Terbit *online* pada laman web jurnal : https://jes-tm.org/index.php/jestm/index

Journal of Engineering Science and Technology Management

| ISSN (Online) 2828 -7886 |

JES-TM

Journal of Engineering Science & Technology Management

Article

A Comparative Study of Cost Significant Model and Artificial Neural Networks Methods for River Retaining Wall Cost Estimation in Grobogan Regency

Rizky Tulus Panuwun[1], Henny Pratiwi Adi[2], Soedarsono[3]

[1,2,3]Master's Program in Civil Engineering, Faculty of Engineering, Sultan Agung Islamic University, Semarang

DOI: 10.31004/jestm.v5i2.306

E-mail: rizkytulus@gmail.com (Corresponding author)

ARTICLE INFORMATION

Volume 5 Issue 2
Received: 14 September 2025
Accepted: 28 September 2025
Publish Online: 30 September 2025
Online: at https:JESTM.org/

Keywords

Artificial Neural Networks
Cost Estimation
Cost Significant Model
River Retaining Wall

A B S T R A C T

Grobogan Regency in Central Java Province has a high level of flood risk, so the construction of river retaining walls is an important infrastructure for disaster mitigation. The estimation of construction costs at the early planning stage plays a crucial role in budgeting and technical decision-making. This study aims to compare the accuracy and consistency of two cost estimation approaches: Cost Significant Model (CSM), based on multiple linear regression, and Artificial Neural Networks (ANN) using the backpropagation algorithm. The dataset comprises 42 Bill of Quantity (BoQ) documents (37 training data and 5 testing data), with additional validation conducted through field surveys at seven proposed retaining wall locations. Model performance was evaluated using Mean Absolute Percentage Error (MAPE) to measure accuracy and Bland–Altman Plot to assess consistency. The results indicate that CSM achieved a MAPE value of 1.70%, which is lower than that of ANN, which yielded 2.50%. The Bland–Altman analysis also shows that CSM demonstrates higher consistency, as the linear regression approach allows prediction beyond the training data range, making it more adaptive to actual conditions. In contrast, ANN tends to be constrained within the normalized training data range, reducing its flexibility when encountering new data variations. Therefore, it can be concluded that CSM performs better than ANN in terms of accuracy and consistency in estimating the construction cost of river retaining walls in Grobogan Regency.

1. Introduction

Grobogan Regency is one of the regions in Central Java Province that has a high level of disaster vulnerability. According to Peraturan Bupati Grobogan No. 48 Tahun 2022 tentang Kajian Risiko Bencana Kabupaten Grobogan Tahun 2022–2027, the overall Disaster Risk Index (DRI) value reaches 131.66. Among various types of disasters assessed, floods contribute a score of 17.33, which falls into the high-risk category. This condition requires planned, systematic, and sustainable mitigation efforts, one of which is through the development of flood control infrastructure. Juliastuti et al., (2024) stated that the construction of retaining walls along riverbanks serves as a practical strategy to resist floodwater from spreading into residential areas.

At the early stage of infrastructure planning, construction cost estimation plays an important role as it affects the accuracy of budgeting and the effectiveness of technical decision-making. However, in practice, the initial cost estimation for retaining wall projects implemented by related department generally still relies on simple approaches based on visual assessment, resulting in relatively limited precision. On the other hand, the detailed or bottom-up estimation method can produce more accurate calculations, but its application requires more time and resources, making it less efficient for use during the early planning phase.

Along with the development of technology, various methods for construction cost estimation have been developed. The Cost Significant Model (CSM) is a parametric approach that identifies the most significant cost components using multiple linear regression analysis (Johari & Almuhsy, 2024). On the other hand, Artificial Neural Networks (ANN), inspired by the mechanism of biological neural systems, offer greater flexibility in modeling nonlinear relationships and have the potential to improve estimation accuracy (Dastres & Soori, 2021). However, studies that specifically compare these two methods for river retaining wall construction projects in disaster-prone areas remain limited.

This research gap forms the basis for a comparative study to test the accuracy and consistency of the two construction cost estimation methods. Using historical project data and field validation, this study aims to provide recommendations for the most appropriate estimation method in the context of river retaining wall planning in Grobogan Regency. Practically, the results of this research are expected to support the local government in improving the effectiveness of budget management and data-based disaster mitigation policies. From an academic perspective, this study contributes to the civil engineering literature by integrating conventional and artificial intelligence-based methods in cost estimation analysis.

In line with the above background, this study specifically aims to:

1. Develop and analyze construction cost estimation models for river retaining walls in Grobogan Regency using CSM and ANN.
2. Compare the accuracy and consistency levels of cost estimation results between the CSM and ANN methods to determine the most appropriate approach for the early planning stage of river retaining wall projects.

2. Literature Review

2.1 Construction Cost Estimation

Construction cost estimation is the process of calculating the estimated amount of funds required to carry out a project based on the available design information, even though such information is often not yet final or complete (Holm & Schaufelberger, 2021). Cost estimation plays an important role in the overall project management cycle, as it determines the accuracy of budgeting, resource allocation, and both technical and financial decision-making (Saeidlou & Ghadiminia, 2024).

