A Machine Learning Approach to Predict Time Delays in Marine Construction Projects

Aymen H. Nassar
Department of Civil Engineering, German University in Cairo, Cairo, Egypt
αyman.hamdy@guc.edu.eg

Ahmed M. Elbisy
Department of Civil Engineering, German University in Cairo, Cairo, Egypt
ahmedelbisy98.29@gmail.com (corresponding author)

Received: 18 June 2024 | Revised: 3 July 2024 | Accepted: 6 July 2024

 DOI: https://doi.org/10.48084/etasr.8173

ABSTRACT

The estimation of time delays in construction projects represents a challenging undertaking, frequently constrained by insufficient data, inherent uncertainties, and potential risks. Nevertheless, it remains a crucial element in ensuring the success of a construction project. Marine construction projects represent a highly specialized subcategory of the construction sector, characterized by a considerable degree of risk and significant financial outlays. Despite the extensive application of Machine Learning (ML) techniques across a range of domains, there is a notable absence of studies evaluating their efficacy, particularly in the context of marine construction project assessment. In light of the above, the present study examines the potential of ML techniques for estimating time delays in marine construction projects. A total of 43 factors that affect marine construction projects in terms of time delay were identified and categorized into nine major groups through a detailed analysis of interviews with experts from the marine construction industry. The relative importance index method was employed to ascertain the relative importance of the factors affecting delays. The factors and groups were then ordered according to their level of impact on time delay. Considering the advancements in ML, this study utilizes General Regression Neural Networks (GRNN), Support Vector Machines (SVM), and Tree Boost functionality to estimate the time delay of marine construction projects. To evaluate the predictive capacity of each model, they were assessed using five statistical features and Taylor diagram visualization. With regard to predicting time delay, the overall performance of the GRNN was found to be more accurate than that of the other models, while the SVM model exhibited the least predictive capabilities. The GRNN model was found to be both efficient and precise and, therefore, may serve as a practical tool for predicting the time delay of marine construction projects.

Keywords-time delay; marine construction projects; relative importance index; Machine Learning (ML)

I. INTRODUCTION

A mere 2.5% of global construction projects have been classified as successful in terms of scope, cost, schedule, and business objectives, as evidenced by empirical data [1]. Marine construction projects represent a particularly notable instance of human-induced environmental alteration. These projects represent a highly specialized division of the construction sector, characterized by a considerable degree of risk and significant financial outlays [2]. Authors in [3] investigated the significant impact on economic factors, which have received less attention than onshore projects. This has resulted in considerable cost overruns. It is evident that marine construction has a significant impact on the global Exclusive Economic Zones, with an average of 1.5% (0.7% to 2.4%) affected, in comparison to the relatively minor impact on urban areas, with an average of 0.02% to 1.7% affected. It is therefore of the utmost importance that marine construction projects are completed successfully. Failure is defined as the inability of a marine construction project to meet its critical objectives, which may include initial targets relating to cost, on-time completion, safety and quality, stakeholder engagement, environmental impact (sustainability), strategic outcomes (e.g. macro-economic targets), lifecycle functionality/performance and aesthetics [4]. Delivering a project on time and on budget is a key objective of project management and is considered essential to project success. However, due to the complexity of construction projects, cost overruns are common. According to the World Bank, 30% to 40% of construction projects worldwide experience cost overruns [5-9].

There are a number of studies that have examined the issue of time delays in marine construction projects. Authors in [10] reported a prominent risk factor affecting marine construction projects, namely the discrepancy between tender documents and actual project conditions. The study identified a number of design and construction related factors, such as design errors, estimation errors, inexperience of the design team, construction errors and the consultant's inability to deal with severe uncertainties. They also found that unreliable geotechnical assumptions are a common cause of time delays and in marine projects [11]. According to [12], the causes of delay vary, depending on the type of project. For example, most authors agree that bad weather causes delays. However, the effect of weather is more significant for maritime projects than for land-based projects. Therefore, it is not sufficient to simply identify the causes of delays, but their impact should be assessed according to the project environment. Authors in [13] studied the causes of delays in the Gulf region for oil and gas projects and identified seven factors as the main causes of project delays, such as poor site management and supervision by contractors, delay in delivery of materials, lack of effective communication between project stakeholders, and poor interaction with suppliers in the engineering and procurement stages. Of these major causes, only poor management of contractors' schedules' shows significant differences in the perceptions of project stakeholders. Authors in [14] investigated the main causes of time delays and cost overruns in Saudi Arabian oil and gas construction projects. A total of 38 causes of time delay and cost overrun were identified through a review of the literature and an interview.

In a related study [15], the factors affecting the delay of oil and gas construction projects in Indonesia were examined. A considerable number of upstream oil and gas projects have been unable to meet the original contract dates. The delays in oil and gas construction projects were classified according to the major factors identified in the construction and engineering plan contracts. In [16], the research focused to the factors that contribute to delays in large-scale natural gas projects. The findings indicated that 72% of the subjects perceived the delay of gas projects in Australia to range between 10% and 30%. In a survey of oil and gas construction projects in Vietnam [17], five major causes of schedule overruns were identified. These were excessive bureaucracy in terms of project approvals from clients, lack of efficient design, lack of competency in different parties, inadequacies in contract constructs and delays in internal approval processes. The latter include delays in the delivery of land, the handover of sites, the granting of work approvals, the approval of drawings and documents, and the delivery of test results. The impact of these factors on marine construction projects can be significant. Authors in [18] identified and analyzed the risks associated with the execution of offshore petroleum and gas projects. The findings revealed that there are certain activities with a markedly high probability of occurrence and impact. These include adverse weather conditions affecting the project, an increase in material costs, delays in the acceptance or approval of contractor submittals by the owner, and the presence of rigorous quality control standards. The survey in [19] was concentrated on the factors that lead to delays in offshore wind turbine projects. The study was based on an analysis of 208,140 historical data points from seven different cases, investigating both onshore and offshore on-site assembly locations. The findings indicated that the primary contributing factors to delays were "planning" at the onshore assembly location and "previous task" at the offshore assembly location. This challenges the current assumption that weather is the primary cause of delays in offshore projects.

The application of ML has the potential to significantly impact project management, with the automation of repetitive tasks, the generation of predictions, and the provision of insights for enhanced decision-making representing key areas of benefit. The successful implementation of ML in project management is contingent upon the availability of pertinent and high-quality data, as well as the careful consideration of ethical and privacy concerns [20]. Furthermore, it is essential to continuously refine and validate ML models with new data to guarantee their accuracy and efficacy in supporting project management activities. The inaccuracy and unreliability of time and cost predictions in construction projects [21], along with their tendency to exceed initial estimates, are common occurrences that present a significant challenge [22]. The uniqueness, diversity and complexity of a project, coupled with the ever-present risks, make it challenging to establish a model that can accurately assess project costs. This is why many researchers often direct their attention to this problem, employing a range of approaches and methods that are adapted to specific types of buildings and structures [23]. Authors in [24] developed Artificial Neural Networks (ANNs) models for forecasting the final budget and duration of a highway construction project during the construction stage. The results may potentially serve as an early warning of an impending budgetary overspending or schedule delay. A total of 92 building projects were used as a basis for the application of ANNs and SVM in order to predict the success of cost and schedule projections at the conceptual stage. The model demonstrated a 92% and 80% accuracy in predicting cost and schedule success, respectively [25]. Additionally, in [26] two ML approaches (with an accuracy value of 79.41% and 73.52% for decision tree and Bayes models, respectively) were employed in order to predict delays in construction logistics in Qatar.

