RESEARCH

Harnessing machine learning and structural equation modelling to quantify the cost impact of rework in bridge projects

Sara M. Elseufy[1] · Ayman Hussein[2] · Mohamed Badawy[2]

Abstract
The study addresses the significant challenge of rework in the construction industry by leveraging machine learning techniques. Specifically, the aim is to develop models that accurately classify the impact of rework causes on the cost performance of bridge projects using objective data sources. Pertinent rework sources and determinants were identified, and a multivariate dataset of prior projects' cost performance was assembled. Additionally, a structural equation model was developed to calculate the impact of these factors on cost performance in bridge projects. To create a suitable dataset for machine learning, 272 responses from subject matter experts were utilized. The study explores Ensemble techniques, K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM). Cross-validation tests were conducted to assess the predictive abilities of the models, and the evaluation results indicated that the SVM model provides superior predictive performance for the dataset examined. SVM achieves 98.53% (89.54%) accuracy in training (testing) with a 1.47% (10.46%) misclassification error. Comparisons were made regarding the impact of rework on cost, with SVM achieving the highest recognition rate across all data divisions, followed by ANN. Conversely, KNN exhibited the lowest recognition rate among the classifiers. With a maximum recognition rate of 97%, SVM emerged as the best classifier. The optimal data separation for testing and training data was determined to be 10% and 90%, respectively. Overall, this study harnesses the power of machine learning to facilitate evidence-based decision-making, enabling proactive prediction of the impact of rework on cost performance in bridge projects.

Keywords Rework · Structural equation modeling · Machine learning · KNN · SVM · ANN

Introduction

The construction industry is plagued by the problem of rework, which results in significant cost overruns, delays, and lower quality of work (Flapper et al., 2002); (Mohamed et al., 2021). Despite the efforts of various quality management philosophies such as lean and total quality management, rework remains a prevalent issue (Khalesi et al.,

✉ Sara M. Elseufy
 sara.elseufy@eng.asu.edu.eg

 Ayman Hussein
 Ayman@adec-arabia.com

 Mohamed Badawy
 Mohamed_Badawy@eng.asu.edu.eg

[1] Department of Structural Engineering, Ain Shams University, Cairo, Egypt

[2] Structural Engineering Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt

2020). The cost of rework represents a substantial portion of a project's overall cost, and it can even result in the failure of an entire project (P. Love & Smith, 2018); (Al-Janabi et al., 2020). Rework can be reduced, and construction cost performance can be improved by understanding the influence of rework costs and its (Hwang et al., 2009). Despite the extensive research on building cost estimation, forecasting construction cost overruns caused by rework has received very little attention (Ye et al., 2015); (Shoar et al., 2022); (de Oliveira et al.'s, 2024)]. According to (Jafari & Rodchua, 2014) the nonconformance cost accounted for more than 12% of the overall contract value. (P. E. D. Love & Sing, 2013) determined the direct and indirect impacts of rework on overall construction cost to be 26% and 52%, respectively, by a questionnaire survey on various project types and procurement routes. Rework expenses hurt the associated plans, such as those for time, cost, and human resources, which lowers construction productivity and costs the project participants money and harms their reputation.

Contractors frequently use an internal quality assurance and control system and proactive methods to foresee potential rework and associated costs to lessen the severity of this issue. Therefore, it is crucial to estimate the Cost for each work item to improve decision-making and planning, as well as to raise the likelihood that a construction project will be successful (Badawy et al., 2021).

Recent developments in civil engineering technology necessitate higher levels of accuracy, efficiency, and speed in the analysis and design of the relevant systems (A Kaveh et al., 2017). Artificial intelligence (AI) and structural equation modeling (SEM) were chosen for testing due to their distinct capabilities in addressing different aspects of the research problem. Firstly, artificial intelligence, including techniques such as support vector machines (SVM), artificial neural networks (ANN), and k-nearest neighbors (kNN), offers a powerful toolset for data analysis and pattern recognition. These algorithms can handle complex, nonlinear relationships between variables and have been successfully applied in various predictive modeling tasks. In the context of our research, AI provides a robust framework for forecasting the impact of rework on construction costs by leveraging historical data and identifying underlying patterns.

On the other hand, structural equation modeling (SEM) is a statistical technique used to test and validate complex theoretical models. SEM allows researchers to examine the relationships between latent variables and observed indicators, providing insights into the underlying structure of a phenomenon. In our study, SEM enables us to explore the causal relationships between different factors contributing to rework and its impact on project costs. By incorporating SEM into our analysis, we can assess the theoretical validity of our proposed model and identify key drivers of rework-related cost overruns in bridge projects. Overall, the combination of artificial intelligence and structural equation modeling offers a comprehensive approach to understanding and predicting the impact of rework on construction costs. By harnessing the strengths of both techniques, we aim to develop a robust predictive model that can inform decision-making and resource allocation in bridge construction projects.

Machine learning is a reliable tool for informed decision-making based on historical data, but there's a lack of advanced strategies like ensemble learning in predicting construction costs and rework. Existing studies don't delve into the resilience of these estimators or their applicability across datasets. This research aims to fill this gap by investigating rework causes, using structural equation modeling, and employing machine learning to forecast rework costs in bridge projects. The study employs sophisticated ML perspectives to address the diversity in the construction industry and aims to enhance strategies for mitigating rework and improving cost performance in bridge projects.

Overview about cost of rework in construction

While there has been significant focus on the estimation of construction costs in past studies, there has been comparatively less emphasis on predicting instances of construction cost overruns (Dahanayake & Ramachandra, 2016); (Taha et al., 2022);(Yap et al., 2016); (Fayek et al., 2004). Inadequate attention has been given to accurately estimating the cost overruns associated with construction rework. This review examines the existing literature on predicting the cost impact of construction cost overruns and explores various construction cost estimation methods within a wider context. The literature reviewed in this study shows a consensus regarding the adverse effects of COR on the total construction cost, but the impact percentages vary depending on the characteristics and types of the construction projects being evaluated (Chidiebere & Ebhohimen, 2018). Rework on building projects is directly related to cost and time overruns. This includes the increased time required to complete rework, additional costs associated with fixing the defect, increases in material usage brought on by rework, wastes, and rising labor costs associated with fixing the defects (Palaneeswaran et al., 2005).Rework expenses and initial costs, cost overruns, initial times, and time overruns all significantly correlate with one another (Oke & Ugoje, 2013). (Oyewobi et al., 2016) reported a correlation between time overrun and rework costs that was positive, indicating that processing nonconforming work would cost more and take longer to complete.

