International Journal of Computing and Digital Systems
2025, VOL. 17, NO. 1, 1–12

http://dx.doi.org/10.12785/ijcds/1571018142

Machine Learning based Material Demand Prediction of Construction Equipment for Maintenance

Poonam Katyare[1], Shubhalaxmi S. Joshi[1] and Mrudula Kulkarni[2]

[1]*Department of Computer Science and Application, Dr.Vishwanath Karad MIT World Peace University, Pune, India*
[2]*Department of Civil Engineering, Dr.Vishwanath Karad MIT World Peace University, Pune, India*

Received 11 Apr. 2024, Revised 18 Sept. 2024, Accepted 19 Sept. 2024

Abstract: Construction managers faced Construction Equipment (CE) challenges related to running repair and replacement of spare part materials as well as shortage of materials, sudden damage of spare parts and unavailability of necessary materials at job sites frequently. Regular follow up and track of materials availability and their usage at each stage of requirement phase becomes essential. This study presents Machine Learning (ML) based material demand prediction. Training of ML models utilizes historical maintenance, and procurement periodic data related to materials of the CE. This study highlights the use of Multiple Linear Regression (MLR), Support Vector Regression (SVR), Decision Tree (DT) Regressor and ensemble boosting models as Random Forest (RF) Regressor and Gradient Bosting Regressor (GBR). According to the performance measurement of each model, RF performs better and is used for prediction. Material demand prediction helps in maintenance and operational planning of CE. Subsequently, approach assists in addressing issues early by involving operators and site owners, enabling preventive actions to be taken before the scheduled procurement process. This study addresses the corrective measurement of the model using periodic data. The model performance results indicate that early prediction of maintenance costs based on the quantity of essential materials withdrawn from demand is helpful for budgeting expenditures.

Keywords: Construction Equipment, Machine Learning, Material Demand, Maintenance

1. INTRODUCTION

Construction Equipment is a key driver for executing successful construction projects. The management of CE concerns efficiently overseeing equipment resources to meet the equipment requirements and to gain maximum returns on equipment for the construction project, which is targeted to be executed in a scheduled and economically viable fashion. Major contractors often can acquire, operate, and manage a substantial collection of heavy CE units. Therefore, making decisions regarding routine equipment management responsibilities is essential for overall project management. The daily routine involves the procurement process, maintenance process, equipment allocations, equipment operational activity, and replacement and repair activities of equipment spare parts.

The day-to-day execution of these activities has financial implications for fleet owners because cost is involved in every activity. Proper and effective budgeting of any construction project focuses on CE cost bifurcations on various aspects, which involve the initial acquisition cost of the equipment, operating cost, maintenance and repairs cost, operator and labor wages, depreciation cost, financing costs, interest payments, transportation cost, regulatory compliance costs, technology integration costs, and disposal cost of CE. A critical piece of maintenance costs is fundamentally credited to essential materials in form of spare parts for CE [6]. Site owners should maintain an inventory of spare parts associated with all equipment present and currently working on the job site. It is a big challenge to handle equipment failures and face downtime while performing tasks on-site using equipment. They need to keep records of all materials in the system with their availability quantity, order details, required quantity, withdrawn quantity, and special operating run hours of the equipment. The cost sheet for each quantity is recorded with the date and time. Large numbers of lists of materials are available that are distributed in similar groups of equipment for simplicity of cost computations.

This investigation aims to avoid manual work for computing the demand quantity of materials. The proposed study emphasizes ML-based essential materials demand prediction of CE in advance from a maintenance perspective. This study focuses on the rational study of various ML algorithms.

This paper is arranged as follows. The existing study

E-mail address: poonam.katyare@gmail.com

with limitations is elaborated in part 2. The proposed methodology for predicting the demand quantity of materials with data preprocessing and model fitting is given in part 3. Results and discussion with a comparison of ML algorithm performance are indicated in part 4. The inference with the concluded work is illustrated in part 5.

2. EXISTING STUDY

A significant study was identified related to CE cost prediction, residual value prediction, and sensor-based data analysis. Maintenance of CE study observed using different methods, such as by reviewing the existing techniques used for reliability and fault analysis of CE. ML techniques, graphical methods, fault tree analysis, and probability distribution models have been used; however, ML models have the best accuracy for failure prediction and reliability estimations of CE [3], [38]. The researcher presented a related study on the implications of the Internet of Things with sensor-based technologies attached to the equipment for capturing real-time information of CE with location tracking, movement tracking, working condition of engines, fuel data updates, distance travelled, and battery updates from equipment working on construction job sites. This would help managers analyse the data collected from remote sensors and make proper decisions regarding the equipment's performance. Remote sensing devices identify information related to construction material tracking to handle the supply chain management process along with cloud computing, radio frequency identification, augmented reality, and big data technologies [1],[5],[7].

The existing study presented the prediction of residual values of CE by an Autoregressive Tree algorithm of data mining using equipment age, make, model, region, horsepower, auction year, condition rating, annual construction investment, and Gross Domestic Product features to predict equipment price. This study compared the performance of data mining algorithms with those of neural networks, linear regression algorithms, and deep learning algorithms [11]. The authors demonstrated fuel consumption prediction using ML and live data parameters from smart sensing devices, which indirectly impacted the maintenance cost of CE. Real-time data ensemble methods provide better accuracy than other regression models [2],[4],[35], [36], [37].

Another study demonstrated a prototype model that effectively reduces labor costs and mitigates challenges associated with equipment maintenance decision-making by presenting a data-driven methodology that integrates three key skills reliability maintenance focused on reliability, modeling of building data, and live tracking system. This includes critical components for ensuring the optimal functionality of buildings [8],[9]. Data-driven approaches have been observed in many studies that attempt to manage equipment information data using data analysis from huge amounts of data and manage the data for decision making [10].

Manufacturers in the construction machinery parts industry must manage inventories promptly, optimize production processes for efficient and swift product manufacturing, and promptly deliver finished products to customers. To solve this problem, an existing study elaborated demand estimating for spare parts in the construction machinery industry using regression and artificial intelligence models [12]. Similarly, the heavy equipment of specific group demand forecasting is also performed by the researcher using the Support Vector Machine Regressor, which is very useful for the equipment owner [13]. Multivariant time series analysis performs better for the constriction raw materials of steel products prediction [14]. A related study investigated the prediction of heavy equipment prices with precision by employing ML algorithms on sales data obtained from a website [15].