Moreover, in [27], a hybrid artificial intelligence model (a combination of RF and GA) was developed and achieved an accuracy value of 91.67% for the prediction of delays in Iraq. Authors in [28] developed a multilayer, high-performance ensemble predictive model utilizing hyperparameter-optimized ensemble machine learning algorithms for the purpose of predicting delays in construction projects. The findings indicated that the ensemble algorithms demonstrated superior accuracy in predicting construction project delays when compared to a single algorithm. In order to predict the construction duration an analytical method was employed [29], including ANNs, smoothing techniques, time series analysis, and Bromilow's time–cost model. The findings suggest that smoothing techniques with a constant value of 0.3 are an effective approach, with a notable small error of 1.2%. Furthermore, this technique demonstrates superior outcomes compared to other techniques. Authors in [30] presented evidence of well-established evaluation frameworks for predicting time delay and cost overrun. In order to create hyper-parameter-optimized predictive models, the numerical data was primarily used to train a specific ANNs model. The results demonstrated that the Relative Importance Index (*RII*) value was 0.939 and the Mean Absolute Error (*MAE*) was

0.057 for the data on cost overruns, while it was 0.939 and the *MAE* was 21.709 for the data on time delays.

Marine construction projects have historically been characterized by a range of cost overruns and time delays, which can be attributed to their complex and evolving nature. Cost overruns and time delays are recurring issues, encountered in marine construction projects on a global scale, affecting multitude of endeavors, including petroleum and gas projects, harbors, onshore and offshore structures. The current state of the global marine construction sector gives cause for concern, as it demonstrates a considerable degree of inefficiency. The implementation of this approach will enable a solution to reduce cost overruns and time delays, thereby improving the overall efficiency of the industry's operations. The present study aims to develop an ML model for predicting time delays in marine construction projects, and to provide project management engineers with a robust and accurate ML-based model capable of representing time delays in marine construction projects.

II. METHODOLOGY

Figure 1 illustrates the methodology adopted to achieve the study's objectives.

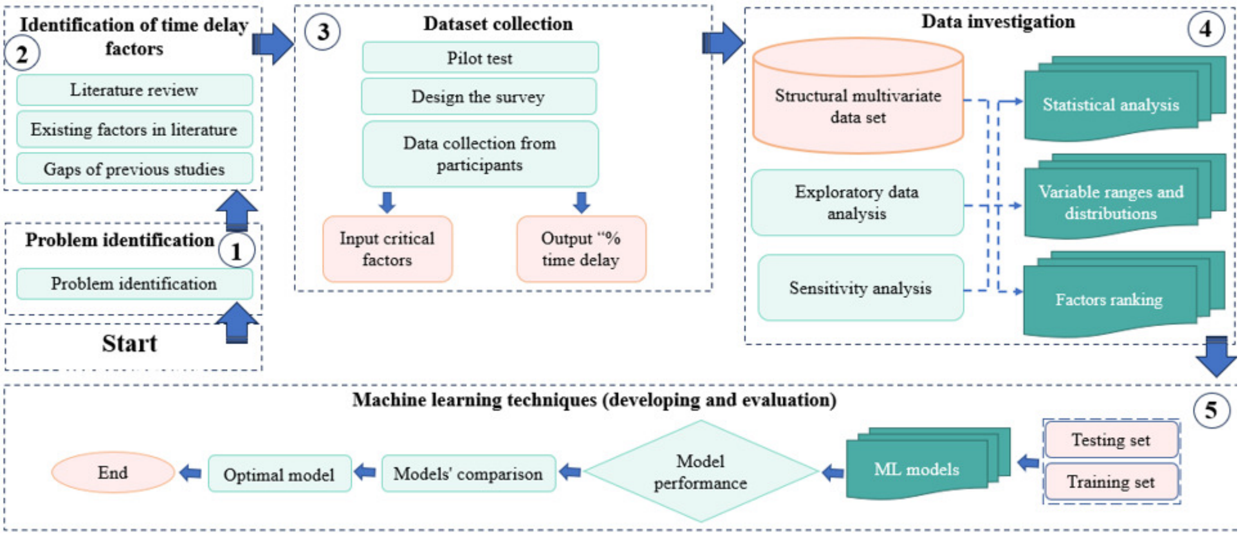


Fig. 1. Methodology process.

III. STUDY CONSTRAINTS

The following limitations of the study should be noted:

- There is a scarcity of marine projects globally.
- This study is limited to identifying, ranking, and analyzing the causes of time delays and cost overruns in marine construction projects in Egypt, Saudi Arabia, the United Arab Emirates (UAE), Kuwait, and Qatar.

IV. IDENTIFICATION OF TIME DELAY FACTORS

This study identifies and analyzes the key factors that influence time delays in marine construction projects, based on a review of prior research. The current study presents a survey based on existing literature, which was carried out in three stages. A comprehensive search was initially conducted to identify articles that included the following key terms: "overrun", "delay", "time delay", "time overrun", "delay risk", "delay factor", "delay source", "delay cause", and "construction" or "construction project". The temporal scope of the search was defined from 1985 to 2023. The second stage of the search, involved a more detailed examination of the identified articles from the first stage. This entailed a review of the titles, abstracts, and keywords of the articles, with the objective of selecting and retaining only those that were deemed to be relevant to the research scope. In the third stage, an in-depth review was carried out on the selected articles from the second stage. This included a detailed examination of previous studies' classifications of time delay factors and their integration into primary sources of delay and overrun risk. The three-stage literature survey identified nine categories of delay, which were subsequently adopted by the present study. These categories are: (a) the owner, (b) the consultant, (c) the contractor, (d) the designer, (e) all parties, (f) the resources (material and equipment), (g) the contract, (h) external, (j) force majeure.

It is notable that the individual causes of time delays identified in the reviewed articles exhibited variations within the different risk source categories. The divergence in findings across the articles can be explained by the differences in construction environments, geographical conditions, political situations, construction methods, resource availabilities, and stakeholder engagements. Consequently, a preliminary meeting was held with 10 construction experts to ascertain the applicability of the identified time delay causes to marine construction projects and to make any necessary modifications. Based on the literature survey and the expert meetings, 43 causes of time delay were identified.