In general, cost estimation methods can be classified into two main categories: traditional methods and modern methods. Traditional methods include analog estimation, which uses comparisons with similar projects; parametric estimation, which is based on mathematical relationships between project parameters and costs; and detailed (bottom-up) estimation, which breaks down all work components based on quantities and unit prices. On the other hand, modern methods involve the use of Building Information Modeling (BIM), Artificial Intelligence (AI) algorithms, and cloud-based software, which enable more efficient, accurate, and collaborative estimation processes (Reddy Anireddy, 2024).

Although modern methods offer a higher level of precision, their implementation requires initial investment, sufficient data availability, and user technical competence. Therefore, the selection of an estimation method should be adjusted to the planning objectives, data availability, and project complexity. In the context of this study, CSM represents a parametric method as a traditional approach, while ANN represent a modern, artificial intelligence–based method that is relevant for performance comparison in construction cost estimation.

2.2 Cost Significant Model

CSM is a cost estimation method that focuses on identifying the most significant components that consistently contribute dominantly to the total project cost. This concept was introduced by Poh & Horner, (1995) through the analysis of Bill of Quantities (BoQ) data using the Pareto principle, which states that approximately 20% of work items contribute to about 80% of the overall cost. After the significant components (cost significant items) are identified, cost estimation is developed using multiple linear regression analysis to establish a mathematical relationship between independent variables (quantities of significant items) and the dependent variable (total project cost).

The main advantages of CSM lie in its simplicity, efficiency, and ability to produce reasonably accurate estimates for similar projects with standardized work components (Johari & Almuhsy, 2024; Setiawan et al., 2023). However, its limitations include dependence on the quality and availability of historical data, as well as reduced accuracy when applied to projects with a high level of variability. Therefore, although considered a traditional method, CSM remains relevant for use in the early planning stage as a practical reference for construction cost control.

2.3 Artificial Neural Networks

ANN are an artificial intelligence–based approach that imitates the working mechanism of human biological neural systems. ANN are built from interconnected processing units (neurons) linked by weights, with a basic structure consisting of an input layer, a hidden layer, and an output layer (Dastres & Soori, 2021; Ighneiwa et al., 2017). The learning process in ANN generally uses the backpropagation algorithm, which adjusts the weights iteratively based on the difference between the predicted output and the target value, enabling the network to learn complex patterns of relationships among variables (Li, 2024).

Overall, the layered structure of Artificial Neural Networks (ANN) allows data to be processed gradually, starting from raw input and resulting in relevant output for decision-making. The main elements of the network include neurons in the input layer (X_i) that receive external data, neurons in the hidden layer (Z_j) that process signals, and neurons in the output layer (Y_k) that produce the network output. This process is supported by inter-layer weights V_{ij} connecting the input layer to the hidden layer, and W_{jk} connecting the hidden layer to the output layer. In addition, there are bias weights (V_{i0} and W_{k0}) that function respectively in the hidden layer and output layer (Tahapari et al., 2021). The network architecture is presented in Figure 1.

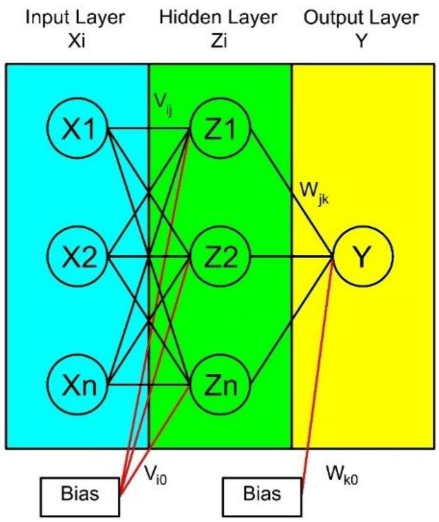

Figure 1 Architecture of ANN Algorithm (Tahapari et al., 2021; Windarto et al., 2020)

The advantage of ANN lies in their flexibility to handle nonlinear data, tolerance to errors, and ability to develop predictive models even when the available information is limited (Ahmad et al., 2022). However, ANN also have several limitations, including a non-transparent decision-making process that is difficult to interpret, high hardware requirements, and uncertainty in training duration (M. Mijwil, 2021). In the context of construction cost estimation, ANN offer high accuracy potential, particularly when the relationships among variables are complex and difficult to represent using conventional approaches.

2.4 River Retaining Walls as Flood Control Infrastructure

Retaining walls are one of the civil engineering structures designed to prevent soil movement and are widely applied in embankment works, excavations, bridge construction, and flood control infrastructure (Pongsagorn et al., 2018). A retaining wall, particularly the gravity-type structure commonly referred to as a *talud* in Indonesia, plays a crucial role in protecting residential areas along riverbanks. In the context of flood mitigation, retaining walls function as barriers to reduce water overflow and strengthen the protection system along river catchment areas (Item et al., 2024).

According to the SNI 8460:2017 tentang Persyaratan Perancangan Geoteknik, the stability of a retaining wall structure is influenced by its self-weight and the supporting soil mass. Therefore, the technical design of retaining walls must consider several key parameters, including maximum water level, subgrade bearing capacity, slope inclination, and the type of construction materials used. These considerations are essential to ensure that the retaining wall performs optimally against hydraulic pressure and external loads, thereby functioning effectively as flood control infrastructure. The typical dimensions of a retaining wall are presented in Figure 2.