Another study uses an artificial neural network-based methodology to measure uncertainties and generate forecasting intervals for predicting prices of construction materials, with a specific focus on asphalt and steel. This study provides supplementary information to enhance the effective management of project cost-related risks through estimate intervals to project managers. The proposed optimal Lower Upper Bound Estimation (LUBE) cost function yields highly precise estimate intervals [16]. The Analytical Hierarchy Process within the thematic domain involves the development of a modified decision model in CE procurement. This model is designed to order parameters that influence equipment procurement. The approach is particularly tailored to address the unstructured aspects of the selection method [17].

Research showcased the prediction of maintenance costs related to breakdown and planned maintenance activity events for essential plant resources, and the developed model exhibited strong predictive accuracy. The methodology integrates a stochastic mathematical modeling approach that considers both unplanned breakdowns and scheduled maintenance. This technique generates a pseudo-random number to simulate the magnitude of an impending maintenance cost event [18]. Time series maintenance and fuel consumption data were used to anticipate the CE cost of maintenance using a neural network time series model [19], [21]. The existing study employs time series and multiple regression models to predict construction material prices. The combination of these statistical methods allows for capturing both time-series trends and relationships between different economic factors, providing a robust prediction framework. The integration of multiple models enhances predictive accuracy and robustness, catering to complex market dynamics [29]. This review discusses various artificial intelligence methods for demand forecasting in supply chain management, including machine learning algorithms like neural networks, support vector machines, and ensemble methods. AI methods can handle large datasets and complex patterns, improving forecast accuracy over traditional statistical methods [30].

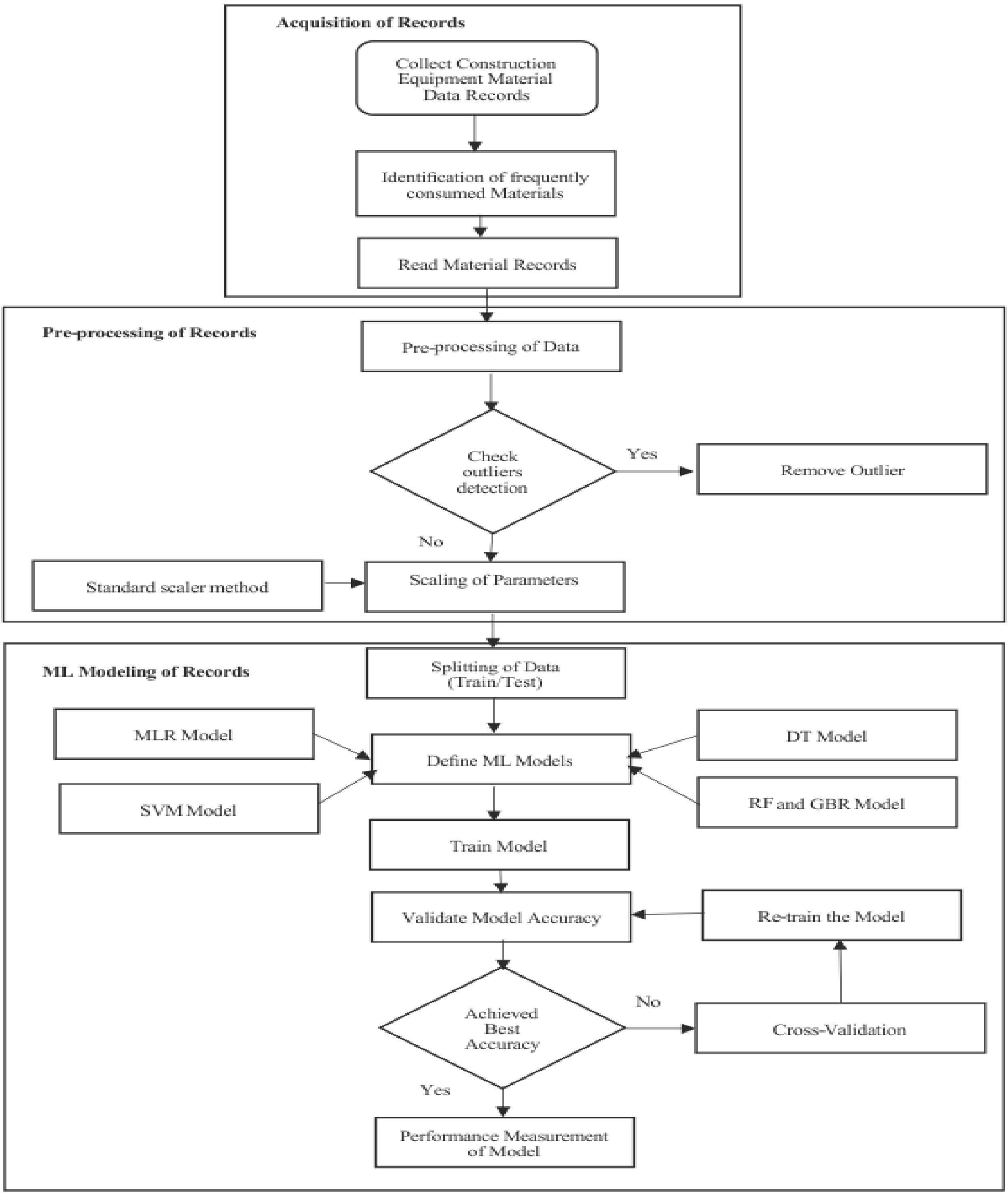

Figure 1. Prediction Model Flow Diagram

This systematic review examines the adoption of machine learning technology for failure prediction in industrial maintenance, emphasizing the use of algorithms such as decision trees, random forests, and neural networks. Machine learning models can analyze historical data to predict equipment failures, thus optimizing maintenance schedules and reducing downtime [33]. This study investigates the impact of increasing raw material prices on construction costs, providing insights into economic factors affecting the construction sector in specific regions. Understanding the economic impact helps in better budgeting and cost management strategies in construction projects [32].

Another study provides a comprehensive review of artificial intelligence and ML techniques used for performance monitoring and failure prediction in industrial equipment. The study highlights the increasing importance of predictive maintenance to improve operational efficiency and reduce downtime in industrial settings. The study discusses various AI and ML algorithms such as neural networks, support vector machines, decision trees, and ensemble methods. It examines their applications in identifying patterns and anomalies in equipment performance data to predict potential failures [31].