V. DATA COLLECTION

In order to facilitate the analysis of the obtained data, a questionnaire comprising closed-type questions was employed. A total of 10 questionnaires were distributed to experts specializing in marine construction with the objective of ensuring that the questionnaire comprehensively addressed the majority of factors that may affect marine construction projects in terms of time delay. Moreover, the input of the experts was taken into account and integrated, along with other related factors and questions. The ranking scale was a 5-point Likert scale, classified as indicated in Table I. The questionnaire was constructed using the online survey tool Google Forms and subsequently distributed to the participants.

TABLE I. LIKERT SCALE BREAKDOWN

Scale	Probability	Impact
1 (Lowest)	Very Rare	Very Low
2	Rare	Low
3	Possible	Moderate
4	Frequent	High
5 (Highest)	Very Frequent	Very High

The determination of an adequate sample size is an essential component of the data collection process. To ensure the findings are applicable to a wider population, it is important to use a sufficiently large sample size. The determination of an appropriate sample size is contingent upon a number of factors, including the size of the population under investigation, the minimum acceptable level of precision, and the desired level of confidence. The sample size in this study was determined using the commonly employed approach of "utilizing as many participants as possible within budgetary constraints" [31]. A survey was distributed to a randomly selected subset of 103 individuals from various segments within the marine construction sector. The respondents' years of experience and level of education are presented in Tables II and III, respectively. The data set comprises 30 projects from the public and private sectors in Egypt (18 projects), Qatar (1 project), Kuwait (2 projects), Saudi Arabia (6 projects), the UAE (2 projects), and Oman (1 project).

TABLE II. RESPONDENTS YEARS OF EXPERIENCE

Years of experience	Percentage of respondents	Number of Respondents
< 5 years	2.91%	3
5 to 9 years	19.42%	20
10 to 14 years	25.24%	26
15 to 19 years	28.16%	29
≥ 20 years	24.27%	25
Total	100 %	103

TABLE III. RESPONDENT'S HIGHEST DEGREE OF EDUCATION

Degree of education	Percentage of respondents	Number of Respondents
Bachelor's Degree	25%	26
Master's Degree	29%	30
Ph.D. or higher	46%	47
Total	100%	103

VI. ANALYSIS TOOLS

The reliability of the questionnaire was evaluated through the application of Cronbach's alpha (α). The reliability of the questionnaire was evaluated, and the computed α values were 0.963 and 0.973 for the severity scale of time delay and the probability scale, respectively. These values were deemed to be acceptable since they exceeded the minimum acceptable level of 0.7 [32]. The Cronbach's alpha values have demonstrated high reliability and internal consistency in each scale [33]. The data were employed to establish a ranking of the factors, from the most to the least important, in accordance with the *RII* method. The factors were then ranked in descending order of importance according to the *RII*:

$$RII = PI \times SI \quad (1)$$

where *PI* is the probability index and *SI* is the severity index.

VII. DATA ANALYSIS

The results of the survey are presented in Table IV In terms of overall importance, the factor "difficulties in project financing by contractor" was identified as the most critical time delay in marine construction projects, with an *RII* score of 14.9. Additionally, "inflation" is the second highest, followed by "poor planning and management of the contractor's schedule" from the consultant category, "fluctuation in cost" from the external category, and "incompetence or inexperience of the contractor" from the contractor category. The third highest factor is the lack of effective communication among project stakeholders, followed by ineffective management, planning, and scheduling of the project. These factors are the fifth, fourth, and third highest in the entire study, respectively. The contractor-related factors group had the highest *RII*, followed by the external-related factors group with an insignificant difference. The groups are ranked in Figure 2.

VIII. MACHINE LEARNING TECHNIQUES

The objective of ML techniques is to identify precise approximations that offer a robust, computationally efficient, and cost-effective solution while conserving computational time.

A. Generalized Regression Neural Network (GRNN)

GRNNs are feedforward networks that employ a nonlinear regression-based learning mechanism, as outlined in [34]. During GRNN training, every distinct pattern is memorized and acts as a single-pass network that does not require a backpropagation algorithm.

A typical GRNN comprises four distinct neuron layers: the input layer, the pattern layer (which employs a radial basis function), the summation layer, and the output layer. These layers are shown in Figure 3. It is noteworthy that the number of neurons in the pattern layer is equal to the number of neurons in the training data [35]. The output of the pattern layer is subsequently presented to the summation layer, comprising neurons for the numerator and denominator. In particular, the neurons of the summation layer are primarily responsible for summarizing the output of the pattern layer, with the output y-vector containing the same number of elements as the number of numerator neurons.

TABLE IV. COMPLETE LIST AND RANKING OF CRITICAL CAUSES OF TIME DELAY IN MARINE CONSTRUCTION PROJECTS

Category	Identified Factors	*RII*
Owner	Owner's financial difficulties.	12.92
	Delay in settlement of contractor's claim by the owner.	13.05
	Interference, change orders, scope variance and slow decisions by owner.	12.59
	Delays in site delivery to contractor.	8.83
Designer	Design errors/incomplete or unclear design drawings and design variation.	10.98
	Weak and insufficient technical studies.	10.16
	Complexity of design.	7.40
	Delays in producing design documents.	9.99
Consultant	Poor planning and management of contractor's schedule.	14.22
	Inadequate quality assurance and control.	10.49
	Delay in approval of completed work (shop drawings, equipment, and material samples).	11.03
Contractor	Difficulties in project financing by contractor.	14.9
	Incompetence or inexperience of contractor (lack of experience, and or managerial skills).	13.73
	Inadequate site investigation.	10.84
	HSE considerations.	9.96
	Inadequate comprehension of scope of work at the bidding stage.	11.75
	Incompetent subcontractors deployed by the contractor.	12.50
	Rework due to non-compliance in quality or poor workmanship.	12.58
	Inappropriate construction methods and work implementation strategies by contractor.	11.60
All Parties	Lack effective communication among project stakeholders.	13.64
	Poor interaction with vendors in the engineering and procurement stages.	10.28
	Ineffective management, planning, and scheduling of project.	13.52
	Legal disputes between project participants.	9.15
Resources	Shortage of skilled labor.	11.36
	Low productivity of labor.	12.18
	Improper equipment or lack of high-tech equipment.	10.66
	Delay in material delivery.	11.56
	Shortage of construction materials—special building materials not available in the local market.	12.87
Contract	Mistakes or discrepancies in contract documents and incomplete scope definition.	9.17
	Unrealistic cost and schedules/overestimating of benefits.	11.94
	Unsuitable type of project bidding and award.	9.95
	Ineffective delay penalties.	10.82
	Poor definition of payment milestones/distribution of cash flow.	11.34
External	Climate and weather conditions.	11.06
	Sea state (waves, tides, and currents).	11.91
	Environmental degradation impacts.	9.142
	Fluctuation in cost.	13.74
	Inflation (e.g., material, equipment, and labor prices).	14.31
	The existence of sanctions and the (technical) inability to import essential goods.	13.01
	Delay in permissions, approvals and statutory compliance from authorities.	13.15
	Unforeseen Geotechnical issues.	11.08
Force Majeure	Spreading of disease, epidemic or pandemic.	10.44
	Wars in region.	10.31

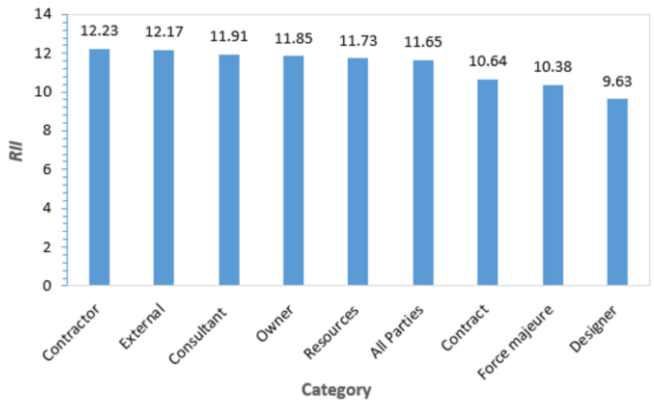


Fig. 2. Ranking categories according to *RII*.

