Research on Engineering Structures & Materials
www.jresm.org
MIM Research Group

Research Article

Development and validation of a random forest model for predicting total construction waste cost in school building renovation projects

Hassanean S. H. Jassim [*,1,a], Ahmed R. Radhi [2,b], Musaab F. Hasan [3,c]

[1]*Department of Civil Engineering, College of Engineering, University of Babylon, Babylon, Iraq*
[2]*College of Engineering Al Musaib, University of Babylon, Babylon, Iraq*
[3]*General Directorate of Education Baghdad Rusafa First, Ministry of Education, Baghdad, Iraq*

Article Info

Article History:

Received 29 May 2026

Accepted 14 Aug 2026

Keywords:

Construction waste; School renovation projects; Random Forest Regression; Waste cost estimation; Sustainable construction

Abstract

The production of construction waste during the renovation of schools causes significant economic and environmental challenges, highlighting the need for a reliable tool to estimate waste costs during the planning phase. Based on project-level variables available during renovation planning, a Random Forest regression model was developed to predict the total construction waste cost of public-school renovation projects. The model used ten derived material-waste ratios together with renovation area as project-level predictors, while the response variable was the total direct construction waste cost. The model was developed using data from 50 completed school renovation projects in Iraq. Its predictive performance was evaluated against Linear Regression, Decision Tree Regression, and Support Vector Regression models using the same training and testing datasets and 5-fold cross-validation. Feature-importance analysis was used to identify the predictors contributing most strongly to the predicted waste cost. Among the evaluated models, Random Forest achieved the best predictive performance, with a Mean Absolute Percentage Error (MAPE) of 11.14% and an R^2 of 0.944. Brick waste ratio and renovation area were identified as the most influential predictors of waste cost. The proposed Random Forest model provides a practical decision-support tool for estimating construction waste costs during the planning phase of school renovation projects, supporting more efficient budgeting and sustainable construction waste management. The findings are applicable within the scope of the investigated dataset and provide a basis for future validation using larger and more diverse project datasets.

1. Introduction

In the 21st century, global solid waste management has become a major challenge facing governments and industry professionals worldwide, especially in developing countries, due to the weak legislation and regulations necessary to support effective strategies against this problem. Furthermore, renovation operations and activities, such as demolition, structural modification, and finishing works, generate substantial quantities of waste materials, known as construction and maintenance debris. Research in this area has shown that solid waste generated from construction, demolition, deconstruction, renovation, periodical maintenance, and remodeling projects amounted to 40 % of the worlds' total solid waste [1]. This percentage is increasing day by day, as the volume of solid waste is expected to exceed two billion tons by 2025 [2]. Therefore, improper management of this waste can lead to inflated project costs as well as having an environmental impact.

*Corresponding author: hassanean.salam.eng@uobabylon.edu.iq
[a]orcid.org/0000-0003-0465-8304; [b]orcid.org/0000-0001-5844-8789; [c]orcid.org/0000-0001-7036-0106
DOI: http://dx.doi.org/10.17515/resm2026-1719st0529rs

Building renovation projects are essential for maintaining existing building structures and ensuring the continuity of public services in order to fulfil the increasing needs of service continuity. Therefore, school buildings often need to be renovated to meet the standards of modern educational life and provide a comfortable environment. These renovation projects often produce a considerable amount of construction material waste that can be categorized into unused, damaged and excessive wastes. The quantity and characteristics of construction waste materials vary significantly from one project to another depending on project conditions, construction practices, and material handling procedures, significantly impacting project costs and environmental sustainability [3]. However, most renovation projects do not include waste-related costs in a systematic way in the budgets.

During renovation works, it is possible to encounter some differences when estimating the items of the Bill of Quantities (BOQ) because of unforeseen structural conditions of buildings and finishing work, which can cause an excess of waste products. Previous studies indicate that these waste materials represent between 10% and 15% of total project costs, depending primarily on the type of project and management practices; therefore, this requires significant attention from participants and stakeholders in construction projects to address the problem of material waste and its negative impact on project costs [4–7]. In many cases, the necessity to remove extra quantities of items listed in the BOQ may even be inevitable if structural members' damage is increased [8]. About 30% of the world's total solid waste is generated from demolition and construction activities, which can raise material costs to more than 40% of the actual cost [9,10]. According to the U.S. Environmental Protection Agency (EPA), construction and demolition waste (CDW) constitutes a significant portion of total municipal solid waste, with concrete materials being one of the most common waste materials [11]. These massive quantities choke the environment and threaten the existence of any project requiring accurate and sound waste management approval [12]. Furthermore, accurately estimating the costs of waste materials is essential not only to ensure sustainability and compliance with global regulations, but also to ensure budget adequacy.

