



## Estimation of Delay in Prefabricated Projects Using Modern Machine Learning Approaches (Case Study: Baghdad)

Mansoor HM<sup>1\*</sup>, Muter RA<sup>2</sup>, Mutar AA<sup>3</sup>, Fadel ZE<sup>4</sup>

DOI:10.5281/zenodo.18013341

<sup>1\*</sup> Hayder M. Mansoor, Department of Civil Engineering, Azad Islamic University, Qazvin, Iran.

<sup>2</sup> Ruqaya A. Muter, Department of Electrical Engineering Techniques, College of Engineering and Technology, Al-Mustaqbal University, Hillah, Iraq.

<sup>3</sup> Ali Ahmed Mutar, College of Cyber Security, Asia Pacific University of Technology and Innovation, Malaysia.

<sup>4</sup> Zahraa Emad Fadel, Electrical Techniques Engineering Department, Technical College Al-Musaib, Al-Furat Al-Awsat University, Hilla, Iraq.

We analyzed the prior literature on estimating delays in project time in the presentation section, and we found that data uncertainty could be minimized using generative adversarial network (GAN) to augment our dataset, which uses data to produce findings that mimicked actual world circumstances. We organized the findings accordingly. In the initial finding, we utilized four (4) algorithms on a dataset of twenty-one features containing 284,807 transactions i.e. multilayer perceptron (MLP) neural network, support vector machine (SVM), decision tree, and k-nearest neighbor (KNN). The findings established that MLP neural network produced the largest accuracy value of (90.72%), followed with SVM (78.43%), Decision Tree (77.64%), and KNN (74.5%).

Next, the GAN was used to augment the dataset to a total of 400,00 transactions, allowing the augmented dataset to result in a number of delay samples of 609. The four (4) algorithms were subsequently re-evaluated with the expanded dataset to classify and identify project delays in the dataset. The results indicated that augmentation using GAN enhanced the accuracy of the models overall. From the first process, using the MLP neural network reached an accuracy of 98.76% and SVM was 82.03%, decision tree was 80.31% and KNN was 79.95%.

**Keywords:** Estimation, Delay, Project, Prefabricated, Utilization, New Approaches, Machine Learning, Baghdad

Corresponding Author	How to Cite this Article	To Browse
Hayder M. Mansoor, Department of Civil Engineering, Azad Islamic University, Qazvin, Iran. Email: <a href="mailto:haalmohana4@gmail.com">haalmohana4@gmail.com</a>	Mansoor HM, Muter RA, Mutar AA, Fadel ZE, Estimation of Delay in Prefabricated Projects Using Modern Machine Learning Approaches (Case Study: Baghdad). Int J Engg Mgmt Res. 2025;15(6):24-32. Available From <a href="https://ijemr.vandanapublications.com/index.php/j/article/view/1823">https://ijemr.vandanapublications.com/index.php/j/article/view/1823</a>	

<b>Manuscript Received</b> 2025-11-03	<b>Review Round 1</b> 2025-11-18	<b>Review Round 2</b>	<b>Review Round 3</b>	<b>Accepted</b> 2025-12-05
<b>Conflict of Interest</b> None	<b>Funding</b> Nil	<b>Ethical Approval</b> Yes	<b>Plagiarism X-checker</b> 4.39	<b>Note</b>



## 1. Introduction

The most important indicator of a project's success, in addition to cost-effectiveness, is its completion within the planned timeframe. In other words, finishing a project on schedule is one of the key criteria for determining its success. However, several issues—such as inappropriate allocation procedures in industrial plans and credit rules, inflation, insufficient financial resources, procurement and contractor execution problems, sharp price fluctuations, inadequate workshop management, and inaccurate estimation of consultant workload, materials, equipment, and project duration—can extend the completion time to more than twice the originally planned schedule. In some cases, this delay results directly from inaccurate initial time predictions. Therefore, developing accurate completion time estimation methods, grounded in the historical impact of various influencing factors on similar projects, is of great importance. For power plant projects in particular, the issue is even more critical, as their sensitive nature, high costs, and significant investment requirements mean that any delay can lead to substantial cost increases. Consequently, identifying and analyzing the factors that contribute to delays in such projects is essential, highlighting the importance of this research (Fleischman & Seeber, 2016).