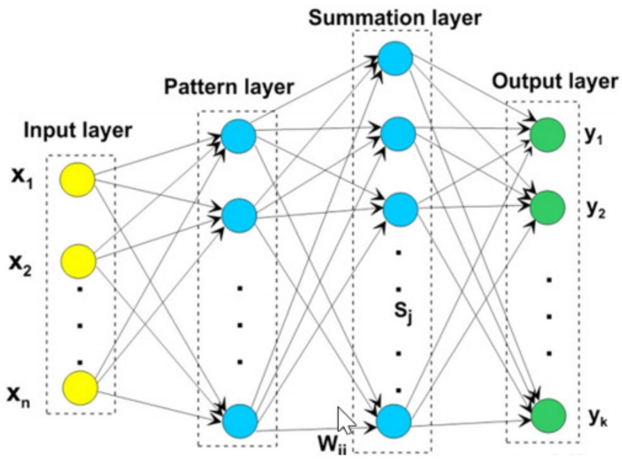


Fig. 3. A GRNN's schematic diagram.

B. *Support Vector Machine Method (SVM)*

The SVM method was initially proposed as a means of solving pattern-recognition problems. However, with the introduction of the ε-insensitive loss function, the SVM method has been expanded to estimate nonlinear regression [38]. The fundamental concept underlying SVM regression is to nonlinearly map the original data, x, into a higher-dimensional feature space, where a linear regression problem can then be solved. The SVM regression function can be expressed as follows [39]:

$$f(x) = [w \cdot \phi(x)] + b \qquad (2)$$

where $\phi(x)$ is the nonlinear mapping function, w is the weight vector and b is the bias term. The values for w and b are estimated by minimizing the regularized risk function.

Hence, the regression function is:

$$f(x) = \sum_{i=1}^{l}(\delta_i - \delta_i^*)K(x_i, x) + b \qquad (3)$$

where $K(x_i, x)$ is the Kernel function,δ_i and $\delta_i^*$are Lagrangian multipliers, and the data points corresponding to $\delta_i - \delta_i^* \neq 0$ are the support vectors. The Radial Basis Kernel Function (RBF) $K(x_i, x) = e^{(-\|x_i - x\|^2/2\partial^2)}$ was used, where δ^2 is the Kernel parameter of the RBF.

C. Tree Boost Method

The Tree Boost method is a form of boosted regression tree modeling. The incorporation of a boosting technique enables the Tree Boost method to enhance the efficacy of the decision tree approach. The fundamental concept is to integrate a collection of relatively weak models to create a robust consensus model, as opposed to developing a single optimized model. The Tree Boost method generates new decision trees in a sequential manner, with the objective of reducing existing residuals. This sequential model construction process may be described as a type of functional gradient descent. The Tree Boost model typically has three major parameters: the learning rate (also known as the shrinkage parameter) (v), tree complexity, and the number of regression trees (also known as the tree size). Cross-validation techniques are employed to evaluate the generalization capacity of the Tree Boost model and to prevent overfitting [40].

The advancement of multiple ML models necessitates two distinct processes: training and testing. The training subset is employed to determine the optimal model parameters. Subsequently, the testing process is performed to evaluate the models' capacity to generalize the knowledge acquired during the training process in cases that were not included in the training data. Approximately 80% of the data were randomly selected for model training, while the remaining 20% were employed for model testing. The predictive modeling software, Decision Tree Generator (DTREG) is employed for the purpose of predicting time delays and cost overruns for each model [41]. A summary of the training parameters employed in the ML models is provided in Table V.

TABLE V. MODEL PARAMETERS OF ML MODELS

Model	Parameters
GRNN	- Training method: conjugate gradient algorithm. - Kernel function: Gaussian; sigma (σ) = 0.0001:10 - Validation method: Leave-one-out - Number of neurons = 27
SVM	- Type of SVM model: Regression (*Epsilon*-SVR) - Kernel function: Radial Basis Function (RBF) - Free parameters of kernel function: ε = 0.5, C = 1.10717318, γ = 1.12323995, *Epsilon* = 0.001, and P = 0.02929247 - Validation method: cross-validation, and number of cross-validation folds = 10
Tree Boost	Type of model: Tree Boost series of trees Maximum trees in Tree Boost series= 400 Maximum splitting levels: = 5 Minimum size node to split = 10 Max. categories for continuous predictors: 1000 Tree pruning criterion: Minimum absolute error. Average number of group splits in each tree: 1.2 Huber's quantile cutoff = 0.95 Influence trimming factor = 0.01 Learning rate = 0.05 Validation method: Random sampling (20%)

The statistical measures were used to evaluate the ML-based models examined in this study are: Mean Square Error (*MSE*), *MAE*, Root Mean Square Error (*RMSE*), Scatter Index (*SCI*), and the Correlation Coefficient (*R*). These five criteria are defined as follows [42]:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2 \quad (4)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - O_i| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \quad (6)$$

$$SCI = \frac{RMSE}{\bar{O}} \quad (7)$$

$$R = \frac{\sum_{i=1}^{N}(P_i-\bar{P})(O_i-\bar{O})}{\sqrt{\sum_{i=1}^{N}(P_i-\bar{P})^2\sum_{i=1}^{N}(O_i-\bar{O})^2}} \quad (8)$$

where O_i is the observed value, P_i is the predicted value, N is the total number of data points, $\bar{O}$ is the observation mean value, and $\bar{P}$ is the prediction mean value.