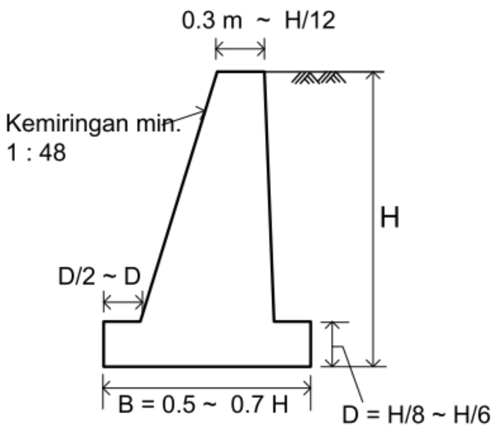

Figure 2. Typical Dimensions of Gravity Retaining Wall
(SNI 8460:2017)

2.5 Evaluation of Estimation Model Performance

The evaluation of estimation model performance is essential to assess the reliability of a method in producing cost predictions that closely reflect actual conditions. One of the most widely used measures is MAPE, which calculates the average relative error between estimated and actual values. This indicator is considered effective because it provides a clear measure of how much the prediction deviates from the real data in percentage form (Khair et al., 2017). The smaller the MAPE value, the higher the accuracy of the estimation model.

In addition to accuracy, consistency is also an important aspect of model evaluation. The Bland–Altman Plot, introduced by Bland & Altman, (1999), is a method used to assess the agreement between two measurement techniques by analyzing the differences between their results relative to their mean. This approach helps identify the presence of systematic bias and variations between the methods being compared.

3. Research Methodology

3.1 Research Design

This study employs a quantitative–comparative approach, focusing on numerical data processing to compare two construction cost estimation methods. This approach was selected because it aligns with the research objective, which is to examine the differences in accuracy and consistency between the two methods in estimating the construction cost of river retaining walls.

Data sampling was carried out using a mixed purposive–convenience sampling method, which combines sample selection based on research relevance and ease of data access. Through this method, the samples used consist of *Bill of Quantity* (BoQ) of river retaining wall projects in Grobogan Regency that meet the criteria of validity, accessibility, and relevance to the research problem.

3.2 Research Data

The research data consist of secondary data in the form of 42 *Bill of Quantity* (BoQ) documents of river retaining wall construction projects in Grobogan Regency, comprising 37 datasets for training and 5 datasets for testing. In addition, primary data obtained from field surveys at seven proposed retaining wall locations were used for validation to ensure the model's consistency with actual site conditions.

The independent variables in this study are the item volumes listed in the BoQ, including clearing and stripping (X1), profiles (X2), mobilization (X3), project signboard (X4), occupational safety and health (X5), soil excavation (X6), backfilling (X7), stone masonry (X8), plastering (X9), cement finishing (X10), weep hole installation (X11), and documentation and as-built drawing (X12). Meanwhile, the dependent variable (Y) represents the total project cost of the river retaining wall, which serves as the output of both CSM and ANN estimation methods.

3.3 Data Processing Methods

The data processing in this study began with the adjustment of cost values in the BoQ documents to account for inflation using the concept of the *time value of money*. This adjustment was made to ensure that all cost data could be analyzed on a comparable price basis. The calculation was carried out using the *Inflation Adjustment Factor* (IAF) formula, expressed as follows:

$$IAF_{t\to T} = \prod_{i=t}^{T-1}(1 + inflation_i) \quad (1)$$

where t = base year, T = projection year, and *inflation* = annual inflation rate in decimal form. The calculation was performed by multiplying the inflation factor for each year, starting from the base year up to the year preceding the projection year. Subsequently, the identification of Cost Significant Items (CSI) was conducted as independent variables in the CSM method to filter the most influential work components contributing to the total project cost.

In the ANN method, data were normalized to a range of 0–1 using the modified *min–max normalization* technique to ensure scale uniformity among variables and to prevent bias during the training process. The normalization formula is as follows:

$$x' = 0.8.\frac{(x - x_{min})}{(x_{max} - x_{min})} + 0.1 \quad (2)$$

Once the ANN output values were obtained, a denormalization process was carried out using the following equation:

$$x = x_{min} + \frac{(x' - 0.1)}{0.8}.(x_{max} - x_{min}) \quad (3)$$

where x = original value, x' = normalized value, x_{min} = minimum variable value in the training data, and x_{max} = maximum variable value in the training data. This formula produces a data distribution within the interval of 0.1 to 0.9. The lower and upper bounds are intentionally not set exactly at 0 and 1 to avoid convergence issues in the sigmoid activation function used in the neural network.

3.4 Analysis Methods

The analytical method in this study was conducted by comparing two construction cost estimation approaches. In CSM, analysis was performed using multiple linear regression to establish a mathematical relationship between the Cost Significant Items (CSI) and the total project cost. A partial hypothesis testing (t-test) was employed to evaluate the significance of each independent variable, while a simultan test (F-test) was used to assess the joint influence of all independent variables. All regression analyses were carried out using the SPSS software.