## 2. Literature Review

### 2.1 Prefabricated

Structures can generally be divided into two main groups:

**1. Structural members** – These are usually reinforced, pre-stressed, or post-tensioned elements. The pre-stressed parts themselves can be classified into two categories.

**2. Components of prefabricated structures** – These typically include floors, slabs, roofs, columns, beams, infill walls, piles, and stairs (Mohammadipour & Sadjadi, 2016).

### 2.2 Project Management

A project is a temporary endeavor undertaken to deliver a unique product, service, or result. Its success depends not only on cost-effectiveness but also on timely completion, which is often regarded as a critical measure of performance.

However, several challenges (Mohammadipour & Sadjadi, 2016)—such as inefficient allocation procedures, financial limitations, inflation, contractor-related issues, material shortages, price fluctuations, and inadequate planning or workload estimation—can significantly delay completion, sometimes doubling the originally planned duration. In many cases, such delays stem from inaccurate time estimation at the project's outset.

Accurate prediction of completion time, based on lessons learned from similar projects and the influence of multiple factors, is therefore essential for achieving timely delivery. This is particularly critical in power plant projects, where the high costs, large investments, and sensitive nature of operations mean that delays can lead to substantial financial and operational losses. Identifying and analyzing the key factors that contribute to delays thus becomes a necessity, underscoring the importance of this research (Suk et al., 2017).

### 2.3 Major Causes of Delays and Failures during the Project

Project delivery delays are among the most common causes of complications in construction, particularly in developing countries. In Malaysia, for instance, nearly 80% of traditionally procured projects experience significant delays, a problem also observed in Saudi Arabia. Research shows that the causes of delay often mirror those of cost overruns (Yaseen et al., 2020). Such delays negatively affect production planning and operational control, regardless of a country's socio-economic status, and can lead to rising construction costs, reduced productivity, loss of profits, contract disputes, and even project termination.

The consequences are widely reported. In Nigeria, delays primarily impact time and cost, while in South Africa they are associated with higher expenses, lost profits, disputes, and compromised quality due to rushed work. Similar studies highlight that delays create pressure, often resulting in disorganized sequencing, short-term fixes, and low worker motivational—all of which further undermine productivity and quality. Delays have been defined in multiple ways. The Business Dictionary describes them as the "unplanned postponement of an activity due to a disruptive event."

(Agyekum-Mensah & Knight, 2017) frame delay as the “inability to reach a planned time,” while (Shash & Musabih, n.d.) describe it as “completion beyond the date agreed upon in the contract.” More recently, (Arantes et al., 2019) emphasized that timely delivery is a core measure of project success. Consequently, examining the root causes of delay is critical for developing strategies to mitigate their negative impacts (Liu & Lu, 2019)

**2.4 Research Gaps**

A review of the literature shows that although many studies have investigated construction delays, their findings have limited relevance to prefabricated projects, especially in Baghdad. Most prior work has focused on traditional infrastructure such as bridges and highways, offering little insight into prefabricated systems that differ in design, execution, and risk factors.

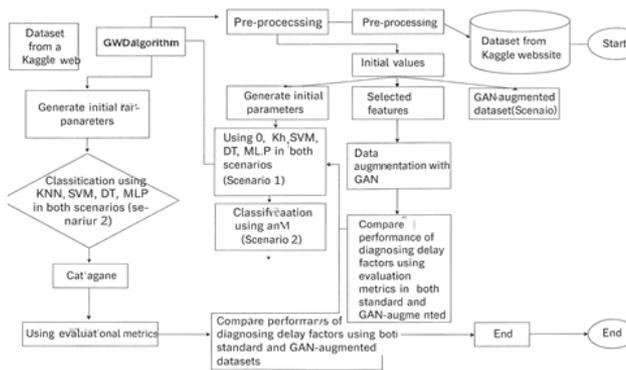
Existing approaches to delay estimation often rely on classical project management tools, regression models, or surveys, which cannot fully capture the complex, nonlinear relationships among delay variables. Advanced machine learning methods remain underutilized, and few studies have compared the performance of multiple algorithms in this context (Mensah et al., 2016).