IX. RESULT ANALYSIS

The selection of input and output variables represents a crucial aspect of the process of developing ML models. The input parameters comprise 43 factors that were previously identified as influencing time delays. The output of the ML models was a percentage value representing the estimated time delay in marine construction projects. In order to evaluate the quality of the ML models, it is essential to assess their predictive performance. In cases where an SVM approach was employed, the results were obtained through the use of a 10-fold cross-validation based on the average results obtained for test data. The cross-validation method was employed for the training and testing of the model, ensuring that all instances of the data set were applied in both stages. During the training phase, the SVM model yielded the following results: *MSE* = 107.213, *MAE* = 7.948, *RMSE* = 10.354, and *SCI* = 0.226, and *R* = 0.898. However, during the testing phase, the corresponding values were 219.622, 12.276, 14.820, 0.280, and 0.669, respectively. The results demonstrated that the *RMSE* and *SCI* calculated in the training data exhibited a 30.131% and 19.30% reduction, respectively, in comparison to those observed in the test data. Figure 4 illustrates the scatter plot of the time delay values measured and predicted by the SVM model.

The GRNN model was validated using a leave-one-out approach, with the GRNN and conjugate gradient algorithm employed for training. The training data yielded the following results: *MSE* = 21.385, *MAE* = 3.926, *RMSE* = 4.624, *SCI* = 0.101, and *R* = 0.980 for the predicted time delay values. Furthermore, regarding the test data outcomes, *MSE* was found to be 53.604, *MAE* was 7.050, *RMSE* was 7.321, *SCI* was 0.139, and *R* was 0.926 for the predicted time delay values. The index values of the test data were found to be comparable to those of the training database. The results demonstrated that *RMSE* and *SCI* values calculated in the training data were, respectively, 36.838% and 27.047% lower than those obtained in the test data. In relation to the experimental findings, the predicted values were observed to be relatively proximate to the corresponding actual values. The results demonstrate that the GRNN model is an effective predictor of the time delay associated with marine construction projects. Figure 5 illustrates the correlation between the time delay values measured and those predicted by the GRNN.

(a)

(b)

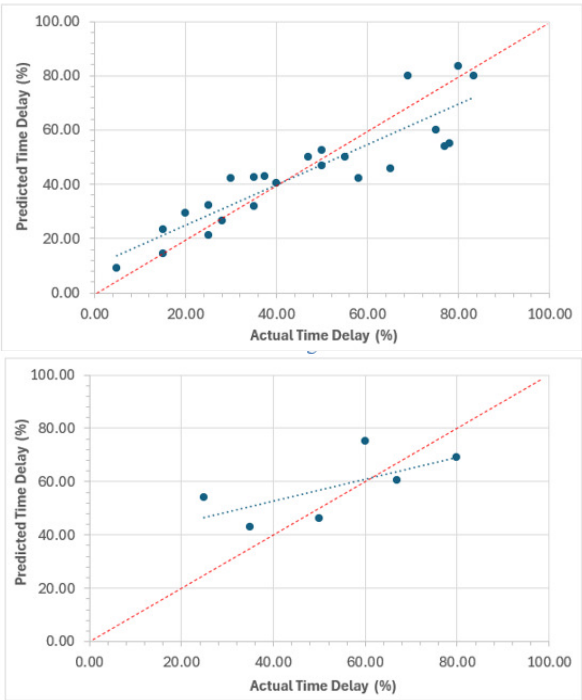


Fig. 4. Scatter plot of the measured and predicted time delay values for SVM model: (a) training data set, (b) test data set.

(a)

(b)

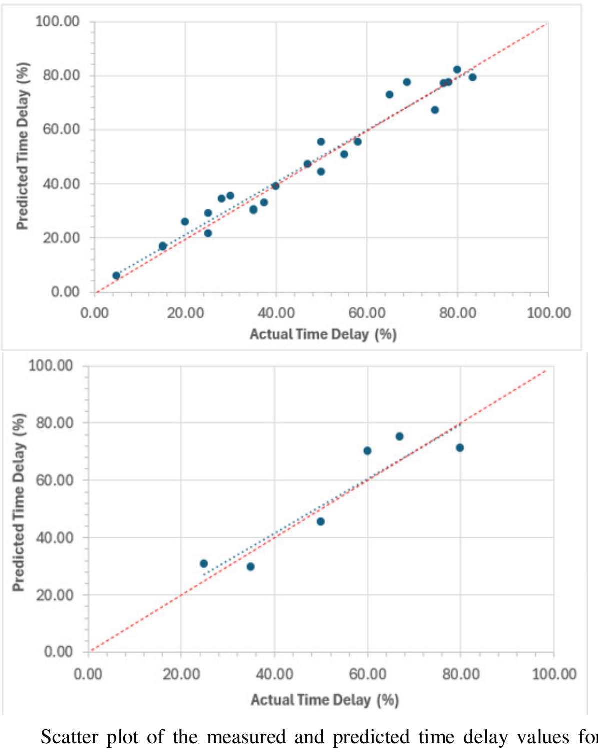


Fig. 5. Scatter plot of the measured and predicted time delay values for GRNN model: (a) training data set, (b) test data set.

The Tree Boost model represents a further extension of the predictive modeling toolkit. The statistical results for the training data are presented as follows: *MSE* is 30.386, *MAE* is 4.905, *RMSE* is 5.512, *SCI* is 0.121, and *R* is 0.972 for the predicted time delay of marine construction projects. Furthermore, with regard to the test data outcomes, *MSE* was found to be 82.757, *MAE* was 8.681, *RMSE* was 9.907, *SCI* was 0.172, and *R* was 0.892. The results demonstrated that the *RMSE* and *SCI* values derived from the training data exhibited a 39.405% and 30.013% reduction, respectively, in comparison to those obtained from the test data. Figure 6 shows the scatter plot of the time delay values observed and predicted by the Tree Boost model.

(a)

(b)

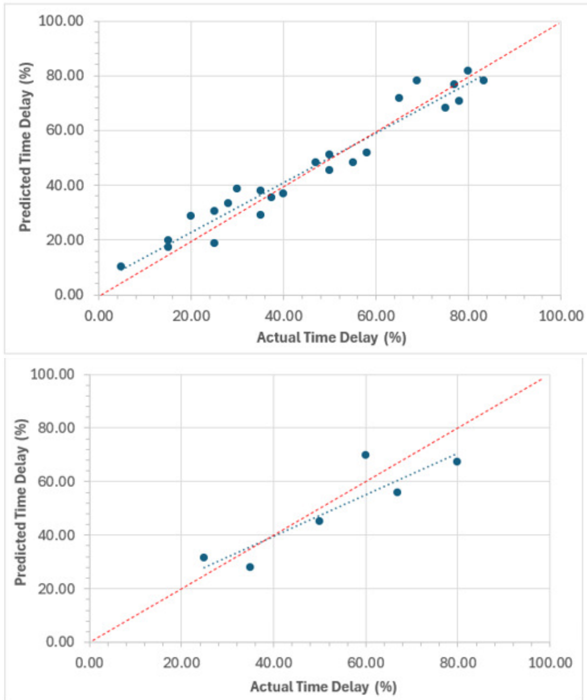


Fig. 6. Scatter plot of the measured and predicted time delay values for Tree Boost model: (a) training data set, (b) test data set.