A substantial proportion of school infrastructure in developing countries is aging and continues to increase over time, especially with the increase in population [13]. In this context, the School Construction Authority, particularly in Iraq, oversees a large number of renovation projects, including those of obsolescent educational facilities. Therefore, managing waste materials from renovation projects is a key criterion for achieving sustainable building value, something that remains elusive. On the other hand, most renovations involve major construction work, such as replacing floors, roofs, HVAC (heat, ventilation and air conditioning) systems, redesign of the interior environment necessary to meet flexible learning requirements and electrical systems [14, 15]. Thus, effective management of this issue can turn it from a conflict into a great opportunity in terms of cost, materials, time and human effort [16]. However, validated predictive models for estimating the total construction waste cost during the planning stage of school renovation projects remain limited, particularly in developing countries such as Iraq. It should be emphasized that the proposed Random Forest model predicts the total construction waste cost at the project level. The renovation project area with waste-related variables, including the derived waste ratios for different construction materials, is used solely as predictor (input) variables, whereas the target variable of the model is the total construction waste cost.

The remainder of this paper is organized as follows: The theoretical background and literature review of construction waste management and construction cost estimation models are presented in Section 2. Methodology used for the research is detailed in Section 3, which consists of data collection, selection and analysis of factors, development of the Random Forest model, and validation of the model. The results of the study are presented and discussed in Section 4, highlighting practice and theory implications. Finally, Section 5 summarizes the conclusions and provides recommendations for future research.

2. Construction Waste Management and Problem Statement

Construction waste can also be defined as any losses resulting from construction activities that generate direct or indirect costs but do not add any value to the planned final product [17,18]. It is worth noting that construction waste differs from municipal waste and typically comes from renovation, construction, modification, and demolition of roads, buildings, and other facilities [19]. Numerous studies have been conducted in America, Japan, England, France, Germany, Denmark, India and other developed and developing countries which have shown the possibility of recycling and reusing construction waste materials from construction site, such as bricks, bitumen and other components of waste materials from construction projects [20], which certainly leads to cost savings and a reduced environmental impact from the construction sector. Basically, the waste material from these processes and activities primarily includes concrete, asphalt, wood, metals, gypsum boards, and roofing [21]. In this context, there are a number of studies that have focused on identifying the main components of construction waste materials in building projects [22]. For example, Atabi et al., 2023 [23] reported that soil waste constitutes the majority of the total waste generated during building construction, followed by brick, concrete, mortar, ceramics, marble, gypsum, steel, wood, and subbase. In Shenzhen, southern China, concrete waste accounts for the largest proportion of total waste generated from residential building construction projects. In Thailand, waste materials such as metal components, wood, and concrete materials are the main contributors to construction waste, which can add value to the project when efficiently estimated [24]. Therefore, construction companies around the world should pay more attention to activities involving concrete, steel, formwork, sand, soil, and piles as the factors that most contribute to waste generation at construction sites [25].

In Jordan, a study was conducted to identify several factors contributing to material waste at construction sites based on a survey to find out the total material waste at Jordanian construction sites, which ranged between 15% - 21% of total waste generated [5]. Researchers J. Radosavljevic and A. Vukadinovic [26] examined construction, demolition, and renovation waste in Serbia, classifying it according to the European Waste Catalogue, and highlighting regulatory procedures and obstacles to waste recycling, which vary depending on materials, technologies, and project types. Zunjarrao [27] emphasized the importance of implementing various remedial techniques and measures that can control the impact of waste materials on construction project costs and schedules in India. In the United Arab Emirates, waste material management in the construction industry was investigated to identify the causes behind waste generation while also revealing the waste management process, improvement methods, professionals' responsibilities, and their perspectives on the subject [28]. A quantitative model developed to evaluate the social performance of construction waste management using a system dynamics approach through an empirical case study was able to demonstrate its potential to enhance such practices [29]. Overall, despite the significant contributions of such studies to knowledge of WG (Waste Generation) and WM (Waste Management), most studies are descriptive/survey based with little development of predictive models that could estimate waste cost at the project level. Furthermore, existing studies are mainly concerned with waste characterization, causes or percentage estimation but have not tried to combine waste quantity and cost in a single predictive framework.

To capture the differences between previous studies and the current study, the differences are summarized in Table 1 below in respect of country, methodology, size of dataset, predicted output, and limitations.

As illustrated in Table 1, the present study differs from previous research by developing a Random Forest machine learning model to predict the total construction waste cost at the project level using data collected from completed public school renovation projects in Iraq. Iraq is one of the developing countries experiencing an increasing number of construction and renovation projects due to obsolescence of existing school buildings and the rise in living standards associated with modern educational requirements, which has led to increased construction waste requiring improved cost estimation.