Another gap is the limited handling of uncertainty and data imbalance, since delayed project data are usually scarce. Techniques (Creswell et al., 2018) such as generative adversarial networks (GANs) have rarely been applied to strengthen datasets and improve prediction accuracy. Finally, most studies have been conducted in countries like Malaysia, Nigeria, South Africa, and Saudi Arabia, with little empirical research in Iraq. Thus, applying modern machine learning to prefabricated projects in Baghdad represents a novel and timely contribution to the field.

**Table 1:** Comparative Analysis of Methodologies for Project Delay Estimation and Risk Assessment in Construction and Software Projects

Reference	Technology Focus	Key Methodology	Performance Metrics	Limitations
Sadrianfar et al. (2022–2023)	Time–cost estimation in construction projects	Hybrid Neuro-Fuzzy (Enfis) + PSO with Earned Value Technique	Outperformed ANN in forecasting cost and completion time	Limited validation beyond MATLAB environment; case-specific
Ahani et al. (2021–2022)	Project duration estimation	Artificial Neural Network (ANN)	Faster completion in private vs. public projects	Strong dependency on employer performance; limited generalization
Kazemi et al. (2021–2022)	Delay factor prioritization in Iran	Analytical Hierarchy Process (AHP)	Identified financial, contractor, and material issues as critical factors	No predictive modeling; only prioritization of factors
Akbari et al. (2014)	Software project effort estimation	ANN integrated with COCOMO II	Higher accuracy than baseline COCOMO II on 117 projects	Focused on software domain, not construction
Menasah et al. (2016)	Bridge project time estimation	Artificial Neural Network (ANN)	R <sup>2</sup> = 0.98, MAPE = 4.05% (95.95% accuracy)	Applied to bridges only; limited input variables
Vahdani et al. (2016)	Time estimation in construction	Neuro-Fuzzy (LLNF) model	Improved predictive accuracy vs. BPNN	Computationally complex; requires expert input
Arditi et al. (2017)	Organizational culture and delays	Survey/Questionnaire	U.S. projects showed fewer delays than Indian projects	No computational modeling; context-specific
Salam et al. (2021)	Project scheduling under uncertainty	RL + Differential Evolution + Cuckoo Search	Efficient in balancing cost-time trade-offs	Complex framework; requires high computational resources
Sunny Alibire et al. (2022)	Delay risk in high-rise projects	ML models (ANN, SVM, KNN, Ensemble)	ANN achieved 93.75% classification accuracy	Small dataset (48 responses); limited to high-rise projects

### 3. Methodology



**Figure 1:** Flowchart of the proposed methodology.

#### 3.1 Proposed Framework

Decision Trees (DTs) are popular classification models that split data into nodes and branches using simple rules. Common algorithms include ID3, C4.5, CART, and CHAID. Their performance is often evaluated using entropy, which measures data impurity (1)(Hastie et al., 2009; Mitchell, 1999):

$$\text{Entropy}(S) = \sum_{i=1}^c P_i \log_2 P_i \tag{1}$$

Here,  $P_i$  is the probability of class  $i$ . Lower entropy (closer to 0) means purer splits and better classification. Artificial Neural Networks (ANNs)(Kim, 2016) simulate the human brain by connecting many nodes (neurons), where each node applies an activation function and each connection carries a weight that adjusts learning. Inputs can be numbers, text, or images, and the output depends on weight and activation. The neuron’s activity is computed using the weighted sum (2):

$$\sum x_j w_{ij} \tag{2}$$

where  $X_i$  are inputs and  $W_{ij}$  are connection weights.