A. Comparison between ML Models

Figure 7 shows the relationships between the actual and predicted time delay values using the various models, with a continuous diagonal line. The actual time delay at any point along the reference line is identical to that predicted by the various models for a specific observation. Should the predicted time delay values fall below the aforementioned reference line, it may be posited that the model in question is conservative in its approach. Conversely, if the predicted values exceed the reference line, it can be assumed that the model has overestimated the time delay. To facilitate the interpretation of the results, a trend line can be plotted for comparison with the reference diagonal line. In general, a model will be more accurate or more effective at predicting a time delay if the discrete points lie closer to the trend line. The trend line for the GRNN model is in close proximity to the reference line, with the trend lines for the Tree Boost and SVM models being the second and third closest, respectively. The SVM kernel model exhibits the poorest performance in terms of predicting the time delay in marine construction, as evidenced by its distance from the reference line and the observable discrepancy between its discrete points and the trend line.

(a)

(b)

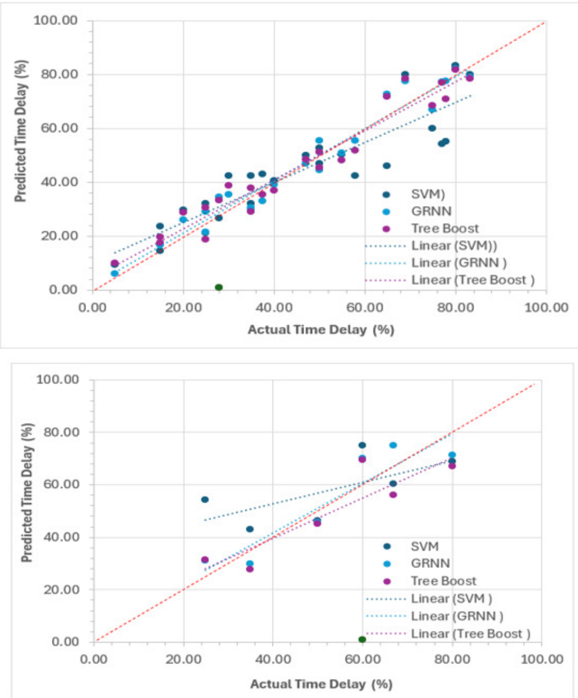


Fig. 7. Scatter plot of the measured and predicted time delay values for various ML models: (a) training data set, (b) test data set.

The GRNN model demonstrates superior predictive performance in comparison to other models, as evidenced by lower *MSE*, *MAE*, *RMSE*, *SCI*, and *R* values. The Tree Boost model exhibits comparable performance as the second-best model. With regard to the prediction of time delays, the results indicate that the GRNN model exhibits superior performance in comparison to the other models. Furthermore, the findings suggest that the GRNN model markedly minimizes the overall error and provides precise predictions regarding the time delay of marine construction projects. The SVM model displays the poorest predictive performance, indicating that it is unable to accurately represent the time delay. In conclusion, the results demonstrate that the GRNN model outperforms the other models. Moreover, the high accuracy of the GRNN and Tree Boost models is corroborated by these results. In comparison to the other models, the GRNN model produces *SCI* values that are 19.518% and 50.596% lower than those of the Tree Boost and SVM models, respectively, for the test dataset, as shown in Figure 8. In general, the outcomes of the GRNN model are more accurate than those of other models, with the GRNN model accurately predicting the time delay. Furthermore, the results demonstrate that the GRNN model is the most reliable for predicting the time delay of marine construction projects in terms of predictive accuracy, convenience, and interpretability.

The Taylor diagram offers a visual representation of the relative suitability of different ANN and SVM models, based on the values of R^2, *RMSE*, and standard deviation. A Taylor diagram is a two-dimensional space in which the predicted and actual values are positioned according to their degree of coordination. The standard deviation, *RMSE*, and R^2 are represented by the horizontal and vertical axes, circular lines, and radial lines, respectively. The accuracy of a model is determined by the extent to which it approximates the actual value. A model that is in closer alignment with the actual data is deemed to be more reliable. The GRNN model exhibits a higher *RII* score and standard deviation and a lower *RMSE* than the other models. It is situated closer to the actual values, as illustrated schematically in Figure 9, and is therefore deemed to be marginally more reliable

(a)

(b)

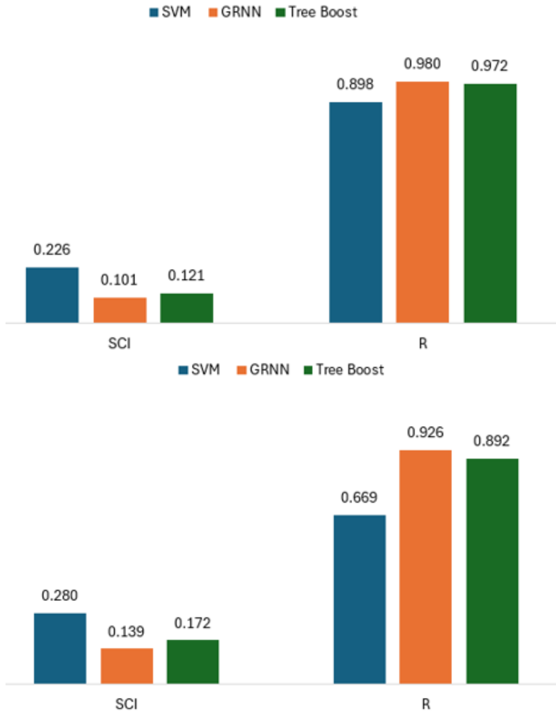
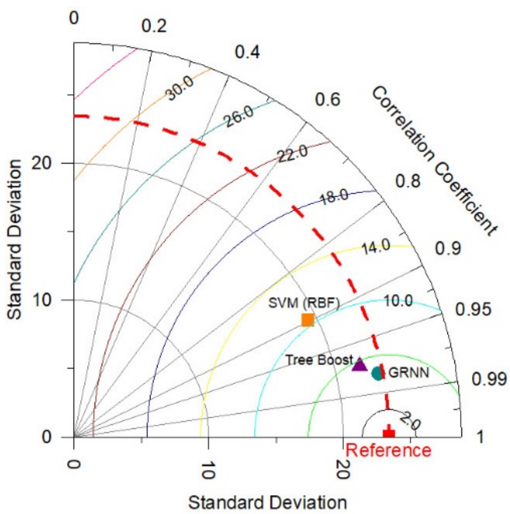


Fig. 8. Comparison of the various ML models: (a) training data set, (b) test data set.

(a)



(b)

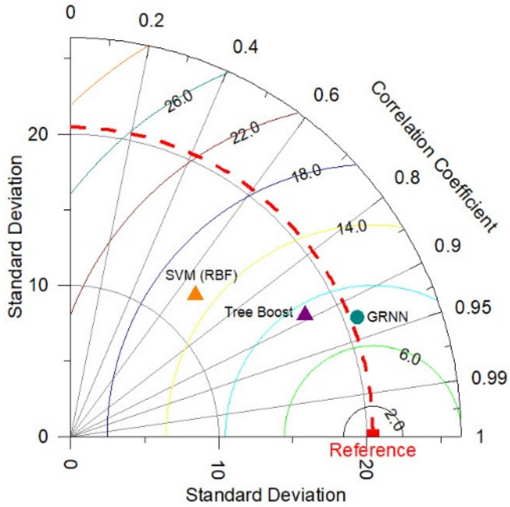


Fig. 9. Taylor diagram visualization of ML models performance for time delays prediction: (a) training data set, (b) test data set.