It's important to note that rework is not always a frustrating problem in building projects—quite the opposite, in fact (Shoar et al., 2022). Rework is caused by four main factors: mistakes, omissions, revisions, and damages during construction. A thorough evaluation of the rework literature in construction projects found that expenses associated with rework ranged from 1 to 20% of the contract price (P. Love & Smith, 2018). Various evaluations have been done on the cost of rework in the construction industry; for instance, it has been projected that the overall cost of rework in infrastructure projects will be close to 10% of the contract value (P. E. D. Love et al., 2010). The mean cost of rework in residential projects is between 10 and 15% (Mahamid, 2022). (Meshksar, 2012) declared that rework cost between 1.30% and 3.30% of the contract value and took between 3.0% and 8.0% of the project's total time to complete. (Wasfy, 2010)discover that the rework expansions in the expense as 2–30% and schedule delay ranged between 10–70% for various work classes. According to (Abeku et al., 2016), a building project in Nigeria had rework costs of 12.58% and time overruns of

38%. Typically, the cost of rework includes both direct and indirect costs. The indirect costs of rework include costs that have no replacement in terms of money, like wear and tear on employees, inertia with clients, and customer dissatisfaction at the project level, as well as diminished benefits, conflicts, and lost opportunities for future work at the organizational level. Rework can cost up to (3–6) times as much directly as actual work (P. E. d. Love, 2002b). According to (P. E. D. Love, 2002a), the direct and indirect costs of rework were 6.4% and 5.6% of the total contract value. The average cost of rework in New South Wales, Australia, was 5.5% of the contract's value; it consisted of 2.75% direct costs, 1.75% indirect costs for key project participants, and 1% indirect costs for individual subcontractors (Marosszeky, 2006).

Overview about machine learning techniques

One of the most exciting techniques in predictive data analytics is machine learning. To extract trends, correlations, patterns of interest, and practical insights from large, complex data sets, it combines techniques from statistics, database analysis, data mining, pattern recognition, and artificial intelligence (Flath et al., 2012). Metaheuristics are algorithms used to tackle complex optimization problems that are challenging for traditional techniques (A Kaveh et al., 2023). Neural networks estimate seismic damage in structures (Rofooei et al., 2011). Neural networks are trained and applied for predicting natural frequencies (A Kaveh et al., 2015). A summary of the popular modern machine learning algorithms employed in this work (ANN, KNN, and SVM) is provided in this section. References that are pertinent will provide detailed mathematical descriptions of these strategies (Witten et al., 2005); (Olatunji, 2017); (Sethi et al., 2017). Based on classes related to the k-instances in the training set, the k-nearest neighbour (KNN) classifier predicts new classes using a simple majority decision method. KNN is a popular and straightforward technique that uses historical data to locate a data point's closest neighbours (Wauters & Vanhoucke, 2017). The value of k has a significant impact on how accurately predictions made using the k-NN model turn out (Sethi et al., 2017) For instance, when there are k143 neighbours, the instance belongs to the class that the three closest neighbours make up the majority (Larrañaga et al., 2018).

According to (Witten et al., 2005), support vector machines combine instance-based learning with linear modelling. Support Vector Machine (SVM) analysis is conducted on the standard penetration test (SPT) dataset, both with and without incorporating weights obtained from all the utilized objective weighting methods (Alla et al., 2023). SVM is one of the most popular machine learning methods now used for both classification and regression issues (Olatunji, 2017) In order to identify the hyperplane (decision boundaries) that results in an effective separation of classes, the theoretical foundation of SVM is built on structural risk minimization and statistical learning theory (Sethi et al., 2017) The borders are created by choosing a few crucial boundary instances known as support vectors. Using this, a linear discriminant function is constructed to separate them as broadly as possible. By employing additional nonlinear elements in the function through its instance-based methodology, SVM can create quadratic, cubic, and higher-order decision limits (Witten et al., 2005) Rework literature lacks exploration of ML techniques like KNN and SVM. Their effectiveness in cost performance compared to existing methods needs examination. ML can enhance construction cost estimation. (Yi & luo, 2023). Prior studies examined rework causes' impact on cost, with limited machine learning use or accuracy comparisons to traditional methods. Existing classification methods may be subjective or lack complexity consideration. This study explores machine learning's potential to predict rework impact on bridge project cost, contributing to proactive understanding. Insights gained may extend to applying machine learning in various industries or contexts.

Methodology

Significant machine learning technology is used with structural equation modelling (SEM) in this study. The analysis is divided into two stages. The first phase is concerned with SEM, which is separated into two steps: measurement model validation and structural model hypothesis testing. Machine learning (ML) is used in the second phase to train and test the proposed model, as well as to quantify the effectiveness of independent factors on the dependent factor. Machine learning may be beneficial in research contexts with predictive scope, weak theory, and little requirement for comprehension of underlying links.

The primary causes of rework on bridge constructions will be initially determined through this study. Then, based on the likelihood and seriousness of the measured variables and the latent variables, create a model utilizing the structural equation model to forecast the overall impact of rework in bridge projects. By contrasting the SEM model's output with the actual output from the site, the model will then be validated. The stages of a comprehensive approach to this research are depicted in Fig. 1. The following is a list of the study's itemized methodology:

1. Defining the research problem and goals, which involve investigating the specific causes and determinants of
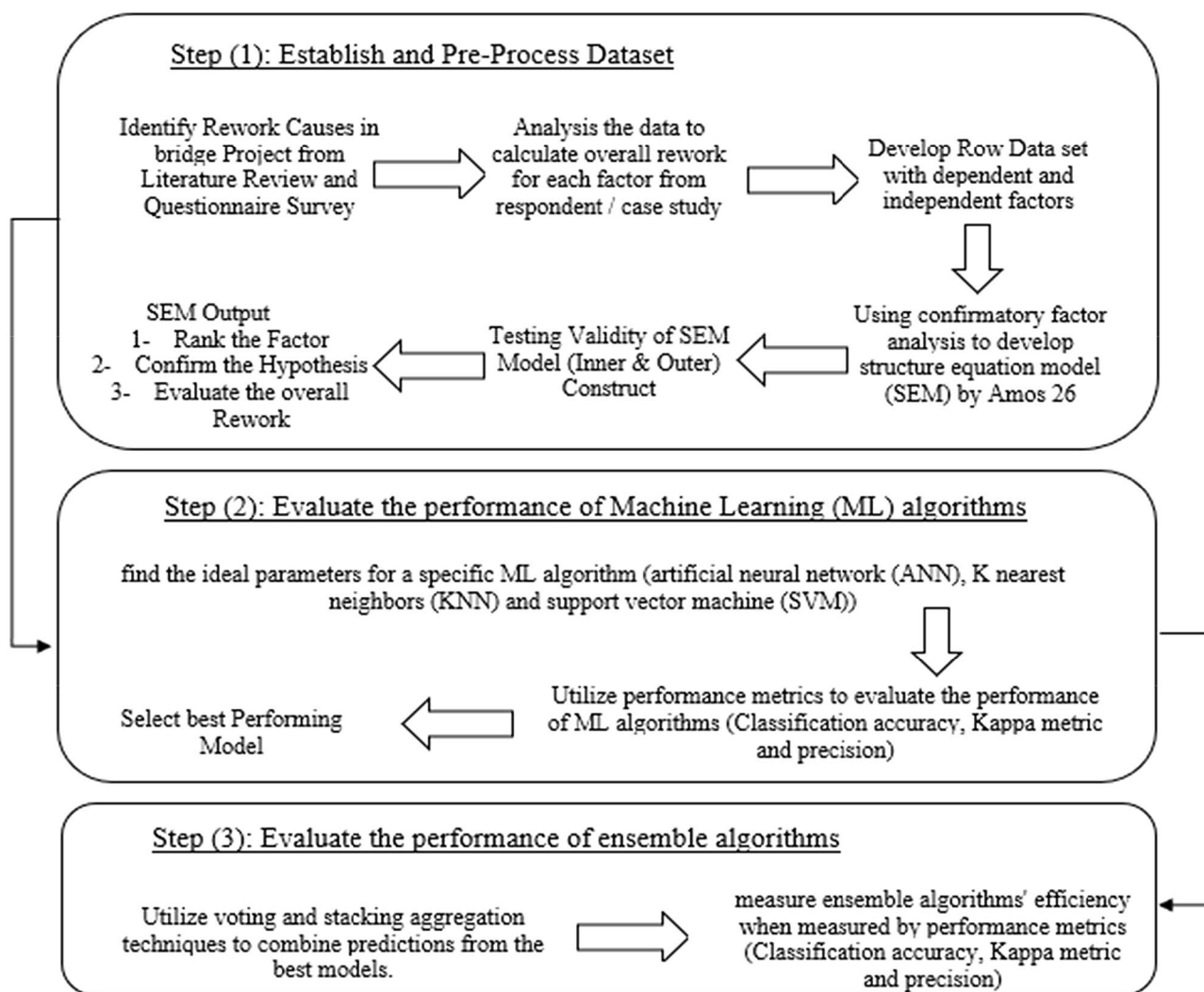
Fig. 1 The Methodology of research