In contrast, the ANN method was developed using the backpropagation algorithm implemented in MATLAB. The neural network model consisted of an input layer, a hidden layer, and an output layer, with data normalized through the modified *min–max* normalization technique. The training process was performed using the training dataset, while testing was conducted using the testing dataset to obtain the cost estimation results.

4. Results and Discussion

4.1 Research Data Description

The research data used for developing the cost estimation models consisted of 42 BoQ documents from river retaining wall construction projects in Grobogan Regency. These documents were selected based on their validity and relevance to the research objectives. To ensure comparability, all cost values in the BoQ were first standardized using the Inflation Adjustment Factor (IAF) as formulated in Equation 1, thereby representing equivalent cost levels across different project years. The details of the collected BoQ data are presented in Table 1.

Table 1. Summary of Collected BoQ Data

No	Year	Number of Projects	Inflation Rate	IAF to 2025
1	2023	18	2.96	1.0472
2	2024	14	1.71	1.0171
3	2025	10		
Total		42		

In addition to secondary data obtained from BoQ documents, this study also utilized primary data collected through field surveys at seven proposed river retaining wall construction locations in Grobogan Regency. The surveys were conducted using direct measurement methods to determine the required wall length and height at each site, resulting in basic dimensional data that represent the actual construction requirements.

4.2 Cost Estimation Model Using CSM

The analysis using CSM began with the identification of Cost Significant Items (CSI) based on 37 BoQ documents used as training data, with all cost values adjusted using the *time value of money* concept. The fundamental principle of CSI states that approximately 80% of the total project cost is concentrated in 20% of the work items with the largest cost values. The proportion of the cost components is presented in Table 2.

Table 2. Proportion of Cost Components

No.	Work Item	Total Cost (Rp)	Percentage (%)
1	Clearing & striping	35,706,778.86	0.55
2	Profiles	42,227,543.04	0.65
3	Mobilization	66,655,944.04	1.03
4	Project signboard	15,610,750.83	0.24
5	Occupational safety and health	70,274,409.57	1.09
6	Soil excavation	230,904,916.51	3.57
7	Backfilling	21,509,119.56	0.33
8	Stone masonry	5,163,714,377.63	79.75
9	Plastering	429,206,259.20	6.63
10	Cement finishing	238,100,728.46	3.68
11	Weep hole installation	103,931,277.40	1.61
12	Documentation and as-built drawing	57,133,666.32	0.88
13	Total project cost	6,474,975,771.42	100.00

Based on Table 2, the work item with the highest cost proportion is stone masonry, accounting for 79.75% of the total project cost, followed by plastering as the second largest item at 6.63%. Combined, these two components contribute 86.38% of the total project cost, which exceeds the 80% threshold according to the CSI principle. Therefore, the independent variables used in the CSM model are the dimensions of stone masonry (X8) and plastering (X9).

Before performing the regression analysis, the data were tested through a series of classical assumption tests to ensure that the dataset met the statistical prerequisites. These tests were conducted to verify that the regression model is valid, unbiased, and capable of producing reliable cost estimation results.

a. Normality Test

The normality test was conducted to ensure that the residual values were normally distributed. The results of the normality test, presented through the Normal P-P Plot shown in Figure 2, indicate that the residual data points lie close to the diagonal line, suggesting that the residuals are normally distributed.

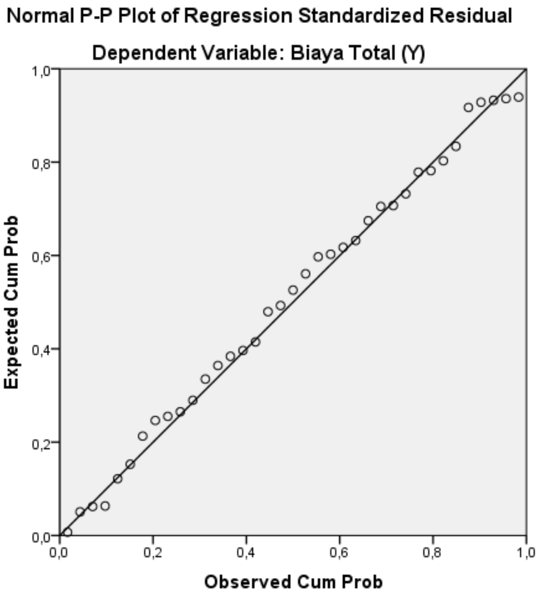

Figure 3 Normal P–P Plot of Regression Standardized Residual

b. Multicollinearity Test

The multicollinearity test was conducted to identify whether strong correlations existed among the independent variables in the regression model. A good dataset should be free from multicollinearity so that each independent variable can influence the dependent variable independently. Data are considered free from multicollinearity if the *Tolerance* value is greater than 0.10 and the *Variance Inflation Factor (VIF)* is less than 10. The results of the multicollinearity test are presented in Table 3.

Tabel 3. Hasil Uji Multikolinearitas

Model		Collinearity Statistics Tolerance	VIF
1	Stone masonry (X8)	0.967	1.034
	Plestering (X9)	0.967	1.034

Dependent Variable: Total project cost (Y)

Based on the results, both independent variables Stone Masonry Work (X8) and Plastering (X9)—have Tolerance values of 0.967 > 0.10 and VIF values of 1.034 < 10. Therefore, it can be concluded that there is no indication of multicollinearity among the independent variables in the regression model.

c. Autocorrelation Test

The autocorrelation test aims to determine whether a correlation exists between the residuals of one observation and those of another within the regression model. A good regression model should be free from autocorrelation to ensure that the estimated parameters are unbiased. The presence of autocorrelation was examined using the Run Test. Data are considered free from autocorrelation if the significance value is greater than 0.05. The results of the Run Test are presented in Table 4.