X. CONCULSIONS

The current study examined 43 factors contributing to time delays in marine construction projects, using data obtained from structured questionnaires. The Relative Importance Index (*RII*) was calculated and the factors were ordered according to their magnitude. Three Machine Learning (ML) models, namely General Regression Neural Networks (GRNN), Tree Boost and Support Vector Machines (SVM), were employed for the purpose of predicting time delays in marine construction projects. The data from the case studies were employed for the purposes of training and validating the various models. The predictive performance of each model was evaluated using Taylor diagram visualization and five statistical features: Mean Square Error (*MSE*), Mean Absolute Error (*MAE*), Root Mean Square Error (*RMSE*), Scatter Index (*SCI*), and Correlation Coefficient (R). The GRNN yielded highly accurate results in terms of estimating the time delay, with a SCI of 0.139, which was lower than that of the Tree Boost (0.172) and SVM (0.281) models. Furthermore, the *R* of the GRNN (0.926) was higher than that of the Tree Boost (0.892) and SVM (0.669) models. It is noteworthy that the results demonstrated that the GRNN model significantly reduced the overall error and accurately estimated the time delay for marine construction projects, thereby illustrating its superior accuracy and precision in comparison with the other models.

Overall, the GRNN model was found to be the most effective and reliable for predicting time delays in marine construction projects, highlighting its usefulness in improving project management and reducing uncertainties. Based on the results of the study on time delays in marine construction projects, the following key recommendations emerge:

- The use of GRNN for predictive analysis offers superior accuracy in predicting delays. Continuous model validation and updating with real-time data will maintain relevance and accuracy.
- Future research should focus on extending factor analysis to include environmental and regulatory variables, using diverse datasets for broader applicability. Exploring advanced machine learning techniques alongside GRNN will optimize prediction accuracy for different project contexts.

REFERENCES

[1] A. Rivera and J. Kashiwagi, "Identifying the State of the Project Management Profession," *Procedia Engineering*, vol. 145, pp. 1386–1393, Jan. 2016, https://doi.org/10.1016/j.proeng.2016.04.204.

[2] K. D. Johnson and C. B. Tatum, "Technology in Marine Construction Firms," *Journal of Construction Engineering and Management*, vol. 119, no. 1, pp. 148–162, Mar. 1993, https://doi.org/10.1061/(ASCE)0733-9364(1993)119:1(148).

[3] A. B. Bugnot et al., "Current and projected global extent of marine built structures," Nature Sustainability, vol. 4, no. 1, pp. 33–41, Jan. 2021, https://doi.org/10.1038/s41893-020-00595-1.

[4] B. Flyvbjerg, "Cost Overruns and Demand Shortfalls in Urban Rail and Other Infrastructure," *Transportation Planning and Technology*, vol. 30, no. 1, pp. 9–30, Feb. 2007, https://doi.org/10.1080/03081060701207938.

[5] F. A. Shaikh, "Financial Mismanagement: A Leading Cause of Time and Cost Overrun in Mega Construction Projects in Pakistan," *Engineering, Technology & Applied Science Research*, vol. 10, no. 1, pp. 5247–5250, Feb. 2020, https://doi.org/10.48084/etasr.3271.

[6] A. H. Nassar, "Causes and Effects of Delay in the Construction Industry in Egypt," *International Journal of Engineering Research & Technology*, vol. 5, no. 4, Apr. 2016, https://doi.org/10.17577/IJERTV5IS040434.

[7] M. H. A. Hazza, M. Muqtadar, K. E. Salamony, I. F. Bourini, A. Sakhrieh, and M. Alnahhal, "Investigation Study of the Challenges in Green Procurement Implementation in Construction Projects in UAE," *Civil Engineering Journal*, vol. 9, no. 4, pp. 849–859, Apr. 2023, https://doi.org/10.28991/CEJ-2023-09-04-06.

[8] P. L. A. Luthan, N. Sitanggang, and S. Syafriandi, "Reinventing Formulas for Construction Project Delay Index Due to Management and Production," *HighTech and Innovation Journal*, vol. 4, no. 4, pp. 768–778, Dec. 2023, https://doi.org/10.28991/HIJ-2023-04-04-06.

[9] A. Chadee, H. Ali, S. Gallage, and U. Rathnayake, "Modelling the Implications of Delayed Payments on Contractors' Cashflows on Infrastructure Projects," *Civil Engineering Journal*, vol. 9, no. 1, pp. 52–71, Jan. 2023, https://doi.org/10.28991/CEJ-2023-09-01-05.

[10] V. W. Y. Tam and L. Y. Shen, "Risk Management for Contractors in Marine Projects," *Organization, Technology and Management in Construction*, vol. 4, no. 1, pp. 403–410, Jun. 2012, https://doi.org/10.5592/otmcj.2012.1.5.

[11] H. Karami and O. A. Olatunji, "Key success factors in marine infrastructures, a review," in *Proceedings of 42nd Australasian Universities Building Education Association (AUBEA) Conference on Educating Building Professionals for the Future in the Globalised World*, pp. 121–128, 2018.

[12] H. Karami and O. A. Olatunji, "Contractor selection model for marine projects," in *Proceedings of 42nd Australasian Universities Building Education Association (AUBEA) Conference on Educating Building Professionals for the Future in the Globalised World*, vol. 2, pp. 27–37, 2018.

[13] M. Ruqaishi and H. A. Bashir, "Causes of Delay in Construction Projects in the Oil and Gas Industry in the Gulf Cooperation Council Countries: A Case Study," *Journal of Management in Engineering*, vol. 31, no. 3, May 2015, Art. no. 05014017, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000248.

[14] A. B. Seddeeq, S. Assaf, A. Abdallah, and M. A. Hassanain, "Time and Cost Overrun in the Saudi Arabian Oil and Gas Construction Industry," *Buildings*, vol. 9, no. 2, Feb. 2019, Art. no. 41, https://doi.org/10.3390/buildings9020041.

[15] J. T. Simanjuntak and A. Mahendrawan, "Factors Affecting Delay of Upstream Oil and Gas Development Projects in Indonesia: Case Study 2012 - 2013," *presented at the SPE/IATMI Asia Pacific Oil & Gas*

Conference and Exhibition, Oct. 2015, Art. no. SPE-176067-MS, https://doi.org/10.2118/176067-MS.

[16] M. Basak, V. Coffey, and R. K. Perrons, "Risk factors affecting delays in upstream natural gas mega-projects: an Australian perspective," in *SPE Asia Pacific Oil and Gas Conference and Exhibition*, 2018.