rework in bridge projects and understanding their impact on cost performance.

2. Conducting a comprehensive literature review to identify previous studies on rework in bridge projects, particularly those that have used objective data sources and machine learning techniques.

3. Developing a research framework that incorporates structural equation modelling (SEM) and machine learning (ML) to investigate the causes and impact of rework on cost performance in bridge projects.

4. Using a combination of objective data sources, such as project reports and financial statements, and subjective data sources, such as expert opinions and surveys, to identify and classify the specific causes and determinants of rework in bridge projects.

5. Conducting field visits to bridge projects to collect data on the identified causes and determinants of rework and assessing their likelihood and impact using a five-point Likert scale questionnaire.

6. Using SEM to validate the proposed model and test the hypotheses on the relationships between the identified causes and determinants of rework, and their impact on cost performance in bridge projects.

7. Developing ML-based perspectives to forecast the cost of rework associated with various bridge operations, considering both measured and latent variables.

8. Applying the ML-based perspectives to a case study to ensure the validity and accuracy of the proposed model, in addition to the possibility of applying it to future projects.

9. Drawing conclusions and providing recommendations on how to mitigate the risk of rework and improve cost performance in bridge projects based on the findings of the study.

SEM (Structural Equation Modeling) served as the initial step, enabling the modeling of intricate relationships and hypothesis testing about causal connections among variables. This approach aids in identifying latent variables linked to rework costs, integrating them into machine learning models. SEM facilitates the assessment of various factors' impact on rework costs, pinpointing the most influential contributors to overruns. It proves to be a robust tool for comprehending intricate relationships in bridge construction projects, enhancing the accuracy and effectiveness of machine learning models for predicting rework costs.

The study aimed to develop an accurate and implementable predictor for construction cost overrun impact without employing techniques like dimension reduction or class imbalance adjustments. Instead, a voting classifier with various ML models was utilized, adjusting members based on performance to expedite training. ML, in this context, specifically refers to predicting rework costs in bridge construction. The authors leveraged ML to identify rework causes, offering a nuanced understanding of factors influencing cost performance. Essential parameters were set as benchmarks, with default options for others, excluding feature engineering and optimization due to the study's scope. Strong and weak ML models enhanced accuracy and overall model performance without employing complex optimization techniques.

Questionnaire Design

After reviewing prior research on rework in construction projects, 86 variables were identified and grouped into human resources, construction phase, design phase, external factors, and a special group for bridge projects. Thirty-three components remained after omitting those not cited at least three times. Fifteen experts with over 14 years of combined experience in road and bridge projects used Delphi technology to reach consensus on bridge rework reasons in Egypt.
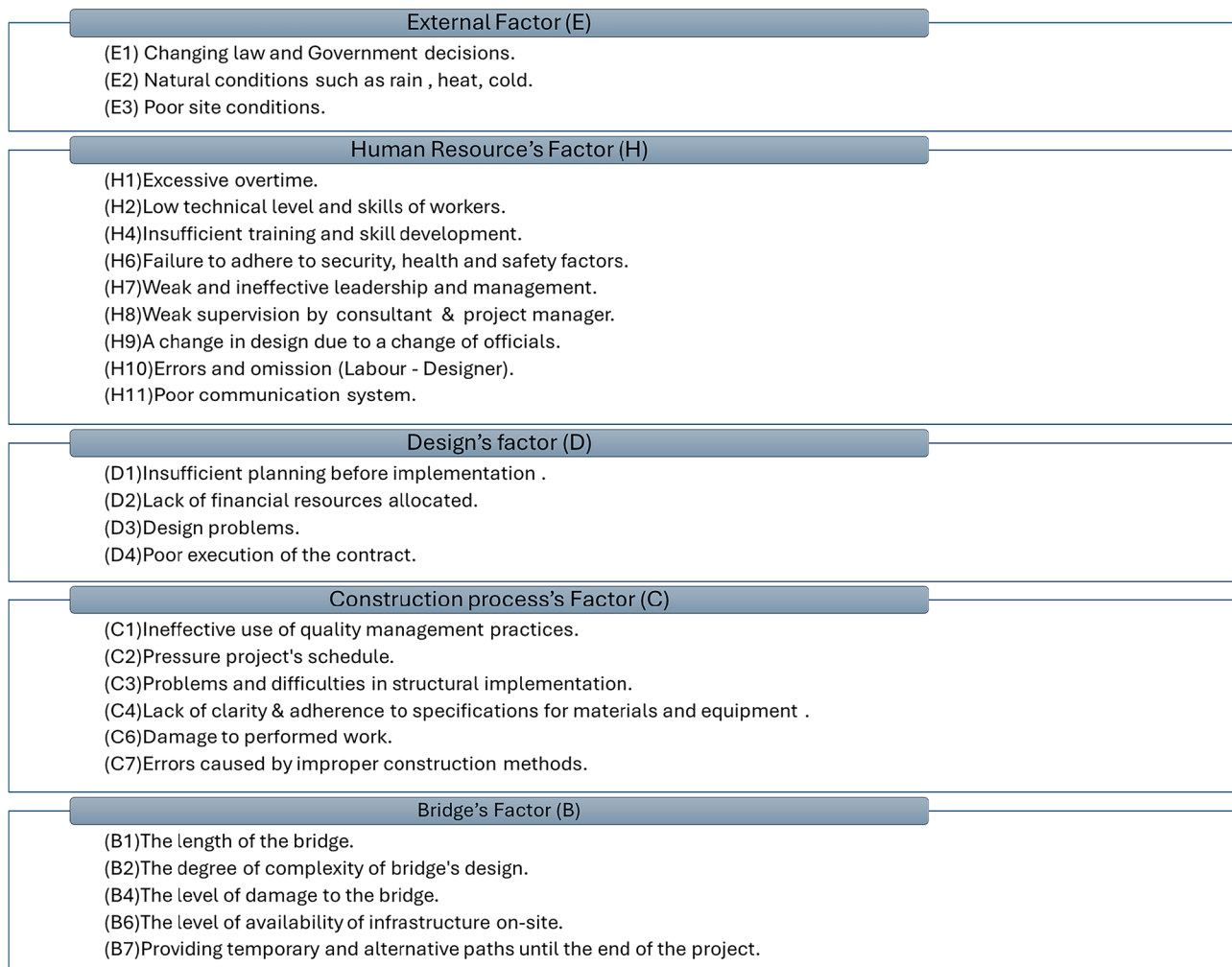
External Factor (E)
(E1) Changing law and Government decisions.
(E2) Natural conditions such as rain , heat, cold.
(E3) Poor site conditions.