A. Limitations of Existing Study

Existing studies have demonstrated the prediction of residual values of equipment, prediction related to the cost of equipment, prediction of failure or breakdown of equipment, fuel consumption, and maintenance cost estimation of equipment. This study presents qualitative and quantitative data analysis using ML models, time series analysis, and factor analysis. Contractors or site owners need to maintain records of equipment spare parts or materials available, required or demand quantity, and how much quantity is utilized manually. They face the issues of failure of materials, replacement of materials, and damage of materials at the site. They should add all these records manually to log sheets, order the materials as per the requirement and wait for procurement of those materials. Meanwhile, there could be downtime at the site for that equipment because of the unavailability of the materials required for that equipment aligned with the working conditions. The frequency of such situations or challenges frequently occurred at the job site. There is a need to estimate such a material demand quantity. Many predictive models require high-quality, granular data for accurate forecasting. Inconsistent or incomplete data can lead to less reliable predictions. The existing system has limitations of Data Quality and Availability. The proposed study predicts the quantity of equipment required for material demand from the operating hours and history data, which would help to maintain the required materials stock in advance at the job site. Machine learning (ML) models are increasingly relevant and effective in addressing a variety of challenges in the construction equipment industry. These challenges range from predictive maintenance and equipment optimization to safety and operational efficiency. Predictive Maintenance is major challenge of equipment downtime due to unexpected failures that can be costly and disruptive. Predictive maintenance uses ML algorithms to analyze data from sensors and historical maintenance records to predict when equipment is likely to fail. This allows for timely maintenance, reducing downtime and repair costs. Techniques such as anomaly detection and time-series analysis are particularly useful here. Managing the supply chain and inventory effectively to avoid delays and excess costs is another challenge. ML can forecast demand for materials and equipment, optimizing inventory levels and supply chain operations. Forecasting models and optimization algorithms are particularly useful in this area. The integration of ML models in the construction equipment industry addresses numerous challenges by improving predictive maintenance, optimizing equipment utilization, enhancing safety, reducing fuel consumption and emissions, managing fleets efficiently, ensuring quality, and optimizing supply chain and inventory management. As data availability and computational power continue to grow, the relevance and effectiveness of ML in this sector are expected to expand, driving further innovation and efficiency.

3. PROPOSED METHODOLOGY

Analyzing time series data for construction equipment material prediction involves a systematic approach to ensure accurate and reliable forecasting. Analyzing time series data for construction equipment material prediction involves collecting and preprocessing data, performing exploratory analysis, selecting and training appropriate models, evaluating their performance, and deploying and monitoring the models in a production environment. This structured approach ensures accurate material forecasts, leading to optimized resource management and reduced operational costs in the construction industry. This emphasizes analysis of data related to CE materials, detailing the quantity available for each machine, used run hours, the specifics of each machinery order, and the anticipated future demand for each. The goal is to predict the demand for construction machinery. Figure 1. illustrates the Prediction model flow diagram with the detailed steps for estimating the quantity of materials. This flow diagram interprets the Data Acquisition, Data Preprocessing and Data Modelling steps.

A. Acquisition of Records

Daily logs of the repair and replacement of materials are maintained at the construction site. New orders are placed to purchase replacement materials. Industries keep these records in their Enterprise Resource Planning system. The proposed study acquires order and material quantity data from the organizational system from 2017 to 2023 from various sources. Interviews were conducted with experts and contractors working on job sites, and literature review data were used to finalize the features required for the proposed study. The acquired dataset is in a daily basis format considering the days when orders are placed. The collected data has features related to order, material, and material quantity details. Order details include order number, creation date,

TABLE I. Material Groups clusters

No.	Material Groups
1	Spares
2	Structural Steel
3	Welding Materials
4	Pipe and Pipe Fittings
5	Hardware, Painting and Chemicals
6	Electrical Items
7	Rubber Goods
8	Lubricant and Oil
9	Tools
10	Miscellaneous
11	Consumables (Anchor/Pilling/Drilling)

completion date, and requirement date. Material details features represent Site number, Material number, Equipment number, Material group, Operating run hours, and equipment manufacturer. The major material groups are classified into various categories. Material quantity-related features denote the available quantity of material in the inventory, the required quantity at the time of replacement and repair, and the withdrawn quantity of materials representing the total quantity of materials used. Frequently consumed materials were observed during the study. Major material categorial groups of materials are highlighted in the dataset. Material group codes are present in the dataset, and mapping of all materials under groups will be used in future studies. Table I presents the material groups used in the dataset.

B. *Preprocessing of Records*

The demand quantity of material estimation is related to the quantity of material withdrawn from historical records. The available and required quantities are major contributors to predicting the withdrawn quantity demanded. Table II Statistical Description of parameters represents statistical values for major parameters. The attributes relation is identified from the correlation matrix denoted in Table III Correlation matrix of parameters where P1 is withdrawn quantity, P2 is site, P3 is material number, P4 is equipment number, P5 is material group, P6 is run hours, P7 is quantity available and P8 is required quantity of material. Correlation Matrix tests can be used to check whether the information focuses are independent and indistinguishably distributed. It observes the relationship of the independent parameters with the target parameter [4],[13]. In this study, the material was highly correlated with the quantity available, quantity required, and operating run hours. The withdrawn quantity is also related to the available and required material quantity. Finally, the major features selected for the predictive modeling of records are material number, Equipment number, Material group, Operating run hours, quantity available, and required quantity to predict the withdrawn quantity demand of material. Outliers are identified and removed from the data using the quantile method of outlier removal [18].

1) *Outlier Removal Method*

A statistical approach, Interquartile Range (IQR) is used to remove the outliers [26]. This approach identifies the distribution of the mid-fifty percentage of the records. Equation 1 represents the formula for computing IQR as the subtraction of the 75th record percentile as QT3 with the 25th record percentile as QT1.

$$IQR = QT3 - QT1 \tag{1}$$

Where,
QT1=Upper bound with value less than 25% of records lie.
QT3=Upper bound with value less than 75% of records lie.
This approach handles the skewed record distribution with outliers and provides a list of outliers.

