node is analyzed. The construction cost prediction model based on machine learning is expected to help predict and analyze construction costs more accurately and quickly in the future.

Keywords: ANN, cost estimation, model

1. Introduction

Construction costs are affected by various factors. However, until now, there have been many limitations in predicting construction costs in consideration of these various influencing factors. For this reason, in the past, construction costs per unit area are often used as the most common criterion to predict construction costs based on regression analysis. In the case of regression analysis, predictive performance may vary depending on various models, but basically, regression analysis is for the fundamental purpose of analyzing past trends, and if the influence factors vary and the deviation of the result values is large, accuracy will inevitably decreases.

In recent years, with the development of machine learning technology, it has become possible to easily predict construction costs considering various influencing factors depending on the level of data acquisition. Machine learning technology has been established to some extent, and with the development of hardware technology necessary for machine learning, machine learning technology can be used at a practical level. This study attempts to derive an optimal learning model for predicting construction costs based on deep learning technology. To this end, this study will analyze the difference in learning performance according to the structure of the most basic learning model.

2. Literature Review

As utilization of machine learning technology increases, it is being used not only in architecture but also in various fields, and machine learning is being utilized in parallel with big data analysis because it must be used with large amounts of data (Han et al., 2022).

Bowen and Edwards switched from the black box mechanism to a more logical and understandable way, such as the expert system of the late 20th century. It was emphasized that the importance of previous data and expert knowledge should not be ignored in the cost estimation range. In addition, they wanted to integrate the resource allocation system with the current cost model in the future (Bowen and Edwards, 1985).

(Alshamrani 2017) predicted construction projects costs using regression analysis.

(El and Shehatto 2014) analysed construction cost data to extract the main influence input parameters of the fuzzy model. He noted that the use of the double-sided membership function showed better results than other research models. They also suggested that it may be useful to compare the results with other single or combined methods.

(Jin et al., 2012) improved CBR using a Multiple Regression Analysis (MRA) function to improve prediction accuracy. In his study, he suggested considering nominal variables and investigating the cause of the increase in error rates.

(Son et al., 2012) applied a hybrid model of principal component analysis and Support Vector Regression (SVR) and compared it with SVR, ANN, decision tree, and Multiple Linear Regression (MLR), of which eventually showed that the SVR algorithm was excellent.

(Adeli and Wu 1998) considered a normalized neural network during the application of a cost function consisting of standard and normalization errors to simultaneously improve network performance and prevent excessive network mounting.

(Khalaf et al., 2020) applied PSO to estimate the initial construction cost and construction period of 60 construction works. What was inferred in this study is that PSO was well performed with high accurate results and faced with parameters with extensive variability. Another strength of this model is that it is more reliable than projects based on existing projects and based on judgments and experimental cases. This paper investigated a model with a wider range of parameters and tried to apply it to eco-friendly buildings.

(Jiang 2019) studied the application of ANN in cost estimation of construction projects, and as a result of comparing the results with the RBFNN method, ANN showed excellent performance. In addition, research was conducted in a way that optimizes model accuracy and applies it to other types of projects to use different methods for cost factor selection. Finally, the study was conducted in a way that optimizes model accuracy, applies it to different types of projects, and uses different methods for cost factor screening.

(Chandanshive and Kambekar 2019) studied the capabilities of multi-layered feed anterior neural network working model through a backpropagation learning algorithm for estimating the cost of 78 building projects in India, along with testing the effectiveness of the Early Stopping and Bayesian normalization approaches. Along with this, overfitting errors were also studied. He also implemented fuzzy logic to predict the cost of building projects. Because their models are not dynamic in response to market prices, the need for a more agile model is felt. He also used the integration of BP neural networks and genetic algorithms to estimate the cost of residential buildings. The role of GA in the study was to prevent ANN from falling to the local maximum and to improve ANN performance by increasing the convergence rate.

(Alchemosi and Alsaad 2017) reviewed the cost estimation of residential buildings using multi-factor linear regression, which eventually reached an accuracy of about 92%. This study recommended comparing the results with studies implementing neural network technology to identify differences. In fact, this study recommended using a cost estimation model instead of the existing method for construction projects.