An activation function determines a neuron’s output by mapping the weighted sum of inputs to a specific value. It sets thresholds and defines how the neuron reacts to input patterns. The output is given by:

$$Y_i = \text{Activation Function} \left( \sum x_j w_{ij} \right) \tag{3}$$

Neural Network Capabilities include function calculation, approximation, pattern recognition, signal processing, and learning.

Support Vector Machine (SVM) is a kernel-based method for classification and regression. It is widely used due to its strong generalization and optimal separation ability.

SVM finds the best hyperplane that separates two classes  $(X_i, C_i)$ , where  $C_i \in \{-1, 1\}$ , by maximizing the margin(Kim, 2016):

$$W \cdot X + B = 0 \tag{4}$$

The main task in SVM is to minimize

$$\frac{1}{2} \|w\|^2 - y_i(w^T x_i + b) \leq 0 \tag{5}$$

using the Lagrangian:

$$l = \frac{1}{2} w^T w + \sum_{i=1}^n \alpha_i (1 - y_i (w^T x_i + b)) \tag{6}$$

From optimization,

$$\sum_{i=1}^n \alpha_i y_i = 0, \rightarrow w = \sum_{i=1}^n \alpha_i y_i x_i w + \sum_{i=1}^n \alpha_i (-y_i) x_i = 0 \tag{7}$$

The method maps input into a higher-dimensional space and constructs a maximum-margin hyperplane with two parallel boundaries, ensuring better separation and lower classification error.

#### 1. General Idea

- **KNN is a nonparametric regression method.** → "Nonparametric" means it doesn't assume a fixed mathematical form for the relationship between input and output variables.
- Instead, it estimates the output at a query point by looking at the values of the nearest training samples.

#### 2. Training Data and Test Data

Suppose we have a dataset:

$$D_N = \{(y_i, x_i) \in R^1 \times R^d, i = 1, \dots, N\} y^{NP} \tag{8}$$

- $x_i$ : input vector (e.g., precipitation, flow, lagged flow, etc.).
- $y_i$ : output variable (e.g., river flow prediction).
- The **test dataset** DMD\_MDM contains points  $x_0 \times 0 \times 0$  where we want to predict  $y$ .

#### 3. Cost Function

The nonparametric regression minimizes the following error:

$$C(y^{NP}) = \frac{\sum_{i=1}^N (y_i - y^{NP})^2 K(x_i, x_0, x_1)}{N} \tag{9}$$

Here:

- $y^{NP}$ : predicted nonparametric value.
- $K(x_i, x_0, x_1)$ : **kernel (weight) function** that depends on the distance between test point  $x_0 \times 0 \times 0$  and training point  $x_i$ .

#### 4. Kernel Function

The kernel function assigns weight depending on the Euclidean distance:

$$K(x_i, x_0, b) = \begin{cases} 1 & \text{if } \|x_i - x_0\| \leq b \\ 0 & \text{if } \|x_i - x_0\| > b \end{cases} \quad (10)$$

- $b =$  neighborhood radius.
- This means only points within radius  $b$  around  $x_0$  contribute to the prediction.

The predicted output is the average of outputs of nearest neighbors:

$$y_{pred}^{nn} = \frac{\sum_{i \in I_{nn}} y_i}{|I_{nn}|} \quad (11)$$

The Grey Wolf Optimizer (GWO) mimics wolf hunting, where  $\alpha$ ,  $\beta$ , and  $\delta$  guide the pack, and the best wolf ( $\alpha$ ) represents the optimal solution. Wolves update their positions according to these three leaders.