A study in Iraq [30] investigated factors contributing to construction waste management in Sulaymaniyah city and reported that traditional construction methods are the main reason for generating large amounts of waste materials. Inadequate infrastructure, insufficient regulations, and low public awareness were highlighted as major challenges requiring modern infrastructure, public education, and waste-to-energy solutions [31]. In another study [32], a construction material waste management system was proposed for local construction projects in Iraq and was positively evaluated for improving waste reduction through integration with MS Project. Furthermore, another study [33] indicated the significant effects of material damage and dual handling tasks on construction waste generation through 15 factors grouped into on-site material management, material handling and transportation, and site management practices.

Table **1**. Comparative analysis of previous studies and the proposed framework

Study	Country	Method	Dataset	Output	Limitation
[5]	Jordan	Survey/statistics	Unknown	Waste %	No prediction model
[26]	Serbia	Classification	Literature-based	Waste types	No cost estimation
[27]	India	Qualitative	Not specified	Waste factors	No ML
[28]	UAE	Descriptive analysis	Case study	Causes of waste	No predictive model
[29]	Mixed	System dynamics	Empirical case	Social performance	Not project-level cost
This study	Iraq	Random Forest	50 projects	Total construction waste cost	Limited dataset but validated

Construction and renovation work in public educational facilities such as school buildings is typically subject to limited budgets and strict accountability requirements [34]. Thus, it is crucial to estimate construction waste amounts and costs accurately to effectively plan the budget, use resources in a sustainable way and meet environmental requirements. While many studies have helped identify waste generation factors, waste classification, waste recycling policies and better management practices [23–35] few have focused on providing practical predictive models, capable of estimating waste quantities and costs in the planning phase of renovations.

While previous studies have made important contributions to the estimation of construction waste or to its management, most of them have employed statistical or qualitative methods with only a few using machine learning techniques. Furthermore, most of the models available are based on waste generation factors, as opposed to combining waste quantity and cost in a single predictive model. Most studies rely on fairly large data sets in developed countries, and there is little research on project-level data sets of high quality and small size in developing countries like Iraq.

Based on those limitations, the present study designs and validates an integrated machine learning model using the Random Forest method to predict the cost of construction waste generated in school building renovation projects. The model was developed based on data gathered from fifty school renovation projects that have been completed in Baghdad, and includes the renovation area with the key materials that contribute to the generation of waste identified from the literature review and expert consultation. To address project-scale differences, renovation area was incorporated as an explicit predictor alongside the material waste ratios, while material quantities and unit prices were retained for calculating the observed waste-cost target rather than used as independent predictors. Unlike previous studies that primarily focused on waste generation or material-level analyses, the proposed model provides a practical decision-support tool for estimating total construction waste cost during the planning stage of renovation projects. The model also incorporates feature importance analysis to identify the project variables that most strongly influence the predicted waste cost, thereby supporting more effective budgeting and construction waste management. The study fills the gap of knowledge providing a statistically validated predictive method that is especially applicable for school renovation projects in Iraq where there are no prediction models.

The proposed model was developed from the data extracted from completed work of school renovation in Baghdad, Iraq, mostly in finishing and maintenance rather than major structural rehabilitations. Thus, the results of this study may be only generalizable with regard to other types of projects, construction conditions or geographical areas. Furthermore, the long-term validity of the developed cost estimation model could be affected by the price changes of materials and labor over time, and its use may need to be recalibrated from time to time to ensure its applicability in changing economic conditions.

3. Methodology

The study methodology was designed to achieve the research objective of developing an effective machine learning model for predicting the total construction waste cost of school renovation projects. The methodology began with a comprehensive review of the literature on construction waste management and cost prediction models, followed by the collection and analysis of data from completed public school renovation projects in various Iraqi governorates (Fig. 1). The renovation area together with waste-related project variables derived from material consumption records were analyzed to identify the most influential factors affecting the predicted construction waste cost and to develop a reliable project-level cost prediction model. The data and information for the study were collected in Baghdad to serve the purpose of this study because of its high population density of Iraqi residents. In addition, in terms of the number of buildings in populated areas, Baghdad is the densest region in Iraq, with various types of educational school buildings that are in a state of obsolescence. From this point of view, it is believed that the models developed from this study will represent the investigated school renovation projects in Baghdad and provide a useful basis for future validation in other regions of Iraq.