(Bala et al., 2014) applied the Back Propagation Artificial Neural Network (BPANN) to predict the cost of construction projects in Nigeria, but this model could only be applied in type of institutional building and other types of buildings or other projects that were difficult to estimate in this way. In addition, it is judged that prediction errors and other evaluation means were not considered as criteria for model performance.

(El-Sawalhi and Shehatto 2014) investigated the factors that had the greatest influence on the cost estimation process, then developed an ANN model, and finally performed sensitivity analysis. He noted that reliable results can be obtained by applying MLP neural networks in the early stages of the project.

Regression analysis was previously the most powerful method applied to cost estimation studies, and methods such as SVM, PSO, RBFNN, and fuzzy ANN were used in building projects. However, in recent years, ANN has been used the most among all methods in predicting construction costs. This shows that the power of the neural network as an artificial intelligence tool is very strong in predicting construction costs. In addition, (Hashemi et al., 2019) showed parameter method as most used case.

(Marzouk and Elkadi 2016)"conducted a study to determine effective variables in the cost estimation process and implement factor variable reduction through exploratory factor analysis (EFA). In their study, the best ANN with an error of nearly 22% was defined.

Looking at the existing research cases, regression analysis was previously used to predict construction costs, but it can be seen that ANN technology has been most used in recent years. However, various detailed methodologies are used among ANNs depending on the type of target facility. There are still many differences in accuracy depending on the target facility, influencing factor, and application methodology, so it seems that research shall be conducted to optimize the configuration of model and parameter to improve prediction accuracy.

3. Machine Learning-based Construction Cost Prediction Performance Analysis

In this study, a learning model was constructed to predict construction costs using machine learning technology, and the learning performance of these models was analyzed.

In machine learning, the Keras library was used in the development environment of Visual Studio Code, and in the construction cost prediction learning model, 251 learning data were used. These data were prepared by extracting only the secured sufficient learning data from 450 cases ordered by the Public Building Corporation for three years from 2017 to 2019. The 11 influencing factors on the construction cost of the building were set, and these were judge to be the factors that determine the basic characteristics and scale of the construction work. The factors influencing the construction cost determined in this way are as follows.

Table 1. Factors influencing construction costs

Influence Factor	Description	Example
Building Classification	Types of facilities classified into 10 types.	Research facilities and institutes; general government buildings and offices; medical and enshrinement facilities; dormitory and residential facilities; fire station, police station, and the like; welfare facilities; libraries and schools; control centers; sports facilities; correctional and training facilities; transportation facilities.
Total Area	The total area of each floor of the building.	
Building Area	The area occupied by the building.	
Total Height	The total height of the building.	
Super Structure Floor	The number of floors on the ground.	
Basement Floor	The number of basement floors.	
Site Total Area	The entire land area.	
Landscape Area	Landscape area.	
Typical Floor Height	Height of the base floor.	
Parking Lot	The number of cars parked.	
Building Year	Construction year.	

The construction cost prediction model defined the four models by dividing the depth of the model and the number of nodes, and the study attempted to analyze the difference in learning performance according to the depth of the model and the number of nodes. In addition, through the analysis results, it was intended to determine the optimal model depth and number of nodes for predicting construction costs. The four models for construction cost prediction were configured as follows, and the same activation function, “Relu”, was applied to these models, and the same number of learning (Epoch) was applied 1000 times.

Table 2. Machine Learning Model Configurations for Cost Estimation

	Node	Level Depth	Activation Function
Model1	6, 200, 100, 64, 32, 16, 1	7	Relu
Model2	6, 100, 64, 32, 16, 1	6	Relu
Model3	6, 100, 64, 32, 1	5	Relu
Model4	6, 32, 16, 1	4	Relu

Fig. 1 shows the results of predicting construction costs for the four learning models. It shows each of the four models, Train Loss, Validation Loss, and Train MAE and Validation MAE. Model 1 consists of the most nodes and depths, and the learning performance of this model is the best. In the case of Model 2, the number of learning increases and the Loss value appears to a level similar to that of Model 1. However, in the case of Model 4, the optimal value was not found in 1,000 studies, but it is judged that the optimal value can be found if the number of studies is increased to more than 3,000 or 4,000 times.