Tabel 4. Run Test

	Unstandardized Residual
Test Value[a]	321565.00887
Cases < Test Value	18
Cases >= Test Value	19
Total Cases	37
Number of Runs	15
Z	-1.330
Asymp. Sig. (2-tailed)	0.183

a. Median

Based on Table 6, the obtained significance value is 0.183, which is greater than 0.05. This indicates that the residuals are randomly distributed, and therefore, it can be concluded that there is no autocorrelation present in the model.

d. Heteroskedasticity Test

The heteroskedasticity test aims to determine whether there is an inequality of residual variance within the regression model. A good model should be free from heteroskedasticity to ensure consistent parameter estimation. The detection was carried out using the Glejser test, by regressing the absolute residual values on the independent variables. The data are considered free from heteroskedasticity if the significance value is greater than 0.05. The results of the Glejser test are presented in Table 5

Tabel 5. Glejser Test

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	10136801.588	7665114.756		1.322	0.195
	Stone masonry (X8)	-16657.165	40257.304	-.071	-.414	0.682
	Plestering (X9)	-16701.800	15803.451	-.181	-1.057	0.298

Dependent Variable: ABS_RES

Based on Table 5, the significance value for the Stone masonry (X8) is 0.682 and for the Plastering (X9) is 0.298; both values are greater than 0.05. Therefore, it can be concluded that all independent variables show no signs of heteroscedasticity.

After all data met the classical assumption requirements, the next step was to perform multiple linear regression analysis to determine the effect of the independent variables on the dependent variable. The results of the linear regression analysis are presented in Table 6.

Tabel 6. Linear Regression Analysis Results

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	**t**	**Sig.**
1	(Constant)	35161210.746	13170618.823		2.670	0.012
	Stone masonry (X8)	796912.012	69172.299	0.905	11.521	0.000
	Plastering (X9)	79672.028	27154.353	0.231	2.934	0.006

Dependent Variable: Total project cost (Y)

Based on Table 6, the linear regression equation is obtained as follows:

$$\hat{Y} = 35{,}161{,}210.75 + 796{,}912.012X8 + 79{,}672.028X9 \quad (4)$$

where $\hat{Y}$ = Total project cost estimate (Rp), X8 = stone masonry dimension (m^3), dan X9 = plastering (m^2).

Furthermore, Table 6 also shows the results of the partial hypothesis testing (t-test), indicating that for the stone masonry variable (X8), the t-statistic value is 11.521 > t-critical value of 1.69092, and for the Plastering variable (X9), the t-statistic value is 2.934 > t-critical value of 1.69092. Therefore, each independent variable is proven to have a significant effect on the Total Cost (Y).

Subsequently, an F-test was conducted to determine the simultaneous effect of the independent variables on the dependent variable in the regression model.

Table 7. F-Test

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	3.288E+15	2	1.644E+15	66.723	0.000
Residual	8.376E+15	34	2.464E+13		
Total	4.125E+15	36			

Based on the F-test output, an F-statistic value of 66.723 with a significance level of 0.000 was obtained. The F-critical value is 3.28. Since the F-statistic (66.723) > F-critical value (3.28), it can be concluded that the independent variables simultaneously have a significant effect on the dependent variable.

4.3 Cost Estimation Model Using ANN

Prior to data normalization for training input, it is necessary to check for constant features, which are variables with constant values across all observations. Such variables do not provide useful information for the machine learning model in predicting the target (Garg, 2021) and should be removed to avoid unnecessary computations and improve algorithm accuracy (Afshar & Usefi, 2022). In this study, the constant variables consist of work items with a lumpsum (ls) unit, namely mobilization (X3), project signboard (X4), occupational safety and health (X5), as well as documentation and as-built drawings (X12).

After removing the constant variables, the next step is to perform data normalization using Equation 2. The variables used, along with their minimum values, maximum values, and respective normalization results, are presented in Table 8.

During the training process, the ANN model using the backpropagation algorithm to iteratively minimize the error value by adjusting the weights and biases in each network layer. The training results indicate that the network architecture with the smallest MSE value is the 8-3-1 configuration, comprising 8 input variables, 3 neurons in a single hidden layer, and 1 output layer.