Human Resource's Factor (H)
(H1)Excessive overtime.
(H2)Low technical level and skills of workers.
(H4)Insufficient training and skill development.
(H6)Failure to adhere to security, health and safety factors.
(H7)Weak and ineffective leadership and management.
(H8)Weak supervision by consultant & project manager.
(H9)A change in design due to a change of officials.
(H10)Errors and omission (Labour - Designer).
(H11)Poor communication system.

Design's factor (D)
(D1)Insufficient planning before implementation .
(D2)Lack of financial resources allocated.
(D3)Design problems.
(D4)Poor execution of the contract.

Construction process's Factor (C)
(C1)Ineffective use of quality management practices.
(C2)Pressure project's schedule.
(C3)Problems and difficulties in structural implementation.
(C4)Lack of clarity & adherence to specifications for materials and equipment .
(C6)Damage to performed work.
(C7)Errors caused by improper construction methods.

Bridge's Factor (B)
(B1)The length of the bridge.
(B2)The degree of complexity of bridge's design.
(B4)The level of damage to the bridge.
(B6)The level of availability of infrastructure on-site.
(B7)Providing temporary and alternative paths until the end of the project.

Fig. 2 Causes of rework in final structural equation modeling

Through a three-step process, the initial list was refined to 27 relevant elements in (Fig. 2).

The questionnaire comprises two main sections. The first covers participants' introductory characteristics, while the second explores rework problems in Egypt's bridge projects. For each element, respondents answer three questions regarding frequency, cost impact, and timeline effect. Weights are assigned using a five-point Likert scale: 1 for insignificance, 2 for low significance, 3 for medium significance, 4 for high significance, and 5 for potential criticality.

Data collection

The survey targeted diverse respondents, including owners, consultants, and contractors engaged in bridge projects, aiming for varied perspectives and data richness (Elseufy et al., 2022). To ensure broad participation, 300 survey forms were emailed, with a sample size determined using the Cochrane formula. Follow-up reminders were employed, resulting in a 86.7% (260) genuine response rate within 8 weeks, surpassing the 30% threshold for reliable statistical analysis (Ye et al., 2015). The respondents comprised 24.6% owners, 38.5% consultants, and 36.9% contractors. On average, 39.2% of Egyptian bridge projects had at least 14 years of experience. More than 55.4% of respondents were involved in new bridge construction projects, contrasting with the 44.6% participating in bridge restoration projects.

Data Analysis

To make sure the information from the questionnaire was accurate and could be applied to structural equation modelling, it was analysed using the SPSS program Version 26 (Elseufy et al., 2022). The validity of the data was examined, as well as the compatibility of all participants in completing the questionnaire. The indications from the questionnaire yielded a Cronbach's alpha coefficient of 0.917, which is more than the lower limit for constructing a good internal consistency (0.7) (Hair et al., 2010). The study data was then further analysed utilizing supplementary methods and statistical measures as described by (Yap et al., 2019).

Spearman's rank correlation, a non-parametric measure, assessed the relationship strength between variables' groups on an ordinal scale. The Relative Significance Index (RII), depicting Spearman test results, indicates strong agreement among project partners on factors influencing bridge project rework. Owners and contractors exhibit the lowest agreement (around 76%), while consultants and contractors demonstrate the highest (approximately 89%), enhancing result consistency.

Model Specification

According to (Brown et al., 2017), AMOS is a program that integrates multivariate frameworks for inspection and analysis, such as factor correlation, variance analysis, regression, and factor analysis (Elseufy et al., 2022). Models produced by AMOS are more precise than those obtained by using conventional multivariate statistical techniques. AMOS's graphical user interface makes it simple to use (Arbuckle, 2011). Even when conventional indicators are not met, a structural equation model based on speculative projections and prior empirical findings is initially appropriate (Brown et al., 2017). The latter model ought to meet the suggested goodness of fit metrics (GOF). AMOS 26.0 was used to create the suggested second-order primary confirmation factor analysis model. Structural equation modeling was utilized to investigate the interconnections between these criteria and the overall desirability (Ojghaz et al., 2023).

The 27 rework causes in bridge projects were categorized into five groups. A second-order confirmatory factor analysis model, with overall rework as a second-order latent variable, included human resources, construction, design process, external factors, and bridge groups as first-order latent variables. Both exogenous and endogenous constructs influenced rework. Table 1 outlines the study's hypotheses.

Model Refinements

The goal of the outer measurement model is to determine the validity, internal consistency, and reliability of both observable and latent variables. The reliability test, which

Table 1 The model hypotheses

Hypothesis	The latent variable	Descriptions
Hypothesis (H1)	The human resources factors	has a significant and positive effect on the overall impact of rework on cost's bridge project performance
Hypothesis (H2)	The construction process factors	
Hypothesis (H3)	The design process factors	
Hypothesis (H4)	The external factors	
Hypothesis (H5)	The bridge-related factors	

is monitored by the physical model, is dependent on the assessment of consistency between latent components. To evaluate legitimacy, convergent and discriminatory validity are used. Variables reported with an outer loading of 0.50 or above are regarded as highly accepted (Hair et al., 2012). It is recommended to remove any factor with an outside loading of less than 0.50 (Chen, 2007). Figure 3 shows the SEM Model after deleting the elements that have an influence factor of less than 0.5.

One of the final model's outputs is the regression analysis for overall rework impact on cost. Equation No. 5 expresses the effect of rework on the cost performance of bridge projects (Elseufy et al., 2022).

$$\text{Re} - work = 0.93\,C + 0.69\,E + 0.85\,H + 0.52\,B + 0.43\,D \tag{1}$$

After verifying the accuracy and reliability of measurements in the outer model, the inner structural model's validity was assessed, focusing on predictive utility and inter-construct connections. The primary evaluation criterion is the Goodness-of-Fit (GOF) index, ranging from 0 to 1. A model

Table 2 Goodness-of-Fit

GOF measures	Final SEM
The goodness-of-fit index (GFI)	0.926
Root means-square-error-of-approximation (RMSEA)	0.057
The comparative-fit-index (CFI)	0.964
The Tucker-Lewis coefficient (TLI)	0.942
The normed-fit-index (NFI)	0.927
The incremental-fit-index (IFI)	0.965

is deemed reasonable and realistic with strong fit indicators (Hair et al., 2012). Table 2 presents the GOF measures, ensuring the model faithfully represents observed data.

