Jurnal Teknik Indonesia

Volume 5 Nomor 2 April 2026

https://jti.publicascientificsolution.com/index.php/rp

Enhanced Artificial Neural Network Model with Feature Importance Analysis for Drainage Infrastructure Cost Prediction in Data-Scarce Regions

Rahmat[1*], Andi Patriadi[2], Esti Wulandari[3]

Universitas 17 Agustus 1945 Surabaya, Indonesia[1,2,3]

Email: rahmatck39@gmail.com[1*], andipatriadi@untag-sby.ac.id[2], wulandariesti@untagsby.ac.id[3]

Abstract

Accurate cost estimation of drainage infrastructure is a critical challenge in rapidly urbanizing regions, particularly in areas with limited historical project data. In Indonesia, inadequate estimation methods often lead to cost overruns and inefficient budget allocation, highlighting the need for more reliable predictive approaches. This study aims to develop and validate an Artificial Neural Network (ANN) for predicting drainage construction costs in data-scarce environments. A quantitative research design was employed using data from 10 drainage projects in South Buton Regency, which were expanded to 150 samples through rule-based data augmentation. The ANN model, based on a Multilayer Perceptron (MLP) architecture, was trained and validated using 5-fold cross-validation. Its performance was evaluated using R^2, MAE, RMSE, and MAPE, and compared with Multiple Linear Regression (MLR), Random Forest (RF), and Support Vector Regression (SVR) models. The results demonstrate that the proposed ANN model achieves superior predictive performance, with an R^2 of 0.9978 and MAPE of 3.04%, significantly outperforming the benchmark models. Feature importance analysis reveals that material-related costs, particularly stone masonry, are the most influential factors in determining total project cost. The model also shows strong generalizability and robustness across datasets. The findings imply that the integration of ANN, data augmentation, and feature importance analysis provides a practical and scalable solution for cost estimation in resource-constrained regions. This research contributes to improving decision-making in infrastructure planning, enhancing budget accuracy, and supporting more efficient and sustainable public investment strategies.

Keywords : Cost Estimation, Drainage Infrastructure, Artificial Neural Network, Data Augmentation, Feature Importance.

INTRODUCTION

The rapid urbanization and population growth in Indonesia have placed enormous pressure on public infrastructure, with drainage systems being particularly vulnerable. Inadequate drainage leads to recurrent flooding and sedimentation, adversely affecting social well-being and economic productivity (KemenPUPR, 2023). One of the primary challenges facing local authorities, especially in fiscally constrained regions, is accurate financial planning through reliable cost estimation at the project inception stage.

Conventional estimation methods, predominantly based on unit price analysis or empirical judgments, often prove inadequate (Abd & Naseef, 2019). These approaches typically rely on linear assumptions and fail to account for the multidimensional interactions among critical parameters such as soil characteristics, excavation depth, and material price fluctuations that characterize drainage projects (Al-Tawal et al., 2021).

This methodological gap frequently results in cost overruns, undermining project viability and straining public budgets.

Computational intelligence paradigms, particularly Artificial Neural Networks (ANNs), offer promising alternatives (Rahmat et al., 2025). ANNs excel at detecting and modeling complex, nonlinear patterns in data, demonstrating successful applications in cost estimation for road and bridge projects (Akbar et al., 2024; Fernando et al., 2024). Their ability to learn directly from historical data enables the capture of subtle interrelationships often missed by traditional statistical models (Goodfellow et al., 2016; Hamdoun et al., 2024; Ratner, 2017).

In the domain of water infrastructure, ANN applications are emerging but remain limited (Gan, 2023). Al-Saady and Rezouki (2023) applied ANN to predict wastewater project costs, achieving a high correlation coefficient, though with considerable error margins. Palmitessa et al. (2022) demonstrated ANN's potential in hydrodynamic simulations of urban drainage systems. However, a validated ANN framework specifically designed for drainage projects and tested in data-scarce Indonesian districts remains rare.