The final weights from the input layer to the hidden layer (V_{ij}) along with the bias (V_{io}) are shown in Equations 5 and 6. Meanwhile, the final weights from the hidden layer to the output layer (W_{jk}) along with the bias (W_{ko}) are presented in Equations 7 and 8.

$$V = \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} & v_{16} & v_{17} & v_{18} \\ v_{21} & v_{22} & v_{23} & v_{24} & v_{25} & v_{26} & v_{27} & v_{28} \\ v_{31} & v_{32} & v_{33} & v_{34} & v_{35} & v_{36} & v_{37} & v_{38} \end{bmatrix}$$
$$= \begin{bmatrix} 2{,}0246 & -0{,}45546 & 2{,}4074 & -0{,}21083 & -0{,}48583 & 1{,}0847 & -0{,}074476 & -0{,}019813 \\ -0{,}18082 & 3{,}5251 & -0{,}065785 & -0{,}016526 & 7{,}9409 & -0{,}44996 & 0{,}25409 & 0{,}99666 \\ 0{,}77663 & -2{,}3977 & 0{,}45394 & -0{,}49264 & 3{,}5083 & 0{,}82728 & 0{,}60337 & 0{,}69755 \end{bmatrix} \quad (5)$$

$$V_0 = \begin{bmatrix} v_{10} \\ v_{20} \\ v_{30} \end{bmatrix} = \begin{bmatrix} -2,9518 \\ -1,4518 \\ 0,53886 \end{bmatrix} \quad (6)$$

$$W = [w_{1k} \quad w_{2k} \quad w_{3k}] = [1,2644 \quad 0,83527 \quad 2,3754] \quad (7)$$

$$W_0 = [w_{k0}] = [-2,1759] \quad (8)$$

Thus, the output value (y_k) can be detailed as follows:

$$y_k = \left(W_0 + \sum_{i=1}^{3} Z_i . W_{jk}\right) = w_{k0} + (w_{1k}.Z_1 + w_{2k}.Z_2 + w_{3k}.Z_3) \quad (9)$$

$$Z_i = f\left(\sum_{i=1}^{8} V_{ij}.X_j + V_{i0}\right) \quad (10)$$

where f is the sigmoid activation function (sigmoid logistic), the empirical equation for the ANN method is as follows:

$$\hat{Y} = T_{min} + \left(\frac{y_k - 0,1}{0,8}\right).(T_{max} - T_{min}) \quad (11)$$

where T_{min} = Rp139.343.000,90 and T_{max} = Rp187.781.658,51.

The equation represents a modified min-max denormalization, as per Equation 3. The values T_{min} and T_{max} correspond to the minimum and maximum values of the actual total project cost in the training data.

Tabel 8. Variables Used in the ANN

Variable	Work Item	Dimension	Min (actual)	Max (actual)	Min (norm)	Max (Norm)
X1	Clearing & striping	m^2	0.00	259.20	0,10	0,90
X2	Profiles	m	0.00	208.00	0.10	0.90
X6	Soil excavation	m^3	43.68	122.46	0.10	0.90
X7	Backfilling	m^3	0.00	110.78	0.10	0.90
X8	Stone masonry	m^3	102.96	176.88	0.10	0.90
X9	Plastering	m^2	144.50	291.20	0.10	0.90
X10	Cement finishing	m^2	144.50	291.20	0.10	0.90
X11	Weep hole installation	m	0.00	98.08	0.10	0.90
Y	Total project cost	Rp	139,343,000.90	187,781,658.51	0.10	0.90

Tabel 9. ANN Training Results

No	Number of Neurons	Number of Iterations	MSE Value
1	1 neuron	15	0.00121610
2	2 neurons	11	0.00129080
3	3 neurons	53	0.00053178
4	4 neurons	6	0.00098079
5	5 neurons	7	0.00191870
6	6 neurons	5	0.00093819

4.4 Accuracy Comparison of CSM and ANN

The accuracy assessment in this study was conducted using the MAPE method to compare the actual total cost with the estimated results from both methods. The MAPE formula is defined as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| x100\% \quad (12)$$

where y_i is the actual value, $\hat{y}_i$ is the estimated value, and n is the number of data points. External validation was performed using 5 BoQ documents as test samples. The calculation results are presented in Table 10.

Tabel 10. Analisis Nilai MAPE

No	Project	Actual total project cost (Rp)	Estimate total project cost with CSM (Rp)	Estimate total project cost with ANN (Rp)	APE CSM (%)	APE ANN (%)
1	Package 38	174,908,732.97	170.496.768,63	171.267.504,45	2.52	2.08
2	Package 39	173,631,142.12	173.236.264,03	173.829.352,63	0.23	0.11
3	Package 40	173,530,343.63	169.672.140,13	168.551.694,87	2.22	2.87
4	Package 41	157,776,718.19	163.009.891,56	158.973.688,76	3.32	0.76
5	Package 42	148,911,408.90	149.188.127,53	138.966.514,29	0.19	6.68
				MAPE	**1.70**	**2.50**

4.5 Consistency Comparison of CSM and ANN

The data used were obtained from field measurement surveys, which were subsequently processed to estimate the total cost using both methods. The survey was conducted at seven locations by calculating the quantity required for each work item. The value of each variable for every survey location is displayed in Table 11.

Tabel 11. Work Item Dimensions at the Survey Locations

Locations	Clearing & Striping X1 (m^2)	Profiles (m)	Soil Excavation X6 (m^3)	Backfilling X7 (m^3)	Stone Masonry X8 (m^3)	Plastering X9 (m^2)	Cement Finishing X10 (m^2)	Weep Hole Instalation X11 (m)
Location 1	84.00	42.90	58.80	19.60	157.80	246.29	246.29	60.00
Location 2	72.00	35.55	64.80	21.60	155.80	183.61	183.61	40.00
Location 3	60.00	30.10	60.00	20.00	139.80	146.50	146.50	30.00
Location 4	70.00	36.30	49.00	16.33	131.50	205.24	205.24	50.00
Location 5	120.00	58.80	72.00	24.00	192.00	352.98	352.98	100.00
Location 6	70.00	35.20	70.00	23.33	168.00	177.77	177.77	35.00
Location 7	72.00	36.00	64.80	21.60	158.40	187.51	187.51	40.00

Based on the data in Table 11, which were used as the independent variables, the total cost estimation for each survey location can be calculated by applying both methods. The differences and averages of the two methods for the other locations are presented in Table 12.