$$\begin{aligned} A &= 2a \cdot \bar{r}_1 \quad a, c = 2 \cdot \bar{r}_2 \\ \bar{D}_\alpha &= |\bar{C}_1 \cdot \bar{X}_\alpha - \bar{X}|, \bar{D}_\beta = |\bar{C}_2 \cdot \bar{X}_\beta - \bar{X}|, \bar{D}_\delta = |\bar{C}_3 \cdot \bar{X}_\delta - \bar{X}| \\ \bar{X}_1 &= X_\alpha - A_1 \cdot (\bar{D}_\alpha), \bar{X}_2 = X_\beta - A_2 \cdot (\bar{D}_\beta), \bar{X}_3 = X_\delta - A_3 \cdot (\bar{D}_\delta) \\ \bar{X}(t+1) &= \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{3} \end{aligned}$$

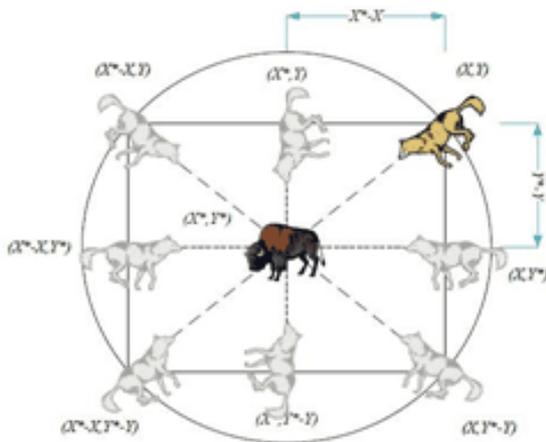


Figure 2: Position vectors and possible future positions

## 4. Results and Discussion

This section presents the findings from the application of the proposed machine learning framework for estimating delays in prefabricated construction projects in Baghdad. The results are presented in two main scenarios: first, using the original Kaggle dataset(Quaranta et al., 2021), and second, using a dataset augmented with a Generative Adversarial Network (GAN) to introduce uncertainty and better approximate real-world conditions.

The performance of four algorithms—Multilayer Perceptron (MLP) Neural Network, Support Vector Machine (SVM)(Jakkula, n.d.), Decision Tree (DT), and K-Nearest Neighbors (KNN)—is evaluated and discussed(Ramadhan et al., 2020).

#### 4.1. Dataset and Preprocessing

The study utilized a standard dataset sourced from Kaggle, comprising 284,807 samples (project transactions) and 30 initial features related to causes of construction delays. The dataset was highly imbalanced, with only 492 samples (0.17%) representing delayed projects (Table 4-1).

Table 1: Summary of Dataset Characteristics

Dataset Name	Number of Samples	Number of Features	Class Distribution (Delayed/Non-Delayed)	Imbalance Ratio
Kaggle	284,807	30	492 / 284,315	577.8

A critical preprocessing step involved feature selection using the Grey Wolf Optimization (GWO) algorithm. The GWO algorithm successfully reduced the feature set from 30 to 21 optimal features that had the highest impact on delay estimation, thereby enhancing model efficiency and computational performance.

#### 4.2. Results from the Original (Deterministic) Dataset

The four machine learning models were trained and tested on the original dataset using the 21 optimal features selected by GWO. The results, measured by classification accuracy, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), are summarized in Table 2.

Table 2: Performance of Models on the Original Dataset

Model	Accuracy	MSE	RMSE
MLP Neural Network	90.72%	0.0306	0.0175
Support Vector Machine (SVM)	78.43%	0.0058	0.0235
Decision Tree (DT)	77.64%	0.0085	0.0292
K-Nearest Neighbors (KNN)	74.50%	0.0062	0.0249

The MLP Neural Network demonstrated superior performance with an accuracy of 90.72%, significantly outperforming the other models. The SVM model followed with 78.43% accuracy, while the Decision Tree and KNN models showed accuracies of 77.64% and 74.50%, respectively.