2) *Feature Scaling*

It is an approach of transforming values of features from records into similar scales, which helps to define the equal contribution of all features. Scaled features have a greater impact on performing ML models accurately.

Standardization is an approach that denotes that the values of features are central to the mean with a unit of standard deviation [24]. This supports the retention of the relationship between record points from the data mentioned in equation 2. It is computed as

$$(DT{-}mean(allDTs))/SD \tag{2}$$

Where,
DT = Data Point
DTs = All Data Points
SD = Standard deviation of all DTs

C. *Preliminary Analysis of Records*

Preliminary analysis of the pre-processed data helps to observe year-wise material usage. This dataset is real-time data of maintenance which involves materials details that were repaired and replaced at the time of maintenance. The issues are related to running repairs, breakdown orders, breakdown repairs, calibration changes, maintenance after specific run hours, order of the machinery, defective material indication, regular servicing, and handling of damaged materials. Every record of the issue along with order details of materials were kept as logs in the ERP system of the organization. This dataset contains Site number, Equipment number, Material number, Material Group, Run hours, Available quantity of materials in stock, Requirement quantity and withdrawal quantity is the quantity used as the major features along with order number, order creation date, and order completion date as the minor features. Figure 2 Year-wise material usage from 2018 to 2023. As per the increment in project scheduling, the increase in the order of material usage is listed. The years 2022 and 2023 highlighted more use of materials than prior years. Key

TABLE II. Statistical description of parameters

	Withdrawn Qty	Material Group	Run Hrs	Quantity Available	Requirement Qty
Mean	1.88	209	6552.90	2.33	2.13
Std	2.14	45.55	4246.44	3.91	3.36
Min	0	200	2	0.004	0
25%	1	200	3350	1	1
50%	1	200	5639	1	1
75%	2	200	8983	2	2
Max	39	500	21022	286	91

TABLE III. Correlation matrix of parameters

	P1	P2	P3	P4	P5	P6	P7	P8
P1	1	0.03	0.32	-0.06	0.32	0	0.31	0.96
P2	0.03	1	0.11	0.19	0.11	0.3	0.05	0.03
P3	0.32	0.11	1	-0.06	1	-0.04	0.39	0.32
P4	-0.06	0.19	-0.06	1	-0.06	-0.21	-0.04	-0.05
P5	0.32	0.11	1	-0.06	1	-0.04	0.39	0.32
P6	0	0.3	-0.04	-0.21	-0.04	1	-0.01	0
P7	0.31	0.05	0.39	-0.04	0.39	-0.01	1	0.32
P8	0.96	0.03	0.32	-0.05	0.32	0	0.32	1

insights from Figure 2 are related to Trends and Patterns with Significant Increase, Fluctuations in Earlier Years, and High Usage. Significant Increase in 2022, with material usage more than doubling compared to 2021.This sharp rise indicates a significant surge in construction activity, likely due to an increase in project scheduling or the initiation of several large-scale projects. Fluctuations in Earlier Years happened Between 2018 and 2021, material usage shows notable fluctuations in a significant increase from 2018 to 2019 and a decline in 2020 and a further drop in 2021. High Usage in 2022 and 2023 with despite a slight decrease in 2023, material usage remains high compared to previous years, indicating a sustained period of increased construction activity.

D. Modelling of Records

ML is a strategy for changing information into noteworthy information. Different directed ML methods are accessible for expectation, which is related to the verifiable information for anticipating new occasions of data of interest with the connection of target factors alongside independent information values. ML model follows information assortment, Information preprocessing, and modeling with different algorithms. Finally, the model with the better measurement is chosen for predicting new occasions of information. We used different ML regression models, for example, MLR, SVR, DT along with ensemble regressors as RF and GBR models for determining the demand quantity of materials.

1) Multiple Linear Regressor (MLR)

MLR is a basic and generally involved method for displaying the association of a dependent variable with at least one independent variable. The model anticipates a linear relationship between the dependent and independent factors, suggesting that they can be represented as a straight line. MLR is a measurable investigation technique used to determine the dependent quantitative connection between at least two factors. Target factors in the regression examination are perceived or assessed [24]. Independent factors are the factors that are remembered to significantly affect the target variable attempted for assessment. Forecasts can be made by estimating the connections within factors using examination. Considering input parameters as X the basic statistical model of MLR is stated by Equation 3.

$$Y(x, c) = c_0 + c_1 x_1 + - - + c_n x_k = c_0 + \sum_{i=1}^{K} c_i x_i \tag{3}$$

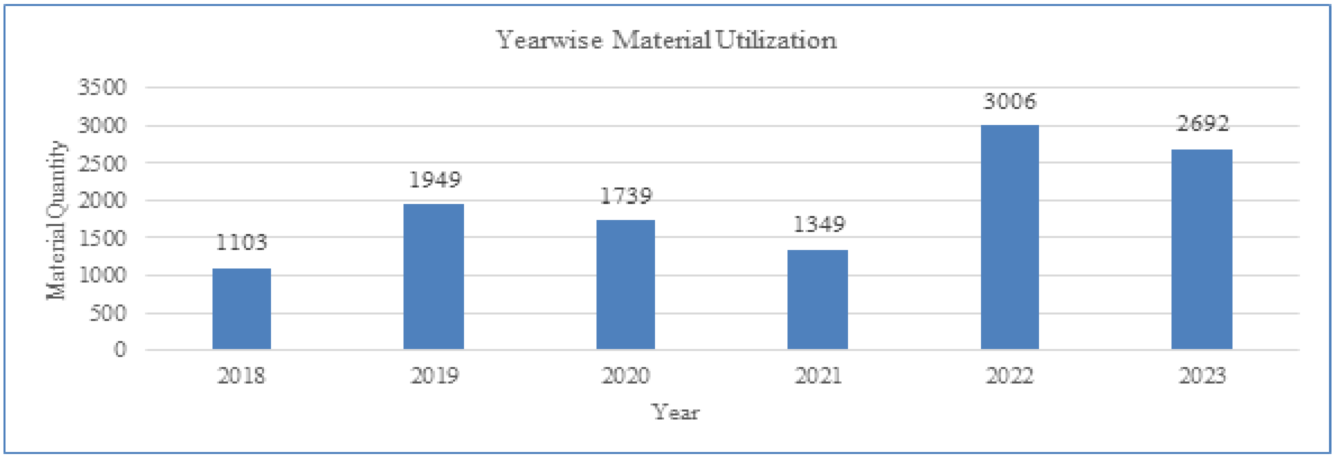

Figure 2. Year wise Materials Utilization

where c can be projected using the least squares method as in Equation 4

$$\hat{c} = argmin\{\sum_{i=1}^{N}(y_j - c_0 - \sum_{i=1}^{K} c_i x_{ji})^2\} \quad (4)$$

where x_1, x_2,. . x_k are the observed values of independent parameters, c_1, c_2,. . $c_k$are the regression coefficients, c_0 is the intercept term, N is the sample count size with K representing input parameters, and y are stated value of the dependent parameter.