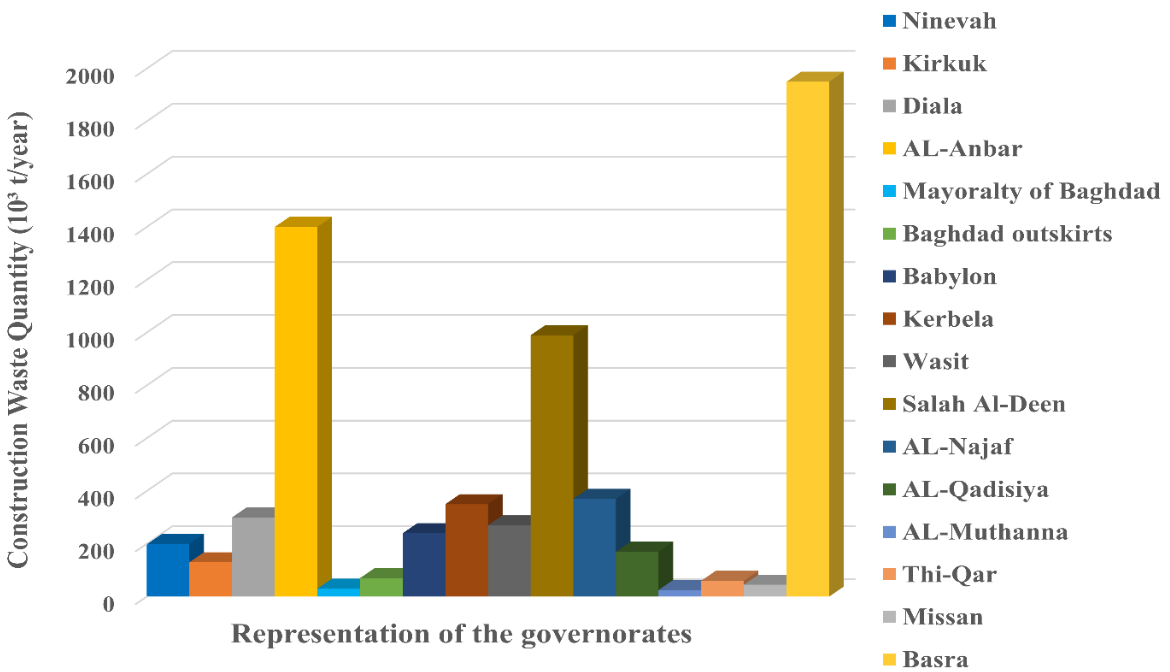

Fig. 1. Construction waste quantities generated in different Iraqi governorates (10^3 t/year). Source: Engineering Departments of the General Directorates of Education in the Iraqi governorates under the Iraqi Ministry of Education.

This methodology is mainly carried out through the following steps: a literature review to investigate the status of waste management methods in construction projects by studying the most relevant research; data collection and analysis, which involves collecting real-world data on construction projects with documented waste problems in different locations in the capital; developing the mathematical model by creating a model to predict the cost of school renovation waste based on historical data using a Random Forest model. Model validation and evaluation can be performed by validating data, which will be analyzed to evaluate the model's performance. Fig. 2 illustrates the overall methodology and details that represent a rigorous framework for understanding and extracting data related to developing a construction waste cost prediction model for school building renovation.