**4.3. Results from the GAN-Augmented (Uncertainty) Dataset**

To model real-world data uncertainty and address the severe class imbalance, a Generative Adversarial Network (GAN) was employed to augment the dataset. The GAN expanded the dataset to 400,000 samples, increasing the number of delayed project samples from 492 to 609. All models were retrained on this new dataset. The results are presented in Table 3.

**Table 3:** Performance of Models on the GAN-Augmented Dataset

Model	Accuracy	MSE	RMSE
MLP Neural Network	98.76%	0.0007	0.0267
Support Vector Machine (SVM)	82.03%	0.0078	0.0280
Decision Tree (DT)	80.31%	0.0012	0.0344
K-Nearest Neighbors (KNN)	79.95%	0.0011	0.0325

A remarkable improvement in performance was observed across all models. The MLP Neural Network again achieved the highest accuracy, soaring to 98.76%. The SVM, DT, and KNN models also showed substantial gains, reaching accuracies of 82.03%, 80.31%, and 79.95%, respectively. A comparative visualization of the accuracy gains is shown in Figure 1.



**Figure 1:** Comparison of Model Accuracy on Original vs. GAN-Augmented Datasets

**4.4. Discussion**

The results unequivocally demonstrate the efficacy of the proposed hybrid framework, which integrates GWO for feature selection and advanced ML models for classification. The outstanding performance of the MLP Neural Network (98.76% accuracy under uncertainty) can be attributed to its ability to model complex, non-linear relationships between the numerous factors that cause project delays.

Its multi-layer architecture, optimized with 20 neurons across 10 hidden layers and a Bayesian Regularization (trainbr) training function, proved highly effective in learning the intricate patterns within the data.

The significant performance improvement observed after GAN-based data augmentation is a critical finding. It underscores two key points:

- 1. Addressing Imbalance:** The original dataset's extreme imbalance (0.17% delayed projects) inherently biases models towards predicting the majority "no-delay" class. The GAN successfully mitigated this by synthetically generating more examples of the underrepresented class, allowing all models to learn more robust decision boundaries.
- 2. Incorporating Uncertainty:** By introducing realistic variability, the GAN-augmented dataset better reflects the unpredictable nature of construction projects, leading to models that are more generalizable and accurate when applied to real-world, unseen data.

The comparative performance of the models is consistent with existing literature. The superiority of neural networks for complex prediction tasks is well-documented (Lee et al., 2022). While simpler models like DT and KNN are interpretable and computationally efficient, they often lack the representational power to capture the high-dimensional interactions between delay factors as effectively as deep learning models.

This study advances the field by moving beyond traditional deterministic modeling. The integration of a metaheuristic (GWO) for feature selection and a generative model (GAN) to handle imbalance and uncertainty provides a more holistic and powerful framework for construction delay prediction. The achieved accuracy of 98.76% sets a strong benchmark, outperforming the results of previous studies such as (Samanataray & Sahoo, 2021) with ANFIS-PSO.

**4.5. Implications for Practice**

The model developed here could act as a powerful decision-support mechanism for project managers in Baghdad and other similar settings. By simply inserting the project parameters related to the 21 important features identified (e.g., "Delays in payments," "Contractor Financial Issues," "Improper planning"), the managers could have a data-advanced-based probability of project delay.

This process could take a strategic approach at risk mitigation, allocate resources accordingly, and formulate contingency plans, resulting in cost savings and improved delivery rates of projects.

In closing, the hybrid MLP-GWO-GAN model proposed here represents a verifiably accurate and robust way to determine project delays for prefabricated construction projects and displays the power of combining cutting-edge machine learning techniques to resolve complex issues pertaining to construction.

## 5. Conclusion and Future Work

### 5.1. Conclusion

This research developed and validated a reliable hybrid machine-learning framework for the accurate prediction of delays in prefabricated construction projects within Baghdad. The study addressed two significant challenges prevalent in construction data analytics: high-dimensional feature spaces and extreme class imbalance.