Machine Learning

Machine learning is an emerging area of artificial intelligence that is used for data modelling, or creating mathematical abstractions of data that computers can use to make
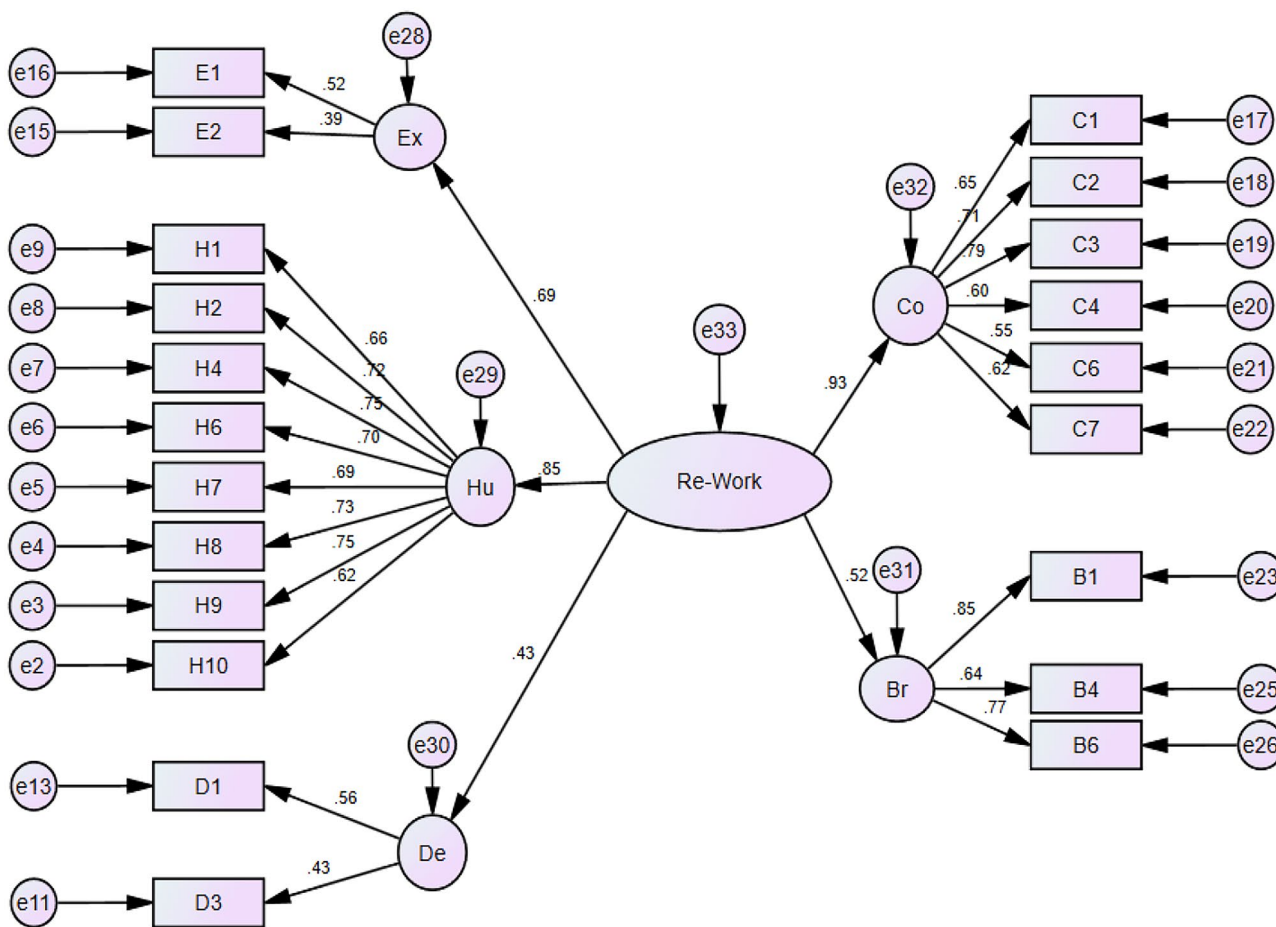

Fig. 3 Causes of rework in bridge projects: structural equation model

precise predictions. A significant subset of machine learning issues is supervised classification. The primary goal of the proposed methodology is to train diverse supervised learning classifiers using a designated training dataset (Entezami et al., 2020). The instance space, which contains a set of independent variables, the label space, which contains the dependent variable for each instance, and the machine learning algorithm are the three basic components of supervised classification (Larrañaga et al., 2018).

Data pre-processing

Identifying the optimal feature combination is crucial for improving a model's predictive ability and addressing the "curse of dimensionality" in machine learning challenges. MATLAB R2022b was employed for the optimization process, leveraging its well-suited features and resources for machine learning applications (A Kaveh et al., 2022). The dataset, consisting of 272 bridge projects with varying rework impacts on cost performance, was used to construct a predictive analytics tool. Each data record represents a unique project linked to six variables. The first variable, representing the predicted impact of rework on bridge costs based on a structural equation model, serves as the dependent variable. The remaining five variables, reflecting rework causes, are considered independent. The dependent variable was categorized into three labels (< 30% LOW, 30–60% MEDIUM, and > 60% HIGH) based on its frequency distribution. The dataset was split into a 70% training and 30% testing ratio.

Evaluate machine learning algorithms

The fine-tuning of optimization hyperparameters affects how well ML algorithms work. In this work, a systematic search was used, and the model was gradually trained using different sets of hyperparameters until sufficient results were attained. In machine learning, ensemble approaches can enhance the performance of base classifiers. Because the classification technique is good at dealing with categorical variables and can manage complex-related variables, it was the focus of the current investigation. This strategy involves combining the base classifiers' predictions through a set

process. The ML algorithms used in this study are described as follow: (artificial neural network (ANN), K nearest neighbors (KNN) and support vector machine (SVM)).

Performance measurement

Confusion matrices were employed to assess algorithm performance, providing values for True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). In binary classification, these metrics indicate correct and incorrect classifications. For multi-class matrices, these values are computed independently for each class, and a weighted average is calculated. TPs are correct classifications, TNs are correctly identified instances outside the considered class, FNs are instances mistakenly categorized, and FPs are samples from other classes. Classification accuracy and Cohen's kappa statistic are used for performance evaluation in cases of asymmetrical class distribution. The confusion matrices for ANN, KNN, and SVM classifiers are presented in Table 3.

$$Classification\,Accuracy = \frac{TP + TN}{N}; \qquad (2)$$

where N = No. of samples & Misclassification Error = (1 _ Accuracy).

$$Cohen's\,kappa\,statistic = \frac{\frac{TP}{N} + \frac{TN}{N} - A}{1 - A}; where$$

$$A = (\frac{FN+TP}{N})\left(\frac{FP+TP}{N}\right) + \left(\frac{FP+TN}{N}\right)(\frac{FN+TN}{N})\,(3)$$

$$Precision\,(Positive\,Predictive\,Value) = \frac{TP}{TP + FP} \qquad (4)$$

$$TP\,Rate\,(Sensitivity) = \frac{TP}{TP + FN} \qquad (5)$$

$$FP\,Rate\,(1\_Specificity) = \frac{FP}{FP + TN} \qquad (6)$$

Validation approaches

Two validation approaches were employed. Initially, models were trained on the complete dataset, predicting class labels of the same data used for training due to dataset limitations.