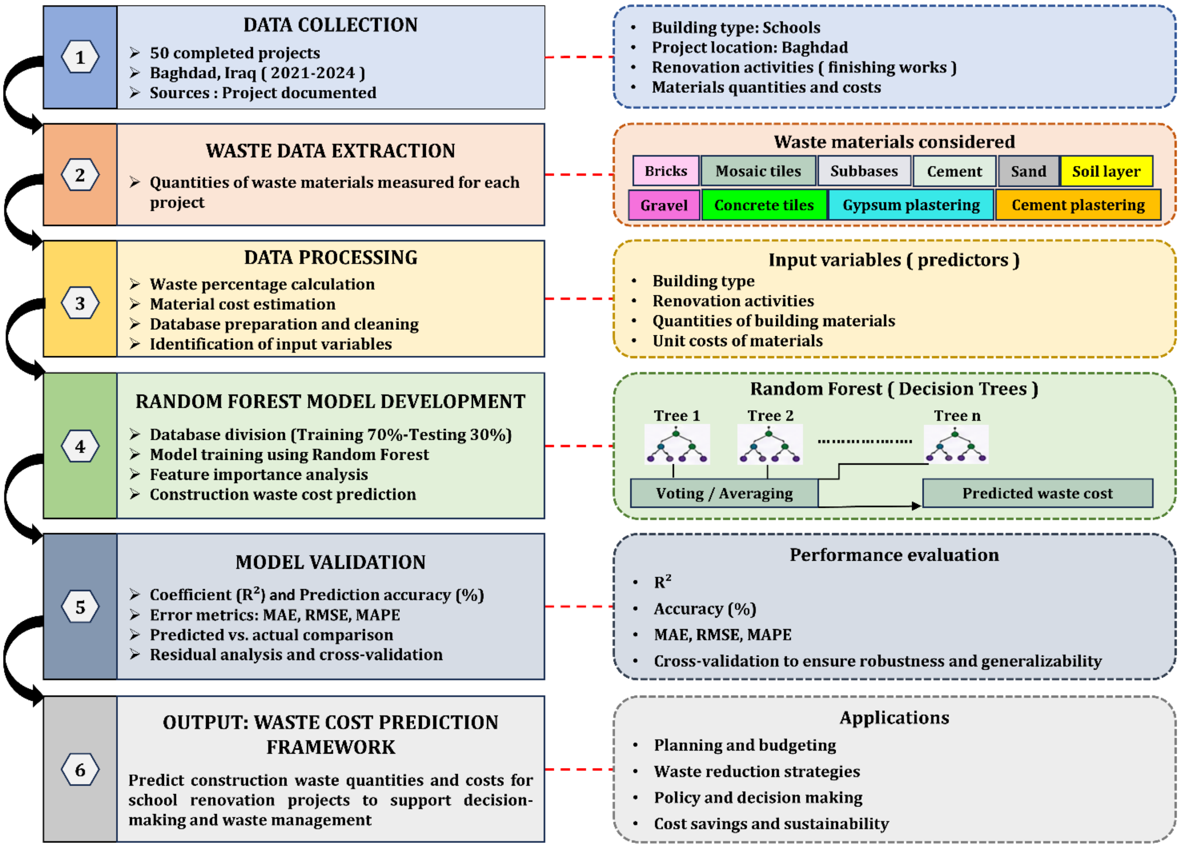

Fig. 2. Methodological framework for project-level construction waste cost prediction in school renovation projects using the Random Forest model.

3.1 Waste Materials and Selection Factors

According to the construction waste management literature, accurate characterization of construction materials for each construction activity is essential for estimating the amount of waste generated during building renovation projects. In this regard, effective waste management should address major waste materials according to the types and characteristics of renovation projects. A variety of materials, such as wood, metal, concrete, and composites, can be generated from renovation activities and may be suitable for reuse or recycling. Table 2 summarizes the major waste materials identified from previous studies of school renovation and demolition projects and considered in the selection of material categories for the present study.

Within the methodology adopted in this study, the proposed Random Forest-based construction waste cost prediction model was developed using project-level data obtained from completed school renovation projects. The model inputs comprised eleven predictors: ten derived material waste ratios and renovation area (RA). The ten waste ratios represented the investigated material categories, namely bricks (BR), mosaic tiles (MT), subbases (SB), cement (CE), sand (SA), gravel (GR), soil layer (SL), concrete tiles (CT), gypsum plastering (GP), and cement plastering (CP). Renovation area (RA) was included as an explicit project-scale predictor. The model output (target variable) was the total construction waste cost (WC) at the project level.

Material quantities, measurement units, consumed quantities, unit prices, and material-level costs were not used as independent predictors. Instead, the relevant material quantities and waste measurements were used to derive the waste ratios and to calculate the observed project-level waste-cost target. Thus, the final model design matrix consisted exclusively of the eleven predictors

BR, MT, SB, CE, SA, GR, SL, CT, GP, CP, and RA, while WC represented the response variable. Accordingly, the final design matrix consisted of eleven predictor variables, expressed as X = [BR, MT, SB, CE, SA, GR, SL, CT, GP, CP, RA], while the response variable was Y = WC, representing the total direct construction material waste cost at the project level. The material-level quantities, physical units, and unit prices shown in Table 5 were used for data preparation and calculation purposes only and were not entered into the Random Forest model as predictors.

Construction and renovation data were collected from completed public-school renovation projects in Iraq using project documents, interviews with engineering personnel, and field observations. The investigated material categories were selected based on the literature review and the available project records. The collected information was processed to derive the material waste ratios used as model predictors and to calculate the corresponding project-level waste cost. Table 3 presents the allowable waste limits for the investigated building materials according to the 2014 Standard Cost Analysis Manual for Building and Construction, Part 1, issued by the Ministry of Construction, Housing, Municipalities and Public Works (MCHMPW) of Iraq [35].

Table **2**. Most common waste materials collected from school renovation projects

Waste Material	Material source	Issue or hazard	Reuse or Disposal	Reference
Wood	Doors, flooring, fixtures	May contain insects or chemicals	Firewood or milled	[36]
Metal	Windows, doors, steel beams, metal framing, garden fence, ducts	Corrosion	Melt and reform new products	[37]
Bricks, concrete and ceramic, and marble tiles	Walls, ceiling, flooring and foundation	None	Landfill and crushed aggregate	[38]
Gypsum products	Ceilings and Wall	Toxic gases in a wet climate	Recycling new construction products	[39]
Electrical and plumbing fittings	Pipes, wires and fixtures	Contaminated fixtures	Reuse and recycling	[40]

Table 3. Permissible limits for waste of construction materials according to (MCHMPW)