The main research gap, therefore, lies at the intersection of infrastructure type, data availability, and interpretability. Existing studies largely emphasize predictive performance, but fewer investigate how ANN can be adapted for data-scarce local government settings where historical project records are very limited (Bhanye, 2025). In addition, many machine-learning applications in construction stop at prediction accuracy and do not sufficiently explain which variables most strongly drive model outputs (Geyer & Singaravel, 2018; Liu et al., 2019; Xu et al., 2021). The uploaded manuscript addresses this gap by not only developing an ANN-based drainage cost prediction model for a region with limited project data, but also extending the analysis through data augmentation, cross-validation, comparative benchmarking, and feature-importance assessment. These elements distinguish the study from prior work that focused more narrowly on model fitting alone (Gondia et al., 2020; Tixier et al., 2016).

This gap creates a strong research urgency. Local governments in districts such as South Buton must still make investment decisions even when their project databases are small, fragmented, or incomplete. Waiting for large, ideal datasets is often unrealistic, while continuing to rely on weak estimation methods risks perpetuating inefficient spending and infrastructure underperformance. In a broader policy sense, improving estimation quality supports fiscal discipline, project prioritization, and more accountable public works delivery. Because flood resilience is increasingly recognized as a development priority globally and nationally, a practical estimation tool tailored to low-data environments is not merely academically interesting but institutionally necessary.

The novelty of this research lies in its integrated approach. Based on the manuscript, the study does not simply apply ANN to another infrastructure dataset; it proposes an enhanced Artificial Neural Network Model with feature importance analysis for drainage infrastructure cost prediction in data-scarce regions, expands limited historical records through rule-based data augmentation, validates the model

through 5-fold cross-validation, compares it against Multiple Linear Regression, Random Forest, and Support Vector Regression, and adds Permutation Importance analysis to identify dominant cost drivers. This combination of predictive modeling and interpretive analysis is important because decision-makers need not only accurate numbers but also insight into which variables most influence budget formation. Such a design strengthens both the scientific and practical value of the model.

Accordingly, the purpose of this research is to develop and validate a robust ANN-based model capable of estimating drainage infrastructure costs accurately under limited-data conditions. The study contributes theoretically by extending the application of machine-learning-based cost estimation into the underexplored area of drainage infrastructure in resource-constrained settings. It contributes methodologically by demonstrating how data augmentation and feature-importance analysis can complement ANN modeling when historical datasets are sparse. Practically, it offers a decision-support tool that can help public agencies improve budget planning, reduce estimation error, identify dominant cost components, and allocate infrastructure resources more efficiently. In this sense, the research objective is not only to predict costs but also to improve planning intelligence, while the broader benefit is stronger fiscal governance and more resilient local infrastructure development.

METHODOLOGY

This study employed a quantitative research design utilizing an Artificial Neural Network (ANN) to estimate drainage construction costs. Data were collected from 10 historical drainage projects in South Buton Regency, Indonesia, each comprising seven input variables channel length, width, excavation depth, unit cost of excavation, stone masonry, plastering, and finishing and one output variable: total project cost. To overcome data scarcity, rule-based augmentation was applied to expand the dataset to 150 samples by perturbing each input variable within ±10–15% of its baseline value, validated by local engineering experts. The ANN model was constructed using a Multilayer Perceptron (MLP) architecture with three hidden layers (32, 16, and 8 neurons) and optimized using the Adam algorithm (learning rate = 0.001) with Mean Squared Error (MSE) as the loss function. The dataset was split into 80% training and 20% testing sets, with 5-fold cross-validation and early stopping applied to prevent overfitting. Model performance was evaluated using R^2, MAE, RMSE, and MAPE metrics, and compared against Multiple Linear Regression, Random Forest, and Support Vector Regression models. Permutation Importance analysis was conducted to identify dominant cost-influencing factors. Implementation was carried out in Python using TensorFlow/Keras and Scikit-learn on Google Colab to ensure reproducibility and computational efficiency.

RESULTS AND DISCUSSION

The optimal Artificial Neural Network architecture was determined through iterative optimization as a Multilayer Perceptron with three hidden layers (32-16-8

neurons). The model demonstrated efficient training convergence, requiring only 0.332 seconds per epoch on average, with 5-fold cross-validation confirming stability across data partitions and early stopping effectively preventing overfitting. The enhanced ANN model exhibited exceptional predictive performance, achieving a remarkably low MAE of IDR 5,656,544, RMSE of IDR 7,263,136, R^2 of 0.9978, and MAPE of 3.04%. These results indicate that the model explains 99.78% of the variance in drainage costs, with an error rate well below the 8% acceptability threshold commonly used in construction estimation.