Table 3 Confusion matrix for all performing models

Actual class	Predicted class											
	MODEL (1) (SVM)				MODEL (2) (KNN)				MODEL (3) (ANN)			
	High	Medium	Low	Total	High	Medium	Low	Total	High	Medium	Low	Total
High	42	0	1	43	43	0	0	43	43	0	0	43
Medium	0	62	2	64	0	60	4	64	0	61	3	64
Low	0	3	162	165	0	5	160	165	0	4	161	165

Table 4 Illustrative example

Given for model (1)	True positive (TP)	False negative (FN)	True negative (TN)	False positive (FP)
High	42	0	229	1
Medium	62	3	205	2
Low	162	3	104	3

Subsequently, class labels were predicted for the remaining dataset (testing set) after training on a subset (training set) for objective evaluation. The holdout approach, randomly dividing the dataset for training and testing, was used. K-fold cross-validation, dividing the dataset into k subsets, was preferred for unbiased results. The process repeats k times, using one-fold as the test set and the remaining k-1 for training. Confusion matrices are created for each repeat, and performance indices are derived. These k indices are averaged for final cross-validation results. This technique benefits from utilizing the full dataset for training and testing in each iteration, contrasting with the holdout approach. The study employed tenfold cross-validation, a common choice for computational efficiency and error estimation accuracy.

Analysis Results and Discussion

The study evaluated the training and testing performances of KNN, ANN, and SVM classifiers through tenfold cross-validation. However, due to the dataset's moderate size (272 records), the model's predictive capacity for the testing set appeared unrepresentative. Three initial models (MOD-1, MOD-2, and MOD-3) were developed and assessed using confusion matrices to derive performance metrics. While MOD-1 and MOD-3 showed similar high classification accuracy (above 98%), MOD-2 performed less well, achieving 97%. Despite MOD-1 and MOD-3's good performance, further investigation explored whether ensemble approaches could enhance outcomes. MOD-1 was identified as the best-performing model, supported by the findings in Table 6. Notably, MOD-1's confusion matrix (Table 4, 5) indicated only six misclassified cases, resulting in a low misclassification error of 1.47%, highlighting its exceptional accuracy.

Table 6 compares SVM, KNN, and ANN classifier performance indices. SVM achieves 98.53% (89.54%) accuracy in training (testing) with a 1.47% (10.46%) misclassification error. KNN shows 97.79% (81.2%) accuracy and 2.21% (18.8%) misclassification error. ANN exhibits 98.28% (87.26%) accuracy and 1.72% (12.74%) misclassification error. SVM outperforms in both training and testing.

In both training and testing scenarios, after conducting parameter analysis for each classifier, a comparison was made regarding the impact of rework on cost. Figure 4 illustrates that SVM achieved the highest quantification of rework impact on cost across all data divisions, followed by ANN with the second highest rate. Conversely, KNN yielded the lowest quantification rate among the classifiers, indicating it as the least favorable option for quantifying the rework impact on the cost of bridge projects. With a maximum quantification rate of 97%, SVM emerged as
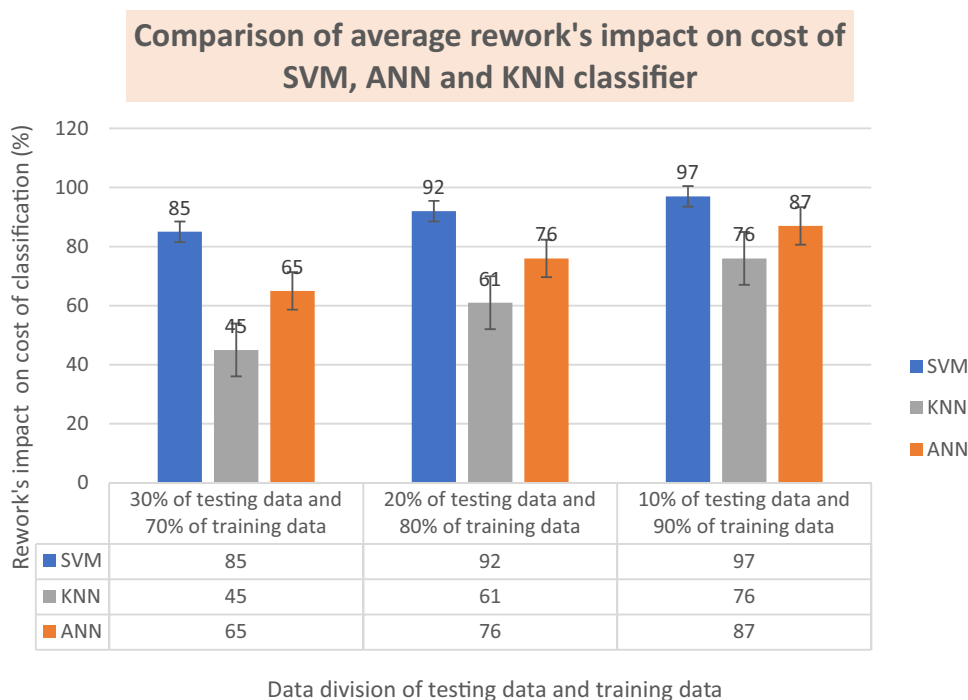
Table 5 The performance measure of the computational model

Performance measure	High	Medium	Low	Weighted average
Classification Accuracy (%)	0.9963	0.9816	0.9779	0.9853
Misclassification Error (%)	0.0037	0.0184	0.0221	0.0147
A	0.7363	0.3506	0.5175	0.5348
Cohen's Kappa Statistic	0.9861	0.9717	0.9543	0.9707
Precision	0.9767	0.9688	0.9818	0.9758
TP Rate	1.0000	0.9538	0.9818	0.9786
FP Rate	0.0043	0.0097	0.0280	0.0140

Table 6 Performance of initial ML models

Performance measure	MOD (1) SVM	MOD (2) KNN	MOD (3) ANN
Classification Accuracy (%)	0.9853	0.9779	0.9828
Misclassification Error (%)	0.0147	0.0221	0.0172
A	0.5349	0.5413	0.3018
Cohen's Kappa Statistic	0.9707	0.9595	0.9811
Precision	0.9758	0.9691	0.9763
TP Rate	0.9786	0.9662	0.9732
FP Rate	0.0140	0.0219	0.0172

Fig. 4 Comparison of average rework's impact on cost of SVM, ANN and KNN classifier

Comparison of average rework's impact on cost of SVM, ANN and KNN classifier

	30% of testing data and 70% of training data	20% of testing data and 80% of training data	10% of testing data and 90% of training data
SVM	85	92	97
KNN	45	61	76
ANN	65	76	87

Data division of testing data and training data

the best classifier. The optimal data separation was determined to be 10% for testing data and 90% for training data.