2) Support Vector Regressor (SVR))

A function provided by the SVR signifies the relationship existing within the dependent and independent parameters with a reducing error factor. The fundamental aim of SVR is to discover a hyperplane with the largest number of points within the decision boundary line or support vectors that should be present within that boundary line. Decision boundaries are used with hyperplanes to anticipate continuous values. This assortment of numerical activities known as kernels is used to change input information into important configurations. SVR attempts to fit between the boundary lines and the hyperplane [4], [24]. The formula for a SVR can be expressed as follows:

$$\hat{y} = \sum_{i=1}^{nsv} a_i K(x_i, x) + b \quad (5)$$

Where,
ŷ - the predicted dependent value.
nsv- the count of support vectors.
x_i- the ith support vector.
b– bias term
K(x_i,x)– function kernel, which calculates similarity between the *i*-th support vector and the input sample x allowing for nonlinear relationships between features. a_i are coefficients associated with the support vectors. A hyperplane is calculated to fit the training data while minimizing margins. This aims to find coefficients a_i and the bias term b that minimize the empirical risk as the variation in the anticipated and real values subject to a margin of tolerance ϵ. This optimization problem is typically described as a quadratic programming problem and is solved using optimization techniques. Common kernel functions include sigmoid, linear, and polynomial

kernels. The choice of the kernel function varies depending on the complexity of the relationships between features and the nature of the data. SVR is intensely efficient for datasets with dense relationships and high-dimensional feature spaces. SVR ensures robust predictions and reduced sensitivity to outliers by expanding the margin among the hyperplane and the data points.

3) Decision Tree Regressor (DT))

It is an extensively applied supervised learning algorithm. It supports regression and classification analysis. A DT is a progressive model used in portraying decisions and their expected results, consolidating chance occasions, asset costs, and utility. This algorithmic model uses contingent control proclamations in the form of statements. It is a nonparametric supervised learning method helpful for both regression and classification analysis. The tree structure contains a root node and subtrees with branches followed by interior nodes, and leaf nodes frame a hierarchical, tree-like construction [25]. The DT regression model can be represented by the following formula:

$$\hat{y} = \sum_{i=1}^{N} w_i \cdot I(x \in \mathbb{R}_i) \quad (6)$$

Where,
ŷ - forecasted target value.
N =total count of leaf nodes in the DT.
R_i = region as leaf node of the feature space stated as the ith leaf node.
wi is the anticipated value correlated with the leaf node.
Input value when lies in the region, Indicator function proceeds to success. The anticipated value of the leaf node is treated as the final prediction. Each region with an

associated leaf node with an anticipated value represents the average of the dependent estimates of the training falling within that region [24]. The DT regressor formula essentially represents a piecewise constant function, where feature space is partitioned into non-overlapping regions and each section is linked with a constant predicted value. The last estimate for a given input sample is the sum of the predicted values of the leaf nodes into which the sample falls [25].

4) Random Forest Regressor (RF)

Ensemble learning models impact the finding of solutions to very complex regression problems. Ensemble learning can be characterized as the method involved in creating different models, such as classifiers, and then accumulating their outcomes to acquire better prescient execution. Two notable outfit-learning techniques are boosting and bagging. In supporting, progressive models add additional load to preparing cases that were erroneously predicted by past models. While making the forecast, a weighted vote is considered. Although progressive models are not reliant upon prior models in bagging, each model is freely developed by a bootstrap test of information. Forecasting is created by considering a basic larger part vote. The ensemble predictive model, RF, is built on a set of decision/regression trees. Rather than basing the forecast on a single tree, a group of trees is used to make the determination. RF adds an extra element of randomization to bagging, which sets it apart from other approaches. Like other bagging models, RF uses a bootstrap of sample data to build each decision/regression tree. The process for creating trees is different [8]. Because of this technique, the RF can withstand overfitting and excel in various problems. In addition, working with the assessment of variable significance and exception recognition are different advantages of this calculation [24]. Moreover, RF is sensibly quick to obtain and can be effortlessly parallelized. By backward eliminating predictors according to the specified variable relevance, RF can be improved. The formula for a RF can be stated as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x) \tag{7}$$

Where,
ŷ - anticipated target value.
N is the entire trees in the RF.
$T_i(x)$ is the prediction of the i-th decision tree for the sample x. This model aggregates estimates of all multiple DTs to make a final prediction. Every DT is trained using a bootstrap sample extracted from the training data and allows the splitting of features randomly. At last estimate is computed by averages of all individuals' predictions. Overfitting reduces using averaging and enhances performance. The final prediction depends on the contribution of every tree, regardless of its individual performance. This ensemble approach makes RFs robust and capable of handling noisy data while providing reliable predictions.

5) Gradient Boosting Regressor (GBR))

It has a place in the class of ensemble learning techniques that explicitly boost calculations. It is known for its high prescient accuracy. It functions admirably with both linear and nonlinear connections between the dependent and target factors. It can deal with complex information with high dimensionality and countless factors. It can detect complex communications among factors and precisely model non-direct connections. It can deal with missing qualities in the dataset without requiring attribution. It divides information by considering accessible elements and continues to prepare the model. It is hearty handles to anomalies in the information. This method uses a collection of weak learners and limits the effect of exceptions to the iterative process. The regressor includes significance scores, allowing comprehension of the elements that are most compelling in forecasting. This can be useful for highlighting determination and identifying hidden examples in the information. The regressor is less inclined to overfit in contrast with other complex models like profound brain organizations. This is because it assembles trees successively, improving the blunders made by the past trees. The regressor considers tweaking hyperparameters like the number of trees, tree profundity, learning rate, and misfortune capability, giving adaptability in model streamlining [39]. It tends to be used for an extensive variety of regression undertakings, including the expectation of nonstop factors. GBR is a flexible and strong model reasonable regression undertaking, particularly when high prescient precision and interpretability are required [22][23]. The formula for a GBR model stated as:

$$\hat{y} = \sum_{i=1}^{M} \gamma_i h_i(x) \tag{8}$$

where:
$\hat{y}$ is the predicted target value.
M is the total count of trees.
$h_i(x)$ is the estimate of the i-th base learner for the input sample x.
γ_i is the learning rate associated with the i-th base learner.