Tabel 12. Calculation Results of the Differences and Averages for CSM and ANN

No	Locations	CSM Estimate (Rp)	ANN Estimate (Rp)	Difference (Rp)	Mean (Rp)
1	Location 1	180.536.357,70	165.846.067,05	14.690.290,65	173.191.212,37
2	Location 2	173.948.729,85	160.841.677,27	13.107.052,58	167.395.203,56
3	Location 3	158.241.376,64	141.642.039,77	16.599.336,87	149.941.708,21
4	Location 4	156.307.166,54	145.572.485,39	10.734.681,15	150.939.825,96
5	Location 5	216.291.125,80	172.657.877,64	43.633.248,16	194.474.501,72
6	Location 6	183.205.928,06	168.692.915,38	14.513.012,68	175.949421,72
7	Location 7	176.331.591,44	163.317.993,91	13.013.597,52	169.824.792,68
				$\bar{d}$ = 18.041.602,80	
				SD = 11.428.884,54	

The next step is to calculate the Limits of Agreement (LoA) using these values, as per the following calculation:

$$LoA = \bar{d} \pm 1{,}96\, xSD \quad (13)$$

$$LoA_{upper} = 18.041.602{,}80 + 1{,}96x11.428.884{,}54 = 40.442.216{,}51$$

$$LoA_{lower} = 18.041.602{,}80 - 1{,}96x11.428.884{,}54 = -4.359.010{,}91$$

A scatter plot was subsequently created, as presented in Figure 3.

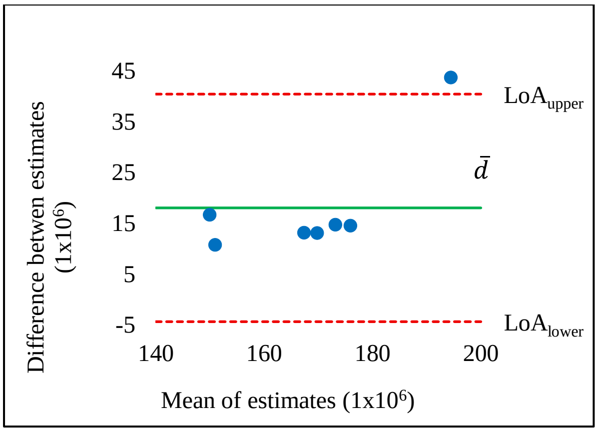

Figure 4 Scatter Plot of the Bland-Altman Plot Method Calculation Results

The analysis results indicate a bias of Rp18,041,602.80, signifying that the ANN method consistently yields lower estimates compared to the CSM method. The standard deviation of the differences was Rp11,428,884.54, showing considerable variation, with Limits of Agreement (LoA) ranging from -Rp4,359,010.91 to Rp40,442,216.51. Out of the seven data points analyzed, six within the limits of agreement, while one data point (the 5th) outside the upper LoA. This discrepancy in the 5th data point is attributed to the dependent variable value exceeding the maximum range of the training data. In the modified min-max normalization process, input values outside this range cannot be proportionally projected, causing the ANN prediction results to be constrained within the training data range and resulting in a substantially larger difference.

5. Conclusions and Recommendations

5.1 Conclusions

Based on the analysis, it can be concluded that the CSM and ANN methods possess distinct characteristics. CSM is a conventional approach based on linear regression, utilizing significant items; in this study, the significant items were stone masonry (X8) and plastering (X9). In contrast, ANN is a modern machine learning-based approach, for which the optimal configuration in this study was 8-3-1 (8 input variables, 3 neurons in the hidden layer, and 1 output layer).

Regarding accuracy, CSM demonstrated better performance with a MAPE value of 1.70% compared to ANN's 2.50%. Furthermore, CSM proved more consistent. The linear regression model is inherently transparent, enabling it to generate estimates beyond the range of the training data and offering greater flexibility for actual conditions. Conversely, ANN, which uses min-max scaling normalization, tends to constrain estimates within the range of the training data, making it less adaptive to data variations outside this range.

5.2 Recommendations

Future research is recommended to use other normalization methods for ANN, such as z-score standardization or robust scaling, to enhance the model's adaptability to varying data distributions and prevent its confinement to a min-max range. Furthermore, expanding the scope of the study by comparing other intelligent computational methods, such as Fuzzy Logic, Support Vector Regression (SVR), or Random Forest Regression, would provide a more comprehensive understanding of the performance of various cost estimation approaches. Increasing the volume of historical project data with a wider range of total costs is also crucial to improve the reliability of the ANN model and strengthen the generalizability of the estimation results.