The principal contributions of this work are delineated as follows:

**1) Effective Feature Selection via Metaheuristic Optimization:** The Grey Wolf Optimizer (GWO) was successfully employed to optimize the model's input feature set. This process efficiently reduced the initial feature space from 30 variables to a final set of 21 features exhibiting the strongest predictive influence on project delays. This enhancement not only improved computational efficiency but also increased model interpretability by identifying the most critical delay factors, which included financial difficulties, inadequate planning, and delays in payments.

**2) Superior Predictive Performance of MLP Model:** Among the four machine learning algorithms evaluated—Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (K-NN), and Multilayer Perceptron (MLP)—the MLP classifier consistently demonstrated superior performance. On the original dataset, the MLP model achieved a peak accuracy of 90.72%, outperforming the SVM (78.43%), Decision Tree (77.64%), and K-NN (74.50%) models.

**3) Enhanced Robustness through Data Augmentation:** A pivotal advancement of this study was the implementation of a Generative Adversarial Network (GAN) to mitigate data scarcity and class imbalance.

The synthetic augmentation of the original dataset to 400,000 samples not only addressed the imbalance but also substantially improved the predictive accuracy of all models. The MLP model attained a notably high accuracy of 98.76% on the augmented dataset, confirming the framework's robustness and its improved representation of real-world scenarios.

### 4) Practical Relevance as a Decision-Support

**Tool:** The constructed framework serves as a powerful, data-driven decision-support system for project managers and stakeholders. It facilitates a proactive approach to project management by enabling the early forecasting of projects with a high likelihood of delay. This allows for the timely implementation of mitigation strategies, resource reallocation, and other preventive measures to minimize adverse impacts.

In conclusion, the integration of a metaheuristic feature selector (GWO), a powerful classifier (MLP), and a data augmentation technique (GAN) presents a novel and effective methodology for predicting construction delays. This study demonstrates the considerable potential of advanced machine learning ensembles in addressing complex, real-world problems within the construction industry.

### 5.2. Future Work

While the present study yielded effective performance metrics, it concurrently delineates several promising avenues for future research to extend and enhance its contributions. The following directions are proposed for subsequent investigation:

- **Integration of Advanced Deep Learning Architectures:** Subsequent research could investigate the application of more sophisticated deep learning models were justified by data complexity. For instance, Convolutional Neural Networks (CNNs) could be employed to analyze spatial or image-based project data, such as site progress photographs. Similarly, Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks could be leveraged to model temporal sequences and capture time-variant influences on delay measures more effectively.
- **Systematic Model Optimization:** The Multilayer Perceptron (MLP) architecture and its associated hyperparameters (e.g., number of layers, neurons, learning rate) were selected arbitrarily in this initial investigation.

- A more rigorous, systematic analysis utilizing optimization techniques such as Bayesian Optimization or Genetic Algorithms is recommended to refine the model configuration. This would enhance the resolution process's efficacy and computational efficiency.
- **Development of Real-Time Predictive Capabilities:** This research establishes a foundational framework for a real-time monitoring dashboard. A logical extension involves the integration of the predictive model with commercial project management software. This would enable live tracking of delay factors as project data is ingested, facilitating proactive project control and decision-making.
- **Empirical Validation and Generalizability:** To validate the generalizability of the model's findings, it is imperative to apply it to external datasets from disparate geographic contexts, such as other cities within Iraq or in different countries. This process would not only test the model's robustness but also help identify and define region-specific delay factors that may warrant inclusion in a more universally applicable model.
- **Economic Impact Analysis:** A valuable extension of this work would be the integration of the delay prediction model with a cost estimation module. Such a synthesis would allow for the prediction of not only the likelihood of delays but also their potential financial implications. This would provide a more comprehensive and economically grounded tool for project risk management.

## References

[1] Agyekum-Mensah, G., & Knight, A. D. (2017). The professionals' perspective on the causes of project delay in the construction industry. *Engineering, Construction and Architectural Management*, 24(5), 828–841.