[17] N. V. Thuyet, S. O. Ogunlana, and P. K. Dey, "Risk management in oil and gas construction projects in Vietnam," *International Journal of Energy Sector Management*, vol. 1, no. 2, pp. 175–194, Jan. 2007, https://doi.org/10.1108/17506220710761582.

[18] M. E. Shehaby, M. Abdel-Hamid, and K. E. Dash, "Risk Assessment for Off-Shore Petroleum and Gas Construction Projects in Egypt," *Engineering Research Journal (Shoubra)*, vol. 51, no. 4, pp. 61–74, Oct. 2022, https://doi.org/10.21608/erjsh.2022.147582.1054.

[19] J. Lerche *et al.*, "Causes of delay in offshore wind turbine construction projects," *Production Planning & Control*, vol. 34, no. 15, pp. 1513–1526, Nov. 2023, https://doi.org/10.1080/09537287.2022.2026673.

[20] S. R. Nagireddy, "Artificial Intelligence and Its Impacts on Project Management", Ph.D dissertation, University of Cumberlands, 2023.

[21] D. Ahiaga-Dagbui and S. Smith, "Neural networks for modelling the final target cost of water projects," in *28th Annual ARCOM Conference*, Edinburgh, UK, Sep. 2012, pp. 307–316, https://doi.org/10.13140/2.1.3660.6402.

[22] V. Žujo, D. Car-Pušić, and A. Brkan-Vejzović, "Contracted price overrun as contracted construction time overrun function," *Tehnički Vjesnik*, vol. 17, no. 1, pp. 23–29, Mar. 2010.

[23] M. Juszczyk and A. Leśniak, "Modelling Construction Site Cost Index Based on Neural Network Ensembles," *Symmetry*, vol. 11, no. 3, Mar. 2019, Art. no. 411, https://doi.org/10.3390/sym11030411.

[24] W. Pewdum, T. Rujirayanyong, and V. Sooksatra, "Forecasting final budget and duration of highway construction projects," *Engineering, Construction and Architectural Management*, vol. 16, no. 6, pp. 544–557, Jan. 2009, https://doi.org/10.1108/09699980911002566.

[25] Y. R. Wang, C. Y. Yu, and H. H. Chan, "Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models," *International Journal of Project Management*, vol. 30, no. 4, pp. 470–478, May 2012, https://doi.org/10.1016/j.ijproman.2011.09.002.

[26] A. Asadi, M. Alsubaey, and C. Makatsoris, "A machine learning approach for predicting delays in construction logistics," *International Journal of Advanced Logistics*, vol. 4, no. 2, pp. 115–130, May 2015, https://doi.org/10.1080/2287108X.2015.1059920.

[27] Z. M. Yaseen, Z. H. Ali, S. Q. Salih, and N. Al-Ansari, "Prediction of Risk Delay in Construction Projects Using a Hybrid Artificial Intelligence Model," Sustainability, vol. 12, no. 4, Jan. 2020, Art. no. 1514, https://doi.org/10.3390/su12041514.

[28] C. N. Egwim, H. Alaka, L. O. Toriola-Coker, H. Balogun, and F. Sunmola, "Applied artificial intelligence for predicting construction projects delay," *Machine Learning with Applications*, vol. 6, Dec. 2021, Art. no. 100166, https://doi.org/10.1016/j.mlwa.2021.100166.

[29] P. Velumani, N. V. N. Nampoothiri, and M. Urbański, "A Comparative Study of Models for the Construction Duration Prediction in Highway Road Projects of India," *Sustainability*, vol. 13, no. 8, Jan. 2021, Art. no. 4552, https://doi.org/10.3390/su13084552.

[30] R. Al mnaseer, S. Al-Smadi, and H. Al-Bdour, "Machine learning-aided time and cost overrun prediction in construction projects: application of artificial neural network," *Asian Journal of Civil Engineering*, vol. 24, no. 7, pp. 2583–2593, Nov. 2023, https://doi.org/10.1007/s42107-023-00665-7.

[31] S. F. Olejnik, "Planning educational research: Determining the necessary sample size," *Journal of Experimental Education*, vol. 53, no. 1, pp. 40–48, 1984, https://doi.org/10.1080/00220973.1984.10806360.

[32] J. C. Nunnally and I. H. Bernstein, "*Psychometric theory McGraw-hill series*," Psychology, vol. 3, 1994.

[33] J. M. Cortina, "What is coefficient alpha? An examination of theory and applications," *Journal of Applied Psychology*, vol. 78, no. 1, pp. 98–104, 1993, https://doi.org/10.1037/0021-9010.78.1.98.

[34] J. Song, C. E. Romero, Z. Yao, and B. He, "A globally enhanced general regression neural network for on-line multiple emissions prediction of utility boiler," *Knowledge-Based Systems*, vol. 118, pp. 4–14, Feb. 2017, https://doi.org/10.1016/j.knosys.2016.11.003.

[35] D. Z. Antanasijević, M. Đ. Ristić, A. A. Perić-Grujić, and V. V. Pocajt, "Forecasting human exposure to PM10 at the national level using an artificial neural network approach," *Journal of Chemometrics*, vol. 27, no. 6, pp. 170–177, 2013, https://doi.org/10.1002/cem.2505.

[36] O. Kisi, M. Tombul, and M. Z. Kermani, "Modeling soil temperatures at different depths by using three different neural computing techniques," *Theoretical and Applied Climatology*, vol. 121, no. 1, pp. 377–387, Jul. 2015, https://doi.org/10.1007/s00704-014-1232-x.

[37] M. M. Lotfinejad, R. Hafezi, M. Khanali, S. S. Hosseini, M. Mehrpooya, and S. Shamshirband, "A Comparative Assessment of Predicting Daily Solar Radiation Using Bat Neural Network (BNN), Generalized Regression Neural Network (GRNN), and Neuro-Fuzzy (NF) System: A Case Study," *Energies*, vol. 11, no. 5, May 2018, Art. no. 1188, https://doi.org/10.3390/en11051188.

[38] R. Barzegar and A. A. Moghaddam, "Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction," *Modeling Earth Systems and Environment*, vol. 2, no. 1, Jan. 2016, Art. no. 26, https://doi.org/10.1007/s40808-015-0072-8.

[39] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation and signal processing," in *Proceedings of the 9th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA, Dec. 1996, pp. 281–287.

[40] J. H. Friedman and J. J. Meulman, "Multiple additive regression trees with application in epidemiology," *Statistics in Medicine*, vol. 22, no. 9, pp. 1365–1381, May 2003, https://doi.org/10.1002/sim.1501.

[41] "Predictive Modeling & Analysis Software," *DTREG*. https://www.dtreg.com/

[42] M. S. Elbisy, "Machine learning techniques for estimating wave-overtopping discharges at coastal structures," *Ocean Engineering*, vol. 273, Apr. 2023, Art. no. 113972, https://doi.org/10.1016/j.oceaneng.2023.113972.