In GBR, the model successively develops an ensemble of weak learners, typically DTs, and joins them with strong learners. Each subsequent base learner focuses on residuals as the variation involving the actual with predicted estimates of the preceding predictions. By iteratively fitting new base learners, GBR gradually improves the model's ability to trap complicated data associations. The key idea behind gradient-boosting regression is to minimize a loss function as squared error loss. Each base learner training helps to reduce the loss concerning residuals of previous predictions. GBR is a compelling method for building predictive models for handling complex datasets. However, it is important to adjust trees to boost iterations and the learning rate to prevent overfitting and achieve optimal performance.

6) Cross Validation (CV) Technique

K-fold CV is a strategy utilized to assess performance by dividing the first dataset into k-equivalent estimated subsamples, called folds. The cycle includes iteratively preparing the model k times. This permits us to obtain k arrangements of assessment measurements, ordinarily finding the middle value to obtain a more reliable prediction. This cross-validation technique of data splitting at the training-validation split can mitigate the overfitting issues and retain a consistent estimate of the model execution. A more robust estimate of the model performance is provided by this technique. It uses multiple training validation splits and averages the performance. This is specifically used to select the most suitable model performance and perform a comparative evaluation of the model measurement [20].

4. Results And Discussion

The study depicts the expectation task completed to examine a bunch of elective models for predicting the material demand quantity of the chose dataset. We assessed the suitability of MLR, SVR, GBR, DT, and RF models for predicting the material demand quantity of CE. Using the real dataset, regression models were trained and evaluated. Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R2 score) are used to assess performance measurement of the regression models presented in equations (3), (4), (5), and (6), respectively. The mean squared variance of actual with projected values assigned as MSE, the mean-variance within the original and estimated values denotes MAE, and the square root of the MSE associated with error rate along with the coefficient of estimated values about the original values is indicated by the R2 score. The percentages represent values between 0 and 1.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(PRED_i - ACT_i)^2 \tag{9}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|PRED_i - ACT_i| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(PRED_i - ACT_i)^2} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(PRED_i - ACT_i)^2}{\sum_{i=1}^{N}(ACT_i - \bar{ACT})^2} \tag{12}$$

where $PRED_i$ and ACT_i denote the i-th predicted and actual material demand quantity values. The comparative examination of the models' performances is shown in Table IV. Performance measurement of models. The MLR, SVR, GBR, DT regression and RF regression models with k-fold cross-validation are measured with MAE, MSE, RMSE, and R^2 score values and compared to predict the withdrawn material quantity demand of the CE.

TABLE IV. Performance measurement of models

Model	MAE	MSE	RMSE	R2
MLR	0.82	3.12	1.74	0.32
SVR	0.28	1.35	1.13	0.52
GBR	0.21	0.37	0.59	0.62
DT	0.06	0.24	0.46	0.65
RF	0.08	0.22	0.42	0.66

MLR can be utilized for material quantity assessment when there is a reasonable direct connection within the input factors as equipment number, material available quantity along with run hours and the target variable as the withdrawn quantity of material. In MLR, the weighted amount of the variables' coefficients is used to predict material quantity. MLR might give a decent beginning stage, yet its capacity to discover complex connections between different highlights might be restricted when predicting the withdrawn quantity of material demand. More complex models may be expected to represent nonlinear impacts. Figure 3 Performance Measurement of models represents a visualization of models with R2 score. SVR with RBF kernel meets a useful ability for anticipating material quantities by really discovering complex relationships between input parameters and the target parameter. Appropriate information preprocessing, model preparation, hyperparameter tuning, and assessment are critical stages in utilizing this methodology for precise expectations in material quantity prediction assignments. Decision trees can deal with different categories of data. The CE materials dataset represents numerical and categorical parameters implications. DT is a very simple method of decision-making at each stage of splitting nodes. DT is inclined to overfitting, particularly when the tree develops intensely in the training data. This can prompt unfortunate speculation execution on inconspicuous information. Ensemble methods such as RF and GBR help for resolving data overfitting and provide better results for forecasting material withdrawn quantities and demand. RF enhances decision trees by joining various trees' forecasts. It acquires complex associations between parameters and material quantities. It can handle multi feature data. Key insights from Comparative Analysis from Table IV

A. Predictive Accuracy representing R^2 (Coefficient of Determination):

Random Forest (RF) and Decision Tree (DT) models exhibit the highest R^2 values (0.66 and 0.65, respectively), indicating they explain most of the variance in the data and provide the most accurate predictions. Gradient Boosting Regressor (GBR) performs well with an R^2 of 0.62.Support Vector Regressor (SVR) shows moderate accuracy with an

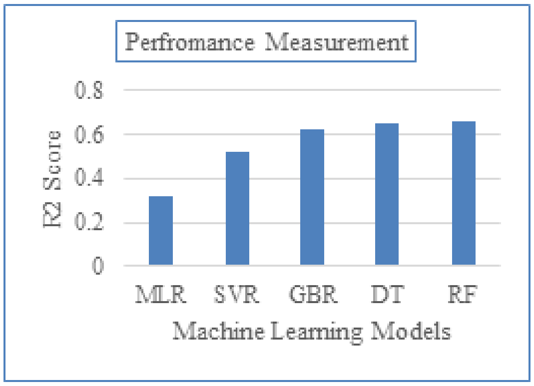

Figure 3. Performance Measurement

R^2 of 0.52. Multiple Linear Regression (MLR) has the lowest R^2 (0.32), suggesting it is less effective at capturing the underlying patterns in the data.