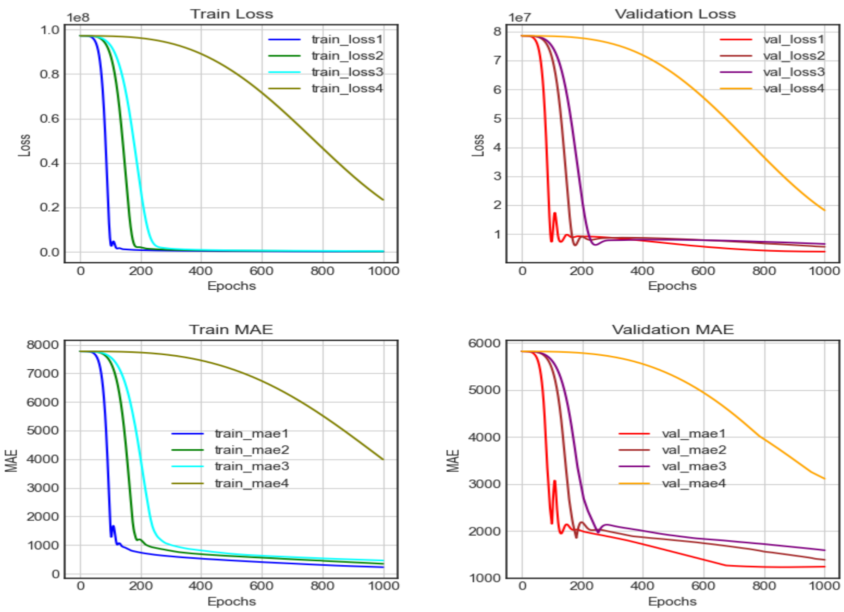

Fig. 1: Train loss and validation loss by models

Fig. 2 shows the comparison of Train and Validation for each model's learning results. Model 1 seems to have the least difference between Train and Validation compared to other models, and it is analyzed that Model 1 shows the closest result.

The Loss value and MAE shown in the learning results show the degree of error in the form of absolute values, and the accuracy of construction costs cannot be determined. To determine the accuracy of the prediction of the construction cost, this study analyzed the error rate of the construction cost by comparing the model's prediction value with the actual construction cost result with the validation data.

Table 3: Average error rate from validation data

	Average Error Rate
Model1	14.92 %
Model2	19.58 %
Model3	15.77 %
Model4	57.62 %

As can be expected, Model 1, which had the best learning performance, had the best prediction error rate of 14.92%.