Material	Waste Percentage (%)	Material	Waste Percentage (%)
Construction Materials			
Poured Concrete	4.0 – 5.0	Concrete Precast Units	3.0 – 5.0
Non-packed Cement	6.0 – 7.0	Finishing Concrete Tiles	5.0 – 10.0
Fine Aggregate	6.0 – 9.0	Marble Tiles	5.0 – 7.0
Coarse Aggregate	6.0 – 9.0	Masonry Blocks	5.0 – 8.0
Subbase	15.0 – 20.0	Angle	3.0 – 4.0
Clay Brick	10.0 – 20.0	Reinforcement Steel	4.0 – 6.0
Forms	5.0 – 7.0	B.R.C	5.0 – 7.0
Insulation Layers	5.0 – 8.0	Galvanized Steel (Ducts)	3.0 – 4.0
Ceramic Tiles	4.0 – 7.0	Binding Wires	3.0 – 5.0
Thermo-Stone	10.0 – 18.0	Lime & Gypsum	10.0 – 15.0
Casting In Situ	5.0 – 10.0	Mosaic Tiles	10.0 – 15.0
Electrical Materials			
Wires	4.0 – 5.0	Electrical Pipe	4.0 – 5.0
Wall Socket Box	1.0	Plug Socket	1.0 – 2.0
Plumbing Materials			
Galvanized Pipe	4.0 – 6.0	Fittings	4.0 – 5.0
Taps	1.0 – 2.0	Washbasins	1.0

3.2 Data Collection

Data were collected as a case study to provide a cost breakdown of school renovation project activities, highlighting indicators of construction waste materials and disposal costs. A total of 50 completed school renovation projects were obtained from the General Directorates of Education in Baghdad, Iraq. The works for all projects were completed in full between 2021 and 2024 and mainly comprised the completion of floor coverings and plastering; there was limited work undertaken on the structural components. Projects were chosen based on availability of complete technical and financial documentation, project drawings, bills of quantities, material procurement records and site execution reports. Projects with missing or incomplete documentation were excluded and only completed school renovation projects with complete and verifiable documentation were included in the study. The projects varied in the quantity of materials involved and the extent of renovation, but each project was of a similar type of school finishing renovation projects, resulting in a uniform dataset for the different projects, where the influential range values of building materials and costs for the aggregated dataset elements are shown in Table 4.

Table 4. Scope of waste construction materials collected according to characteristics of renovated project studies

Statistical range	Renovation area	Bricks	Mosaic Tiles	Subbases	Cement	Sand
Unit	(m^2)	(m^3)	(m^2)	(m^3)	(kg)	(m^3)
Min	91.7	1.372	1.976	1	37.52	1.902
Max	240.12	28.95	61.8	103	1000	46.35

Statistical range	Gravel	Soil Layer	Concrete Tiles	Gypsum Plastering	Cement Plastering	Supply cost	Waste cost
Unit	(m^3)	(m^2)	(m^2)	(m^2)	(m^2)	(IQD)	(IQD)
Min	2	0.684	0.72	4.5016	6.4368	1134560	966775.59
Max	108.8296	100.2215	184.688	350	177.4461	115391450	13070120.5

The dataset was created from a variety of data sources such as project drawings, project reports, material procurement records and verified quantity statements. Recorded quantities and waste estimates were further clarified and verified with one key stakeholder per project (usually the site engineer or project manager) in a semi-structured conversation. The research team applied a standardized data collection form in all 50 projects to ensure method uniformity.

Each record collected was put through a structured data audit process. Based on the official project documents, execution records in the construction site, and procurement's invoices, the material quantities were cross-checked from the reported quantities. All projects were finished, no missing nor incomplete records were found and all data sets were completely validated before analysis. All projects followed the same measurement and verification process to ensure consistency and comparability of the data. The unit prices for construction waste costs were taken from the official bills of quantities and financial records maintained by each education General Directorate for completed projects. The physical waste quantities used in this study were derived from the official material issue and consumption records available in the completed project documentation. These records were used as the basis for calculating the waste ratios of the investigated construction materials. It should be noted that the available project records did not explicitly distinguish discarded materials from reusable leftovers, returned materials, or transferred stock. Therefore, the recorded waste quantities were accepted as documented in the official records. Since the objective of this study was to develop a predictive model using project-specific data rather than to compare project costs across different years, the official project costs recorded at the time of project completion were used without adjustment for inflation. The percentage of material waste (W %) was calculated using Eq. (1).

$$W\% = \frac{SM - CM}{SM} \times 100\% \tag{1}$$

Where, SM: The supplied material quantity of each material needed in any item mentioned in the project bill of quantities, CM: The actual amount of consumed material for the executed work items.

To simplify the explanation of the overall calculation procedure, an example illustrating the calculation of construction waste-related variables for one school renovation project is presented in Table 5. As mentioned earlier, the examined school renovation projects (research sample) are classified as finishing work defect projects. First, Eq. (1) is applied to the primary construction materials used in the renovation works to calculate the waste-related variables. These calculated variables represent the predictor (input) variables used in the proposed machine learning model rather than the model outputs.