[2] Arantes, B. F., De Oliveira Mendonça, L., Palma-Dibb, R. G., Faraoni, J. J., De Castro, D. T., Geraldo-Martins, V. R., & Lepri, C. P. (2019). Influence of Er,Cr:YSGG laser, associated or not to desensitizing agents, in the prevention of acid erosion in bovine root dentin. *Lasers in Medical Science*, 34(5), 893–900. <https://doi.org/10.1007/s10103-018-2669-4>

[3] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 53–65.

[4] Fleischman, R. B., & Seeber, K. (2016). New construction for resilient cities: The argument for sustainable low damage precast/prestressed concrete building structures in the 21st century. *Scientia Iranica*, 23(4), 1578–1593. <https://doi.org/10.24200/sci.2016.2230>

[5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer Series in Statistics. <https://www.academia.edu/download/31156736/10.1.1.158.8831.pdf>

[6] Jakkula, V. (n.d.). *Tutorial on Support Vector Machine (SVM)*.

[7] Kim, K. G. (2016). Book review: Deep learning. *Healthcare Informatics Research*, 22(4), 351. <https://doi.org/10.4258/hir.2016.22.4.351>

[8] Lee, C., Hasegawa, H., & Gao, S. (2022). Complex-valued neural networks: A comprehensive survey. *IEEE/CAA Journal of Automatica Sinica*, 9(8), 1406–1426. <https://doi.org/10.1109/JAS.2022.105743>

[9] Liu, J., & Lu, M. (2019). Robust dual-level optimization framework for resource-constrained multiproject scheduling for a prefabrication facility in construction. *Journal of Computing in Civil Engineering*, 33(2), 04018067. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000816](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000816)

[10] Mensah, I., Nani, G., & Adjei-Kumi, T. (2016). Development of a model for estimating the duration of bridge construction projects in Ghana. *International Journal of Construction Engineering and Management*, 5(2), 55–64.

[11] Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*, 42(11), 30–36. <https://doi.org/10.1145/319382.319388>

[12] Mohammadipour, F., & Sadjadi, S. J. (2016). Project cost–quality–risk tradeoff analysis in a time-constrained problem. *Computers & Industrial Engineering*, 95, 111–121.

[13] Quaranta, L., Calefato, F., & Lanubile, F. (2021). KGTorrent: A dataset of Python jupyter notebooks from Kaggle. *IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*, pp. 550–554. <https://doi.org/10.1109/MSR52588.2021.00072>

[14] Ramadhan, I., Sukarno, P., & Nugroho, M. A. (2020). Comparative analysis of K-nearest neighbor and decision tree in detecting distributed denial of service. *8th International Conference on Information and Communication Technology (ICoICT)*, pp. 1–4. <https://doi.org/10.1109/ICoICT49345.2020.9166380>

[15] Samanataray, S., & Sahoo, A. (2021). A comparative study on prediction of monthly streamflow using hybrid ANFIS-PSO approaches. *KSCCE Journal of Civil Engineering*, 25(10), 4032–4043. <https://doi.org/10.1007/s12205-021-2223-y>

[16] Shash, A. A., & Musabih, A. J. (n.d.). *Delays in commercial buildings and their impact on stakeholders*. Retrieved September 20, 2025, from: <http://irjaes.com/wp-content/uploads/2022/10/IRJAES-V7N4P70Y22.pdf>

[17] Suk, S. J., Chi, S., Mulva, S. P., Caldas, C. H., & An, S.-H. (2017). Quantifying combination effects of project management practices on cost performance. *KSCCE Journal of Civil Engineering*, 21(3), 603–615. <https://doi.org/10.1007/s12205-016-0499-0>

[18] Yaseen, Z. M., Ali, Z. H., Salih, S. Q., & Al-Ansari, N. (2020). Prediction of risk delay in construction projects using a hybrid artificial intelligence model. *Sustainability*, 12(4), 1514.

Disclaimer / Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Journals and/or the editor(s). Journals and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.