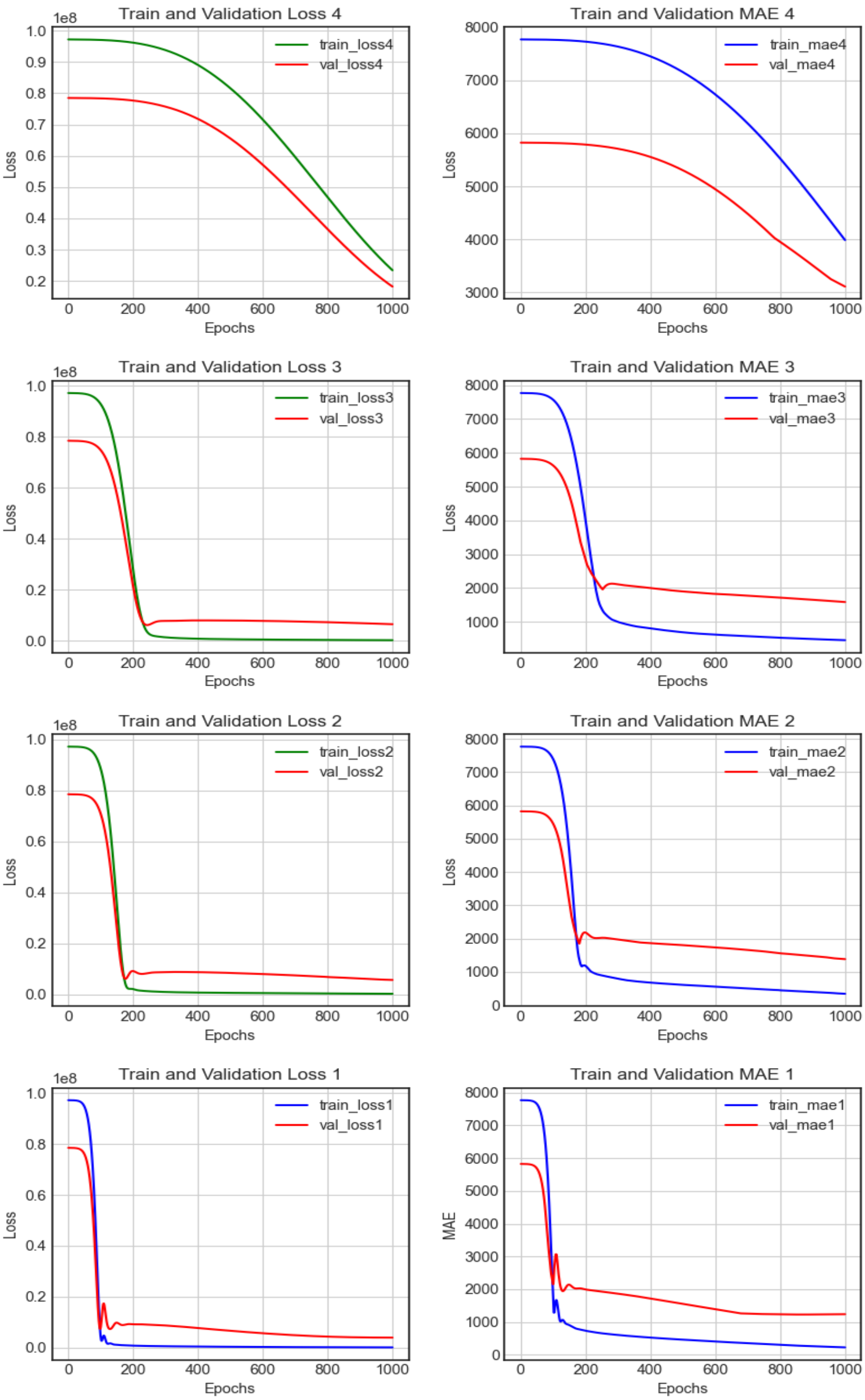

Fig 2: Tran and validation loss

Since the validation data for analyzing the accuracy of construction costs is selected as an arbitrary value, the average error rate was different for each analysis. However, it can be seen that Model 1 had the best relative error rate and Model 4 had the lowest error rate. That does not mean that Model 1 is the best model. In this study, only learning performance according to the depth of nodes and models was compared, and in addition, learning performance varies greatly depending on various Hyper Parameters such as Activation Function and Normalization Method. In the future, it is necessary to derive an optimal construction cost prediction learning model by comprehensively analyzing the learning performance according to this Hyper Parameter.

4. Conclusion

Regression analysis or various approximate methods were used to predict construction costs, but recently, various studies related to prediction of construction costs using ANN have been conducted due to the development of machine learning technology.

In this study, the ANN technique was used to analyze ways to increase the predictive performance of construction costs. There are many ways to increase the predictive performance of construction costs using ANN. First of all, the number and quality of data, the type and configuration of influencing factors, the structure of learning models for machine learning, and the setting of various hyperparameters that can affect the predictive performance of construction costs. This study attempted to derive a learning model suitable for predicting construction costs by analyzing the prediction performance according to the structure of the learning model for predicting construction costs. To this end, in this study, four learning models were defined according to the depth of the model and the number of nodes, and construction cost prediction performance for them was compared.