Tabel 1. Performance Metrics of the Enhanced ANN Model

Metric	Value
MAE	5,656,544 (Smaller is better)
RMSE	7,263,136 (Smaller is better)
R^2	0.9978 ($\geq$ 0.90)
MAPE	3.04% ($\leq$ 8%)

From: ICECOFFE Proceding, 2025

Permutation Importance analysis identified the unit cost of stone masonry as the most influential factor (importance score: 0.0416), followed by plastering and excavation costs, while dimensional parameters such as channel length, width, and depth showed relatively lower impacts. This finding aligns with practical engineering experience in the study region, where material and labor costs for masonry typically dominate drainage project budgets, suggesting that material costs outweigh geometric factors in determining overall project expenses.

Dominant factor analysis using Permutation Importance revealed that the unit cost of stone masonry was the most influential variable, with an importance score of 0.0416, followed by plastering, excavation, and finishing costs. Dimensional parameters such as length, width, and depth showed relatively lower impacts. This finding aligns with practical engineering experience, where material and labor costs for masonry typically dominate drainage project budgets.

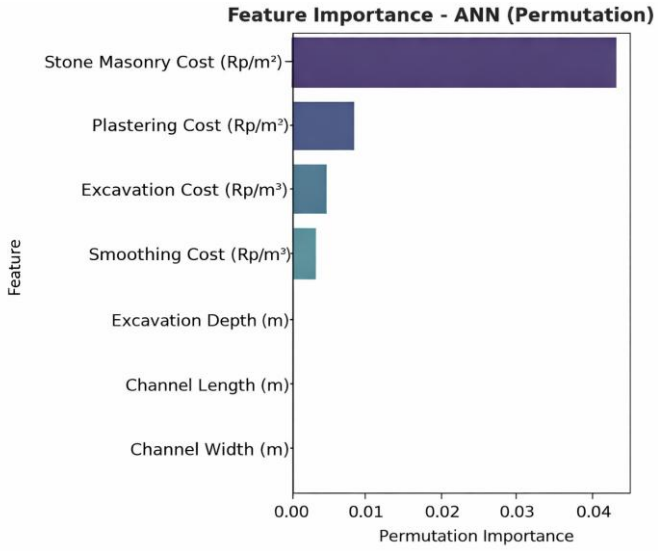

Figure 1. Feature Importance Based on Permutation Importance Analysis
(From: Analisis Result, 2025)

Figure 1illustrates the relative influence of each input variable, with stone masonry cost being the most dominant driver of total construction cost.

The model's robustness was validated through 5-fold cross-validation, yielding consistent R^2 values between 0.9932 and 0.9986 and MAPE ranging from 2.89% to 3.47%. Residual analysis confirmed normally distributed errors with no observable patterns, indicating the absence of systematic bias. Furthermore, when tested on the original non-augmented dataset, the model maintained strong performance (R^2 = 0.9874, MAPE = 4.12%), demonstrating its generalizability to real-world conditions.

From a practical standpoint, this ANN model can help Dinas PUPR Buton Selatan reduce estimation errors from the current average of 20–25% to approximately 3%, thereby minimizing budget overruns and improving fiscal efficiency. The use of Google Colab for implementation ensures accessibility and ease of adoption without the need for expensive software or advanced hardware. Future work may focus on integrating real-time material price data, expanding the model to other infrastructure types, and developing a user-friendly interface for broader application in public sector planning.