References

Afshar, M., & Usefi, H. (2022). Optimizing feature selection methods by removing irrelevant features using sparse least squares. *Expert Systems with Applications*, *200*(August), 116928. https://doi.org/10.1016/j.eswa.2022.116928

Ahmad, M., Al Mehedi, M. A., Yazdan, M. M. S., & Kumar, R. (2022). Development of Machine Learning Flood Model Using Artificial Neural Network (ANN) at Var River. *Liquids*, *2*(3), 147–160. https://doi.org/10.3390/liquids2030010

Badan Standardisasi Nasional. (2017). *SNI 8460:2017 Persyaratan Perancangan Geoteknik*.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Dastres, R., & Soori, M. (2021). Artificial Neural Network Systems. *International Journal of Imaging and Robotics (IJIR)*, *21*(2), 13–25. https://hal.science/hal-03349542v1

Garg, S. (2021). *Dropping Constant Features using VarianceThreshold: Feature Selection -1*. https://medium.com/nerd-for-tech/removing-constant-variables-feature-selection-463e2d6a30d9

Holm, L., & Schaufelberger, J. E. (2021). *Construction Cost Estimating*. Routledge. https://doi.org/10.1201/9781003023494

Ighneiwa, I., Hamidatou, S., & Ben Ismael, F. (2017). Using Artificial Neural Networks (ANN) to Control Chaos. *Libyan Journal for Engineering Research*, *1*(1), 46–49. https://doi.org/10.37376/lyjer.v1i1.374

Item, R., Passambuna, B., & Rumitra, M. (2024). Perencanaan Talud Sungai Pengendali Banjir Di Kelurahan Otomona, Kabupaten Mimika. *Jurnal Teknik AMATA*, *5*(1), 51–55.

Johari, G. J., & Almuhsy, M. R. (2024). Penerapan Metode Cost Significant Model pada Estimasi Biaya Pembangunan Peningkatan Jalan. *Jurnal Konstruksi*, *22*(1), 1–12. https://doi.org/10.33364/konstruksi/v.21-2.1427

Juliastuti, Y., Wijayanti, Y., Fajar, M., & Anda, M. (2024). Hydraulic Modeling-Based Design of Retaining Wall Height for Flood Mitigation. *Journal of Engineering and Sustainable Development*, *28*(6), 710–716. https://doi.org/10.31272/jeasd.28.6.3

Khair, U., Fahmi, H., Hakim, S. Al, & Rahim, R. (2017). Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error. *Journal of Physics: Conference Series*, *930*(1), 0–6. https://doi.org/10.1088/1742-6596/930/1/012002

Li, M. (2024). Comprehensive Review of Backpropagation Neural Networks. *Academic Journal of Science and Technology*, *9*(1), 150–154. https://doi.org/10.54097/51y16r47

M. Mijwil, M. (2021). Artificial Neural Networks Advantages and Disadvantages. *Mesopotamian Journal of Big Data*, *2021*, 29–31. https://doi.org/10.58496/MJBD/2021/006

Pemerintah Kabupaten Grobogan. (2022). *Peraturan Bupati Grobogan No. 48 Tahun 2022 Tentang Kajian Risiko Bencana Kabupaten Grobogan Tahun 2022–2027*.

Poh, P. S. H., & Horner, R. M. W. (1995). Cost-significant modelling - its potential for use in south-east Asia. *Engineering, Construction and Architectural Management*, *2*(2), 121–139. https://doi.org/10.1108/eb021007

Pongsagorn, P., Attaphol, B., & Sorasak, S. (2018). Behaviour Of Dry-Retaining Wall. *International Journal of GEOMATE*, *15*(47), 186–192. https://doi.org/10.21660/2018.47.GTE84

Reddy Anireddy, A. (2024). Comparative Analysis of Traditional vs. Modern Estimating Methods: Pros and Cons of Different Approaches in Construction.

International Journal of Science and Research (IJSR), *13*(11), 1–5. https://doi.org/10.21275/SR241023231113

Saeidlou, S., & Ghadiminia, N. (2024). A construction cost estimation framework using DNN and validation unit. *Building Research and Information*, *52*(1–2), 38–48. https://doi.org/10.1080/09613218.2023.2196388

Setiawan, Y., Yelvi, Broto, A. B., Rafli, M. M., Supriyadi, I., & Salimah, A. (2023). Estimating the Cost of Strauss Pile Retaining Wall Works Using the Cost Significant Method. *IOP Conference Series: Earth and Environmental Science*, *1173*(1), 012062. https://doi.org/10.1088/1755-1315/1173/1/012062

Tahapari, Y., Nugroho, A. S. B., & Suparma, L. B. (2021). Model Estimasi Biaya Dengan Cost Significant Model Dan Artificial Neural Network Proyek Peningkatan Jalan Aspal Di Yogyakarta. *Jurnal Teknik Sipil*, *16*(2), 122–133. https://doi.org/10.24002/jts.v16i2.4778

Windarto, A. P., Nasution, D., Wanto, A., Tambunan, F., Hasibuan, M. S., Siregar, M. N. H., Solikhun, Lubis, M. R., Fadhillah, Y., & Nofriansyah, D. (2020). *Jaringan Saraf Tiruan: Algoritma Prediksi dan Implementasi*. Yayasan Kita Menulis.