The study's goal is a platform for continuous forecasting improvement in cost performance through rework cause adjustments, achieved by training ML models for dynamic analysis, a crucial step toward the long-term objective. The key finding emphasizes the numerous benefits of using machine learning to enhance rework in construction projects:

1. *Improving forecasting and planning*: Machine learning can analyze historical data to predict potential project issues, enabling teams to plan rework more effectively.
2. *Enhancing feasibility estimates*: AI techniques can accurately analyze project costs and financial feasibility, aiding better decisions on whether to continue or rework.
3. *Reducing human errors*: ML minimizes potential human errors in estimating rework percentages, enhancing result accuracy.
4. *Improving risk management*: AI can analyze potential risks, directing attention to areas needing rework and enhancing overall project risk management.
5. *Enhancing time management*: Machine learning analyzes project schedules and forecasts rework time, contributing to improved time management.
6. *Increasing efficiency and performance*: ML helps optimize resource usage and enhance operational performance, reducing the need for rework.
7. *Predicting problems early*: ML aids in early prediction of potential issues, allowing technical teams to intervene effectively and avoid rework later.

8. *Promoting smart technology*: Machine learning integrates smart technologies like intelligent sensing and automated control, boosting construction process efficiency.

This study proposes a novel approach, combining structural equation modeling and machine learning, to forecast rework costs in bridge projects. By leveraging historical data, we aim to provide stakeholders with accurate predictions, facilitating proactive decision-making. Emphasizing the importance of understanding root causes, our research aims to inform targeted interventions, reducing rework occurrence. The application of machine learning, particularly ensemble learning, represents a significant advancement, enhancing the reliability of cost predictions across diverse datasets. Overall, our study underscores the importance of addressing rework and highlights machine learning's potential to improve construction project management.

Conclusion

In this study, machine learning (ML) and structural equation modeling (SEM) were distinct analytical methods. SEM assessed rework impact on project costs initially, validating variables affecting rework. Subsequently, ML classification forecasted fundamental rework causes more precisely. This two-stage predictive analytical methodology provides comprehensive insights, making a substantial methodological contribution. Outputs from the first phase trained and tested

ML classifiers to anticipate rework impact and compare it to overall rework values in bridge projects. The SEM model served as the foundation for ML, emphasizing the value of a dual analytical approach and SEM in assessing total rework.

To assess the overall impact of rework in Egyptian bridge construction, a structural equation model was constructed, examining rework causes based on likelihood and consequences. A theoretical framework was constructed using AMOS 26, employing a second-order model that considers rework occurrence and impact. The root mean square error of approximation for the testing sample was 5.7%, within acceptable bounds. The model highlighted construction process factors as key contributors to rework. An ML-based strategy proved most effective for managing interdependent variables. Three ML algorithms were carefully chosen based on project data characteristics, yielding accurate predictive models. SVM outperformed decision tree models in overall and class performance during tenfold cross-validation, confirming the initial hypothesis of data variables' conditional independence.

The study employed confusion matrices and various performance metrics, including classification accuracy, Cohen's kappa statistic, precision, true positive rate, and false positive rate, to evaluate the effectiveness of KNN, ANN, and SVM classifiers in predicting the impact of rework on project costs. Initial models (MOD-1, MOD-2, and MOD-3) were developed and assessed, with MOD-1 identified as the best-performing model. Further investigation explored ensemble approaches to enhance outcomes. The performance of SVM, KNN, and ANN classifiers was compared, with SVM demonstrating superior accuracy and reliability in both training and testing scenarios. SVM achieved the highest quantification rate for rework impact on project costs, followed by ANN, while KNN exhibited the lowest quantification rate. SVM achieves 98.53% (89.54%) accuracy in training (testing) with a 1.47% (10.46%) misclassification error. This underscores SVM's effectiveness in quantifying rework impact and its potential to inform decision-making in bridge construction projects. Overall, the study highlights the importance of employing advanced machine learning techniques for proactive prediction and evidence-based decision-making to mitigate rework and improve cost performance in construction projects.

However, the methods outlined in the study can be used with data sets from other projects, as opposed to the one assembled here, which was primarily examined to aid comprehension and show the viability of the suggested ML analysis technique. Consequently, it is necessary to make some suggestions for the future use of the approach covered in the study. Also, this model is not limited to implementation in Egypt because the factors are the same in all countries, but the coefficients can differ depending on the impact and the probability of factors, so it must be noted that this model is valid for application in all countries, but some changes and adjustments must be made to suit the conditions of each country.

Declarations

References

Abeku, D. M., Ogunbode, E. B., Salihu, C., Maxwell, S. S., & Kure, M. A. (2016). Projects management and the effect of rework on construction works: A case of selected projects in Abuja Metropolis, Nigeria. *International Journal of Finance and Management in Practice, 4*(1), 329–349.

Al-Janabi, A. M., Abdel-Monem, M. S., & El-Dash, K. M. (2020). Factors causing rework and their impact on projects' performance in Egypt. *Journal of Civil Engineering and Management*. https://doi.org/10.3846/jcem.2020.12916

Alla, V., Sahoo, U. K., & Behera, R. N. (2023). Seismic liquefaction analysis of MCDM weighted SPT data using support vector machine classification. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*. https://doi.org/10.1007/s40996-023-01293-6

Arbuckle, J. L. (2011). *IBM SPSS Amos 20 user's guide* (pp. 226–229). Amos Development Corporation.

Badawy, M., Hussein, A., Elseufy, S. M., & Alnaas, K. (2021). How to predict the rebar labours' production rate by using ANN model? *International Journal of Construction Management, 21*(4), 427–438. https://doi.org/10.1080/15623599.2018.1553573

Brown, G. T., Harris, L. R., O'Quin, C., & Lane, K. E. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: Identifying and understanding non-invariance. *International Journal of Research & Method in Education, 40*(1), 66–90.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504.

Chidiebere, E. E., & Ebhohimen, I. J. (2018). Impact of rework on building project and organisation performance: a view of construction professionalsin Nigeria. *International Journal of Sustainable Construction Engineering & Technology, 9*(1), 29–43.

Dahanayake, B., & Ramachandra, T. (2016). Assessment on defects occurence and rework costs in housing construction sector in Srilanka. *Context, 19*, 86.

de Oliveira Neves, F., & Salgado, E. G. (2024). From uncertainty to precision: Advancing industrial rework rate analysis with fuzzy logic. *International Journal of Business and Management, 18*(6), 119–119.

Elseufy, S. M., Hussein, A., & Badawy, M. (2022, December). A hybrid SEM-ANN model for predicting overall rework impact

on the performance of bridge construction projects. In Structures (Vol. 46, pp. 713–724). Elsevier.

Entezami, A., Shariatmadar, H., & Sarmadi, H. (2020). Condition assessment of civil structures for structural health monitoring using supervised learning classification methods. *Iranian Journal of Science and Technology, Transactions of Civil Engineering, 44*(Suppl 1), 51–66.

Fayek, A. R., Dissanayake, M., & Campero, O. (2004). Developing a standard methodology for measuring and classifying construction field rework. *Canadian Journal of Civil Engineering, 31*(6), 1077–1089. https://doi.org/10.1139/l04-068

Flapper, S. D. P., Fransoo, J. C., Broekmeulen, R. A., & Inderfurth, K. (2002). Planning and control of rework in the process industries: A review. *Production Planning & Control, 13*(1), 26–34.

Flath, C., Nicolay, D., Conte, T., van Dinther, C., & Filipova-Neumann, L. (2012). Cluster analysis of smart metering data. *Business & Information Systems Engineering, 4*(1), 31–39. https://doi.org/10.1007/s12599-011-0201-5

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science, 40*(3), 414–433.

Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. (2010). Multivariate data analysis. Upper Saddle River, NJ: Pearson.

Hwang, B. G., Thomas, S. R., Haas, C. T., & Caldas, C. H. (2009). Measuring the impact of rework on construction cost performance. *Journal of Construction Engineering and Management, 135*(3), 187–198.

Jafari, A., & Rodchua, S. (2014). Survey research on quality costs and problems in the construction environment. *Total Quality Management & Business Excellence, 25*(3–4), 222–234.

Kaveh, A. (2017). *Applications of metaheuristic optimization algorithms in civil engineering*. Springer International Publishing.

Kaveh, A., & Ghaffarian, R. (2015). Shape optimization of arch dams with frequency constraints by enhanced charged system search algorithm and neural network. *Int. J. Civ. Eng, 13*(1), 1–10.

Kaveh, A., Mottaghi, L., & Izadifard, R. A. (2022). Optimal design of a non-prismatic reinforced concrete box girder bridge with three meta-heuristic algorithms. *Scientia Iranica, 29*(3), 1154–1167. https://doi.org/10.24200/sci.2022.59322.6178

Kaveh, A., Eskandari, A., & Movasat, M. (2023, October). Buckling resistance prediction of high-strength steel columns using metaheuristic-trained artificial neural networks. In Structures (Vol. 56, p. 104853). Elsevier.

Khalesi, H., Balali, A., Valipour, A., Antucheviciene, J., Migilinskas, D., & Zigmund, V. (2020). Application of hybrid SWARA–BIM in reducing reworks of building construction projects from the perspective of time. *Sustainability, 12*(21), 8927.

Larrañaga, P., Atienza, D., Diaz-Rozo, J., Ogbechie, A., Puerto-Santana, C., & Bielza, C. (2018). Industrial applications of machine learning. CRC press.c

Love, P. E. D. (2002a). Influence of project type and procurement method on rework costs in building construction projects. *Journal of Construction Engineering and Management, 128*(1), 18–29. https://doi.org/10.1061/(ASCE)07339364(2002)128:1(18)

Love, P. E. D. (2002b). Auditing the indirect consequences of rework in construction: A case-based approach. *Managerial Auditing Journal, 17*(3), 138–146. https://doi.org/10.1108/02686900210419921

Love, P. E., & Sing, C. P. (2013). Determining the probability distribution of rework costs in construction and engineering projects. *Structure and Infrastructure Engineering, 9*(11), 1136–1148. https://doi.org/10.1080/15732479.2012.667420

Love, P. E. D., Edwards, D. J., Watson, H., & Davis, P. (2010). Rework in civil infrastructure projects: Determination of cost predictors. *Journal of Construction Engineering and Management, 136*(3), 275–282.

Love, P. E., Teo, P., Ackermann, F., Smith, J., Alexander, J., Palaneeswaran, E., & Morrison, J. (2018). Reduce rework, improve safety: an empirical inquiry into the precursors to error in construction. *Production Planning & Control, 29*(5), 353–366.

Mahamid, I. (2022). Impact of rework on material waste in building construction projects. *International Journal of Construction Management., 22*(8), 1500–1507.

Marosszeky, M. (2006). Performance Measurement and Visual Feedback for Process Improvement, A Special Invited Lecture presented in the SMILE-SMC 3rd Dissemination Workshop on 11th February 2006, Centre for Infrastructure and Construction Industry Development of The University of Hong Kong, Hong Kong.

Meshksar, S. (2012). Cost and time impacts of reworks in building a reinforced concrete structure (Doctoral dissertation, Eastern Mediterranean University (EMU)).

Mohamed, H. H., Ibrahim, A. H., & Soliman, A. A. (2021). Toward reducing construction project delivery time under limited resources. *Sustainability, 13*(19), 11035.

Ojghaz, A. S., & Heravi, G. (2023). Enhancing residential satisfaction through identifying building and location desirability criteria in Iran: A fuzzy delphi and structural equation modeling analysis. *Iranian Journal of Science and Technology, Transactions of Civil Engineering, 3*, 1–5.

Oke, A. E., & Ugoje, O. F. (2013). Assessment of rework cost of selected building projects in Nigeria. *International Journal of Quality & Reliability Management, 30*(7), 799–810. https://doi.org/10.1108/IJQRM-Jul-2011-0103

Olatunji, S. O. (2017, April). Extreme Learning machines and Support Vector Machines models for email spam detection. In 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE) (pp. 1–6). IEEE.

Oyewobi, L. O., Abiola-Falemu, O., & Ibironke, O. T. (2016). The impact of rework and organisational culture on project delivery. *Journal of Engineering Design and Technology.* https://doi.org/10.1108/JEDT-05-2013-0038

Palaneeswaran, E., Kumaraswamy M., Ng T. and Love P.E.D. (2005). Management of rework in Hong Kong construction projects. Proceedings of Queensland University of Technology Research Week International Conference, (pp. 4–5). Hong Kong.

Rofooei, F. R., Kaveh, A., & Farahani, F. M. (2011). Estimating the vulnerability of the concrete moment resisting frame structures using artificial neural networks. *Int J Optim Civil Eng, 1*(3), 433–448.

Sethi, H., Goraya, A., & Sharma, V. (2017). Artificial Intelligence based Ensemble Model for Diagnosis of Diabetes. International Journal of Advanced Research in Computer Science, 8(5).

Shoar, S., Chileshe, N., & Edwards, J. D. (2022). Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: Application of random forest regression. *Journal of Building Engineering, 50*, 104102.

Taha, G., Sherif, A., & Badawy, M. (2022). Dynamic modeling for analyzing cost overrun risks in residential projects. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part a: Civil Engineering, 8*(3), 04022041.

Wasfy, M. (2010). Severity and impact of rework, a case study of a residential commercial tower project in the Eastern Province-KSA. King Fahd University.

Wauters, M., & Vanhoucke, M. (2017). A nearest neighbour extension to project duration forecasting with artificial intelligence. *European Journal of Operational Research, 259*(3), 1097–1111.

Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & DATA, M. (2005, June). Practical machine learning tools and techniques. In Data Mining (Vol. 2, No. 4).

Yap, J. B. H., Chow, I. N., & Shavarebi, K. (2019). Criticality of construction industry problems in developing countries: Analyzing

Malaysian projects. *Journal of Management in Engineering, 35*(5), 04019020.

Yap, J. B. H., Abdul-Rahman, H., & Wang, C. (2016). A conceptual framework for managing design changes in building construction. In: The 4th International Building Control Conference 2016 (IBCC 2016), 66, 00021. EDP Sciences.

Ye, G., Jin, Z., Xia, B., & Skitmore, M. (2015). Analyzing causes for reworks in construction projects in China. *Journal of Management in Engineering, 31*(6), 04014097. https://doi.org/10.1061/(ASCE) ME.1943-5479.0000347

Yi, Z., & Luo, X. (2024). Construction cost estimation model and dynamic management control analysis based on artificial intelligence. *Iranian Journal of Science and Technology, Transactions of Civil Engineering, 48*(1), 577–588.