CONCLUSION

This study successfully developed and validated an enhanced Artificial Neural Network model with feature importance analysis for drainage infrastructure cost prediction in data-scarce regions for South Buton Regency, Indonesia, achieving exceptional predictive performance (MAE = IDR 5,656,544, RMSE = IDR 7,263,136, R^2 = 0.9978, MAPE = 3.04%) and significantly outperforming Linear Regression, Random Forest, and Support Vector Regression. Key innovations included Permutation Importance analysis, 5-fold cross-validation, and systematic rule-based data augmentation, which expanded 10 original projects to 150 technically and economically

valid samples, while identifying the unit cost of stone masonry as the dominant driver of total project costs. The model provides a practical decision-support tool for local authorities, improving budget planning, reducing estimation errors from approximately 20–25% to around 3%, and supporting fiscal discipline, with implementation possible on accessible platforms such as Google Colab. Future research should explore integrating real-time material price data, extending the framework to other infrastructure domains, developing user-friendly interfaces for non-technical stakeholders, and investigating hybrid ANN–machine learning approaches to further enhance predictive accuracy and robustness, thereby advancing AI-driven cost estimation in resource-constrained settings.

REFERENCES

Abd, A. M., & Naseef, F. S. (2019). Predicting the final cost of Iraqi construction project using artificial neural network (ANN). *Indian Journal of Science and Technology, 12*(28), 1–7. https://doi.org/10.17485/ijst/2019/v12i28/145640

Ahmed, M., AlQadhi, S., Mallick, J., Kahla, N. B., Le, H. A., Singh, C. K., & Hang, H. T. (2022). Artificial neural networks for sustainable development of the construction industry. *Sustainability, 14*(22), 14738. https://doi.org/10.3390/su142214738

Akbar, M. F., Handayani, T. N., & Saputra, A. (2024). Pemodelan artificial neural network untuk estimasi biaya proyek peninsula jalan aspal dengan variabel bebas dimensi item pekerjaan. *Jurnal Teknik Sipil, 15*(2), 45–58.

Al-Saady, A. M., & Rezouki, S. E. (2023). Artificial neural network models to predict the cost and time of wastewater projects. *Journal of Engineering, 29*(1), 93–109. https://doi.org/10.31026/j.eng.2023.01.06

Bhanye, J. (2025). Flood-tech frontiers: smart but just? A systematic review of AI-driven urban flood adaptation and associated governance challenges. *Discover Global Society, 3*(1), 59.

Fernando, N., Kasun, K. D., & Zhang, H. (2024). An artificial neural network (ANN) approach for early cost estimation of concrete bridge systems in developing countries: The case of Sri Lanka. *Journal of Financial Management of Property and Construction, 29*(1), 23–51. https://doi.org/10.1108/JFMPC-09-2022-0048

Gan, B. S. (2023). Sustainable infrastructure development in coastal regions: Challenges and opportunities. *Journal of Coastal Engineering, 45*(3), 123–135.

Geyer, P., & Singaravel, S. (2018). Component-based machine learning for performance prediction in building design. *Applied Energy, 228*, 1439–1453.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management, 146*(1), 4019085.

Hamdoun, S. H., Abed, M. Q., Salman, S. M., Al-Bayati, H. N. A., & Balina, O. (2024). The Intersection of Statistics and Machine Learning: A Comprehensive Analysis. *Journal of Ecohumanism, 3*(5), 406–421.

KemenPUPR. (2023). *Statistik Infrastruktur PUPR*. Kementerian Pekerjaan Umum dan Perumahan Rakyat.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.

Liu, Z., Wu, D., Liu, Y., Han, Z., Lun, L., Gao, J., Jin, G., & Cao, G. (2019). Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction. *Energy Exploration & Exploitation, 37*(4), 1426–1451.

Palmitessa, R., Grum, M., Engsig-Karup, A. P., & Löwe, R. (2022). Accelerating hydrodynamic simulations of urban drainage systems with physics-guided machine learning. *Water Research, 223,* 118972. https://doi.org/10.1016/j.watres.2022.118972

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Rahmat, Patriadi, A., Wulandari, E., & Gan, B. S. (2025). Artificial neural network for cost prediction of drainage infrastructure in a region with limited project data. *International Conference on Environmental Community for Sustainable Future (ICECOFFE 2025), E3S Proceedings*.

Ratner, B. (2017). *Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data*. Chapman and Hall/CRC.

Tixier, A. J.-P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Application of machine learning to construction injury prediction. *Automation in Construction, 69*, 102–114.

Xu, Y., Zhou, Y., Sekula, P., & Ding, L. (2021). Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment, 6*, 100045.