As a result of the study, the model composed of six or seven levels of depth showed similar learning performance, while the learning model composed of three or four levels of depth showed relatively low predictive performance. In the case of the number of nodes, the predictive performance of a model composed of various nodes, such as 200, 100, 64, 32, and 16, was compared to that of a model composed of 100 or more nodes, under the same epoch conditions. Among the four models defined in the study, the construction cost error rate was 14.92% based on the verification data, and it is believed that the prediction performance can be further improved through the tuning of other hyperparameters.

Acknowledgements

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 22AATD-C163269-02).

References

Adeli H. & Wu M. (1998). Regularization neural network for construction cost estimation. *Journal of Construction Engineering and Management*. 124(1), 18-24. DOI:10.1061/(ASCE)0733-9364(1998)124:1(18).

Alshamrani, O. S. (2017). Construction cost prediction model for conventional and sustainable college buildings in North America. *Journal of Taibah University for Science*. 11(2), 315-32. DOI: 10.1016/j.jtusci.2016.01.004.

Alshemosi, A. M. B. & Alsaad, H. S. H. (2017). Cost estimation process for construction residential projects by using multifactor linear regression technique. *International Journal of Science and Research.* 6(6), 151-156. DOI: 10.30574/ijsra.

Bala, K., Ahmad, B. S., & Shehu, W. B. (2014). A computer-based cost prediction model for institutional building projects in Nigeria: An artificial neural network approach. *Journal of Engineering, Design and Technology*. 12(4), 519-530. DOI:10.1108/JEDT-06-2012-0026.

Chandanshive, V. & Kambekar A. R. (2019). Estimation of building construction cost using artificial neural networks. *Journal of Soft Computing in Civil Engineering.* 3(1), 91-107. DOI: 10.22115/SCCE.2019.173862.1098.

Bowen, P. & Edwards, P. (2006). Cost modelling and price forecasting: practice and theory in perspective. *Construction Management and Economics*. 3(3), 199-215. DOI:10.1080/01446198500000015.

El-Sawalhi, N. I. & Shehatto, O. (2014). A neural network model for building construction projects cost estimating. *Journal of Construction Engineering and Project Management*. 4(4), 9-16. DOI:10.6106/JCEPM.2014.4.4.009.

Han J. Jung D & Kim S. (2022). Predicting the OPS of KBS Batters through Big Data Analysis Using Machine Learning. *Journal of Next-generation Convergence Technology Association*. 6(1), 12-18. DOI:10.33097/JNCTA.2022.06.01.12.

Hashemi, S. T., Ebadati, O. M., & Kaur, H. (2019). Cost estimation and prediction in construction projects:a systematic review on machine learning techniques. *SN Applied Sciences*. 2, 1-27. DOI:10.1007/s42452-020-03497-1.

Jiang, Q. (2019). Estimation of construction project building cost by back-propagation neural network. *Journal of Engineering, Design and Technology.* 18(3), 601-609. DOI:10.1108/JEDT-08-2019-0195.

Jin, R., Cho, K., Hyun, C., & Son, M. (2012). MRA-based revised CBR model for cost prediction in the early stage of construction projects. *Expert Systems with Applications*. 39(5), 5214-5222. DOI:10.1016/j.eswa.2011.11.018.

Khalaf, T. Z., Çağlar, H., Çağlar, A., & Hanoon, A. N. (2020). Particle swarm optimization based approach for estimation of costs and duration of construction projects. *Civil Engineering Journal.* 6(2), 384-401. DOI:10.28991/cej-2020-03091478

Marzouk, M. & Elkadi, M. (2016). Estimating water treatment plants costs using factor analysis and artificial neural networks. *Journal of Cleaner Production*. 112, 4540-4549. DOI:10.1016/j.jclepro.2015.09.015.

Sawalhi, N. I. (2012). Modelling the parametric construction project cost estimate using fuzzy logic. *International Journal of Emerging Technology and Advanced Engineering.* 2(4), 4-8. ISSN 2250-2459

Son, H. Kim, C., & Kim, C. (2012). Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables. *Automation in Construction.* 27, 60-66. DOI:10.1016/j.autcon.2012.05.013.