B. Robustness with Consistency and Outlier Sensitivity:

RF and DT models tend to be more robust to outliers and variations in the data due to their ensemble and hierarchical nature. GBR, as an ensemble method, also demonstrates robustness. SVR can be sensitive to the choice of hyperparameters and may not perform as robustly across varying datasets.MLR is the least robust, often influenced by outliers and assumptions about linearity.

C. Computational Efficiency with Training and Prediction Time:

MLR generally the fastest to train and predict due to its simplicity and linear nature. SVR computationally more intensive, especially with larger datasets, due to the kernel trick. GBR has moderate computational efficiency, balancing between accuracy and training time. DT is efficient in training and prediction but can suffer from overfitting if not pruned. RF is Computationally intensive due to training multiple trees, but parallel processing can mitigate this to some extent.

For predictive tasks in material usage forecasting, Random Forest (RF) and Decision Tree (DT) models are the most effective in terms of accuracy and robustness. Gradient Boosting Regressor (GBR) serves as an excellent alternative with a balance of high accuracy and moderate computational demands. Support Vector Regressor (SVR) can be considered for moderate performance needs, while Multiple Linear Regression (MLR) is less suitable for capturing the complexity in this context.

Using machine learning (ML) algorithms for material demand prediction in construction settings can significantly improve efficiency, cost-effectiveness, and project management. Improved Accuracy in Demand Forecasting, Optimized Inventory Management, Enhanced Project Scheduling, and Cost Savings are the practical implications. A large construction company used ML models to predict the demand for materials. By analyzing historical data, weather patterns, and project timelines, the ML model reduced material shortages and overages by 20%, leading to cost savings and smoother project execution. A construction firm with multiple ongoing projects may use the prediction of the exact quantities of materials required at different stages of each project. This enables just-in-time delivery, reducing storage costs and minimizing the risk of material degradation or theft. The accurate predictions allowed for better scheduling of deliveries, avoiding delays caused by material shortages. Furthermore, these accurate forecasts allow companies to optimize inventory management by predicting the exact quantities of materials needed at various stages of a project. This enables a just-in-time delivery approach, reducing storage costs and minimizing risks such as material degradation, theft, or obsolescence. Additionally, by knowing when materials will be required, construction firms can better schedule deliveries, ensuring all resources are available when needed, thereby minimizing delays and enhancing coordination between teams. Ultimately, the integration of ML in material demand prediction not only streamlines operations but also contributes to substantial cost savings and smoother project execution.

5. Conclusions and Future Work

This study focuses on the machine-learning-based material-demand prediction of CE. Maintenance data records were analyzed in this study. The limitations of the existing study were acknowledged and summarized using ML technologies, and a model to predict the material demand quantity was proposed. This study helps to estimate the material demand in advance for maintenance and to maintain the maintenance cost associated with the estimated materials in planning. This study provides various ML-based regression models, such as MLR, SVR, GBR, DT, and RF regression model performance. The results reveal the viability of utilizing ML techniques to overcome the difficulties in predicting material quantities. The RF model predicts material quantities accurately and performs better than other regression models. It is critical to handle real-time data for preprocessing, which involves outlier removal, handling missing values, and scaling the features to acquire accurate data for modeling. ML models are very sensitive to the quality of the dataset. This study demonstrates ML applications for material quantity prediction of CE in the construction industry for maintenance. The estimation of maintenance and operating costs of materials for CE leads to the financial budgeting of the overall construction project at the job site. This study presented the work for limited construction materials data. The Study can be expanded using large materials with similar behavior. There is a challenge to handle the real time data with large volume. Future research would help in providing material prediction for various categories of construction equipment with large volume of data.

References

[1] A. Kumar and O. Shoghli, “A review of iot applications in supply chain optimization of construction materials,” in *ISARC 2018 - 35th*

International Symposium on Automation and Robotics in Construction International AEC/FM Hackathon Future Building Things, July 2018.

[2] P. Katyare, S. Joshi, and M. Kulkarni, "Utilizing machine learning approach to forecast fuel consumption of backhoe loader equipment," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 5, pp. 1194–1201, 2024.

[3] P. Odeyar, D. B. Apel, R. Hall, B. Zon, and K. Skrzypkowski, "A review of reliability and fault analysis methods for heavy equipment and their components used in mining," *Energies*, vol. 15, no. 17, pp. 1–27, 2022.

[4] P. Katyare, S. S. Joshi, and S. Rajapurkar, "Real time data modeling for forecasting fuel consumption of construction equipment using integral approach of iot and ml techniques," *Journal of Information and Optimization Sciences*, vol. 44, no. 3, pp. 427–437, 2023.

[5] P. Katyare and S. S. Joshi, "Construction industry digitization using internet of things technology," in *Proceeding of International Conference on Computational Science and Applications. Algorithms for Intelligent Systems*. Springer, Singapore, 2022, pp. 243–249.

[6] H. Fan, H. Kim, and O. R. Zaïane, "Data warehousing for construction equipment management," *Canadian Journal of Civil Engineering*, vol. 33, no. 12, pp. 1480–1489, 2006.

[7] P. Katyare and S. Joshi, "Construction productivity analysis in construction industry: An indian perspective," in *Proceeding of International Conference on Computational Science and Applications. Algorithms for Intelligent Systems*. Springer, Singapore, 2022.

[8] Z. Ma, Y. Ren, X. Xiang, and Z. Turk, "Data-driven decision-making for equipment maintenance," *Automation in Construction*, vol. 112, p. 103103, 2020.

[9] J. C. P. Cheng, W. Chen, K. Chen, and Q. Wang, "Data-driven predictive maintenance planning framework for mep components based on bim and iot using machine learning algorithms," *Automation in Construction*, vol. 112, p. 103087, 2020.

[10] H. Fan, H. Kim, S. AbouRizk, and S. H. Han, "Decision support in construction equipment management using a nonparametric outlier mining algorithm," *Expert Systems with Applications*, vol. 34, no. 3, pp. 1974–1982, 2008.

[11] O. Alshboul, A. Shehadeh, M. Al-Kasasbeh, R. E. Al Mamlook, N. Halalsheh, and M. Alkasasbeh, "Deep and machine learning approaches for forecasting the residual value of heavy construction equipment: a management decision support model," *Engineering, Construction and Architectural Management*, vol. 29, no. 10, pp. 4153–4176, 2022.