Table 5. Example calculation of waste quantities and costs for a school renovation project

No.	Material Item	Abbrev.	Supply Quantity	Consumed Quantity	Waste Quantity	Unit	Waste ratio (%)	Price (IQD)	Cost (IQD)
1	Bricks	BR	20	17	3	m^3	15.00	79000	237000
2	Mosaic Tile	MT	250	245	5	m^2	2.00	38000	190000
3	Subbases	SB	200	190	10	m^3	5.00	30000	300000
4	Cement	CE	36000	35600	400	kg	1.11	7000	2800000
5	Sand	SA	35	26	9	m^3	25.71	25000	225000
6	Gravel	GR	8	6	2	m^3	25.00	20000	40000
7	Soil Layer	SL	120	100	20	m^3	16.67	5000	100000
8	Concrete Tiles	CT	800	750	50	m^2	6.25	5000	250000
9	Gypsum Plastering	GP	1000	850	150	m^2	15.00	6000	900000
10	Cement Plastering	CP	600	520	80	m^2	13.33	15000	1200000
11	Renovation Area	RA	---	---	---	m^2	---	---	---

Note: Table 5 illustrates the data-preparation procedure for calculating waste ratios and project-level waste cost. Supply quantities, consumed quantities, waste quantities, unit prices, and material-level costs were not used as independent predictors in the Random Forest model. Only the ten derived waste ratios and renovation area (RA) were used as model inputs.

The input variables used in the predictive model represent the project renovation area (RA) with derived waste ratios of ten construction material categories, namely bricks (BR), mosaic tiles (MT), subbases (SB), cement (CE), sand (SA), gravel (GR), soil layer (SL), concrete tiles (CT), gypsum plastering (GP), and cement plastering (CP). Each construction material variable corresponds to the percentage waste calculated using Eq. 1, which was then transformed into decimal form before modeling (e.g., 15% to 0.15). All predictors are therefore made dimensionless and only the derived waste ratios were fed into the model, despite the fact that the material quantities and the physical units (m^3, m^2, kg) are reported in Table 5, except for the renovation area, which is used as an actual value.

The output variable is the total cost of construction waste (WC) for the whole renovation work (in Iraqi Dinars (IQD)). Therefore, the model does not predict material-level waste costs, but rather the overall project level waste cost. In other words, the target variable (WC) represents the total direct construction material waste cost at the project level as shown in Eq. (2). It is calculated exclusively from the monetary value of wasted construction materials recorded in the official Bills of Quantities (BOQs) and verified financial records. Therefore, WC represents the direct purchase value of wasted materials and does not include indirect waste-management costs such as onsite hauling, sorting, temporary storage, landfill charges, recycling operations, replacement, rework, or the residual value of reusable materials. The project-level waste cost is calculated as:

$$WC = \sum_{i=1}^{n} WQ_i \times UP_i \tag{2}$$

Where: WC is total direct construction material waste cost for the project (IQD), WQ_i is wasted quantity of material i, UP_i is unit price of material i (delivered in site) obtained from the project Bills of Quantities (BOQs) and financial records, n = number of investigated material categories.

3.2.1 Data Preprocessing

The final input dataset comprised eleven predictor variables, including ten dimensionless waste ratios and one project-scale variable, namely the renovation area (RA). The ten waste ratios were initially expressed as percentages and were converted to decimal values before model development. In contrast, RA was retained in its original unit of m^2 to preserve the project-scale information required for estimating the absolute waste cost. The target variable, waste cost (WC), was retained in Iraqi dinars (IQD) to ensure direct interpretation of the model predictions.

Feature scaling or normalization was not applied to the predictor variables. This is because the Random Forest (RF) regression algorithm is based on threshold-based recursive partitioning rather than distance- or magnitude-based calculations. Consequently, monotonic transformations such as standardization or normalization do not generally alter the ordering of observations or the resulting tree partitions. The ten waste ratios were therefore represented as decimal values, while RA remained in m^2 and WC remained in IQD throughout model development.

The inclusion of RA as the eleventh predictor was specifically intended to provide a project-scale measure that complements the dimensionless waste ratios. Thus, projects with similar waste-generation ratios but different renovation areas can be distinguished by the model when estimating the corresponding absolute waste cost. No transformation or standardization was applied to WC because retaining the original IQD unit allows the predicted values to be directly interpreted as project-level waste costs.