[12] A. Aktepe, E. Yanık, and S. Ersöz, "Demand forecasting application with regression and artificial intelligence methods in a construction machinery company," *Journal of Intelligent Manufacturing*, vol. 32, no. 6, pp. 1587–1604, 2021.

[13] A. Kargul, A. Glaese, S. Kessler, and W. A. Günthner, "Heavy equipment demand prediction with support vector machine regression towards a strategic equipment management," *International Journal of Structural and Civil Engineering Research*, pp. 137–143, 2017.

[14] C. Lee, J. Won, and E.-B. Lee, "Method for predicting raw material prices for product production over long periods," *Journal of Construction Engineering and Management*, vol. 145, no. 1, pp. 1–8, 2019.

[15] N. Boyko and O. Lukash, "Methodology for estimating the cost of construction equipment based on the analysis of important characteristics using machine learning methods," *Journal of Engineering (United Kingdom)*, 2023.

[16] M. Mir, H. M. D. Kabir, F. Nasirzadeh, and A. Khosravi, "Neural network-based interval forecasting of construction material prices," *Journal of Building Engineering*, vol. 39, p. 102288, 2021.

[17] K. Petroutsatou, I. Ladopoulos, and D. Nalmpantis, "Hierarchizing the criteria of construction equipment procurement decision using the ahp method," *IEEE Transactions on Engineering Management*, pp. 1–12, 2021.

[18] D. J. Edwards and G. D. Holt, "Predicting construction plant maintenance expenditure," *Building Research Information*, vol. 29, no. 6, pp. 417–427, 2001.

[19] H. L. Yip, H. Fan, and Y. H. Chiang, "Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and box-jenkins time series models," *Automation in Construction*, vol. 38, pp. 30–38, 2014.

[20] D. Berrar, "Cross-validation," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1-3, no. April, pp. 542–545, 2018.

[21] N. Makhathini, I. Musonda, and A. Onososen, "Utilisation of remote monitoring systems in construction project management," *Lecture Notes in Civil Engineering*, vol. 245, pp. 93–100, 2023.

[22] G. Guo, W. Zhu, Z. Sun, S. Fu, W. Shen, and J. Cao, "An aero-structure-acoustics evaluation framework of wind turbine blade cross-section based on gradient boosting regression tree," *Composite Structures*, vol. 337, no. June 2023, p. 118055, 2024.

[23] A. Shehadeh, O. Alshboul, R. E. Al Mamlook, and O. Hamedat, "Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, lightgbm, and xgboost regression," *Automation in Construction*, vol. 129, p. 103827, 2021.

[24] Y. Alzubi, "Comparison of various machine learning models for estimating construction projects sales valuation using economic variables and indices," *Journal of Soft Computing in Civil Engineering*, vol. 8, no. 1, pp. 1–32, 2024.

[25] O. Ersöz, A. F. İnal, A. Aktepe, A. K. Türker, and S. Ersöz, "A systematic literature review of the predictive maintenance from transportation systems aspect," *Sustainability*, vol. 14, no. 21, 2022.

[26] P. Parmar. (2021) Outlier detection and removal using the iqr method. Accessed: 2024-09-15. [Online]. Available: https://medium.com/@pp1222001/outlier-detection-and-removal-using-the-iqr-method-6fab2954315d

[27] M. Guerrero Cano, A. Luque Sendra, J. R. Lama Ruiz, and A. Córdoba Roldán, "Predictive maintenance using machine learning techniques," *Proceedings from International Congress on Project Management and Engineering*, 2019.

[28] S. Hosny, E. Elsaid, and H. Hosny, "Prediction of construction material prices using arima and multiple regression models," *Asian Journal of Civil Engineering*, vol. 24, no. 6, pp. 1697–1710, 2023.

[29] M. A. Mediavilla, F. Dietrich, and D. Palm, “Review and analysis of artificial intelligence methods for demand forecasting in supply chain management,” *Procedia CIRP*, vol. 107, pp. 1126–1131, 2022.

[30] M. K. Das and K. Rangarajan, “Performance monitoring and failure prediction of industrial equipments using artificial intelligence and machine learning methods: A survey,” *Proceedings of the 4th International Conference on Computational Methodologies and Communication (ICCMC)*, vol. 2020, pp. 595–602, 2020.

[31] N. Pavičić, Z. Rešetar, and F. Lukić, “The impact of the increase in raw material prices on costs in the construction sector in the city of osijek,” *1st International Scientific Conference on Economy, Management and Information Technologies – ICEMIT 2023*, 2023.

[32] J. Leukel, J. González, and M. Riekert, “Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review,” *Journal of Manufacturing Systems*, vol. 61, no. October, pp. 87–96, 2021.

[33] M. A. Musarat, W. S. Alaloul, A. M. Khan, S. Ayub, and N. Jousseaume, “A survey-based approach of framework development for improving the application of internet of things in the construction industry of malaysia,” *Results in Engineering*, vol. 21, no. January, p. 101823, 2024.

[34] O. T. Sanchez *et al.*, “An iiot-based approach to the integrated management of machinery in the construction industry,” *IEEE Access*, vol. 11, no. January, pp. 6331–6350, 2023.

[35] R. Hidayawanti and Y. Latief, “Raw material optimization with neural network method in concrete production on precast industry,” *International Journal of GEOMATE*, vol. 24, no. 102, pp. 10–17, 2023.

[36] L. Zhang, J. Guo, X. Fu, R. L. K. Tiong, and P. Zhang, “Digital twin enabled real-time advanced control of tbm operation using deep learning methods,” *Automation in Construction*, vol. 158, no. December 2023, p. 105240, 2024.

[37] J. Brozovsky, N. Labonnote, and O. Vigren, “Digital technologies in architecture, engineering, and construction,” *Automation in Construction*, vol. 158, no. November 2023, p. 105212, 2024.

[38] O. Alshboul, A. Shehadeh, M. Al-Kasasbeh, R. E. Al Mamlook, N. Halalsheh, and M. Alkasasbeh, “Deep and machine learning approaches for forecasting the residual value of heavy construction equipment: a management decision support model,” *Engineering, Construction and Architectural Management*, vol. 29, no. 10, pp. 4153–4176, 2022.

[39] H. Yang *et al.*, “Optimization of tight gas reservoir fracturing parameters via gradient boosting regression modeling,” *Heliyon*, vol. 10, no. 5, p. e27015, 2024.