3.3 Random Forest Algorithm

Random Forest Regression (RFR) is a supervised learning regression tree algorithm that has been proven to be useful for predicting complex and nonlinear interactions [41]. A regression tree partitions the data in order to reduce the Prediction Error between actual and predicted value(s), and predicts a numerical result [42]. In regression, one needs to have a dependent variable (the target variable) that is to be predicted, and a set of components to be used for that prediction. The complete dataset includes the target variable and its associated features. The regression tree recursively partitions the data starting from a root node and generates branches until terminal nodes are reached. Each branch is generated by an internal node, which is the point where the tree splits. The closing node on any branch represents the expected value of the target variable and the final prediction. A Random Forest is an ensemble of regression trees that generates predictions by averaging the outputs of the individual trees [43].

The Random Forest algorithm was selected for this study due to its ability to handle nonlinear relationships, high-dimensional datasets, and interactions between multiple input variables without requiring strict statistical assumptions. In addition, Random Forest is less sensitive to overfitting compared to single decision trees and has demonstrated strong performance in construction-related prediction problems in previous studies. Therefore, it was considered suitable for modelling the complex relationship between waste-related predictor variables and the total construction waste cost for renovation projects. Eq. (3) shows the prediction formula based on random forest regression.

$$\hat{y}(x) = \frac{1}{M}\sum_{m=1}^{M} T_m(x) \tag{3}$$

Where, $\hat{y}(x)$: Target variable of the random forest at a new point ▯ is the average of the predictions of all individual trees, M: Number of trees in the forest, $T_m(x)$: Prediction from the m_{th} decision tree.

3.4 Model Development

At this stage of model development, the dataset was randomly divided into two subsets: a training set and an independent testing set. In particular, 86% of the data were used for training the models, and the other 14% for independent testing, which consisted of 7 school projects. This division allows for the model to be tested against data that it has not previously encountered, giving an unbiased estimate of how well the model can predict new data.

Furthermore, the model was cross validated using a 5-fold cross validation (k = 5) on the training set to improve robustness and reliability of the model. The training data was randomly divided into five subsets (folds) in this procedure. The data was split by fold and each fold was used once for validation, while the rest were used for training. The final performance was the mean of the performances obtained from all the folds which minimized sampling bias and increased the stability of the evaluation results.

The model is the Random Forest Regression model which was built in MATLAB with the help of Scikit-learn library. The model was set up with five decision trees (n_estimators = 5) and a maximum depth of three levels (max_depth = 3) to ensure that the model does not overfit and remains manageable. The size of the data set and the computational efficiency were the reasons why these hyperparameters were chosen by an empirical trial and error method, instead of a full grid search. Therefore, the hyperparameters were optimized by evaluating the number of trees (n_estimators) within the range {5, 10, 20, 50, 100, 200}, and examining the maximum tree depth (max_depth), the minimum number of samples required to split an inner node (min_samples_split), and the minimum number of samples required at a leaf node (min_samples_leaf) within predefined candidate values. The model performance for each hyperparameter combination was evaluated using k-fold cross-validation (k = 5) on the training dataset, and the optimal configuration was selected based on the lowest root mean square error (RMSE). Mean Absolute Error (MAE) and coefficient of determination (R^2) were also recorded for comparison. The final configuration is a compromise between predictive accuracy, interpretability and generalization ability. To guarantee the reproducibility of the results in the modeling process, the random seed was fixed (random_state = 42). Hyperparameter tuning was conducted only on the training set with 5-fold cross validation procedure, and the independent testing set was not used at all in the process of tuning hyperparameters, but was only used to evaluate the final model.

The robustness and stability of the model was also tested using the 5-fold cross validation method on the training set only and the results are presented as mean ± standard deviation in Table 6.

Table 6. Cross-validation performance of the Random Forest model (training set)

Metric	Mean	SD
Mean Absolute Error (MAE)	657573	25755
Root Means Square Error (RMSE)	885320	120795
Mean Absolute Percentage Error (MAPE %)	23.31	3.47
Coefficient of Determination (R^2)	0.819	0.283

The Random Forest model performed well with the mean R^2 of cross-validation 0.819 ± 0.283 and MAPE of 11.14 ± 3.47%, both indicating that the model has stable predictive performance across all folds, and that the model was very robust and had only limited sensitivity to data partitioning.

The final predictive performance was tested on the independent test set, and the benchmark models were tested on the same training, testing, and validation framework to make a fair comparison. A comparative assessment of all the benchmark models is given in Table 8 with the same test conditions. The detailed performance of the Random Forest model and the summary metrics are provided in the next section (Tables 9 and 10).The model performance of the Random Forest demonstrated satisfactory predictive performance on the investigated dataset at the cross-validation and independent test stages and seemed to be not significantly overfitted from the investigated dataset.

Finally, the estimated construction waste cost was calculated as the average of the five decision trees outputs as shown in Eq. (4), based on the Random Forest model of the developed model, which is presented in Eqs. (5 -9) (numerical inequality), which shows the “if” logical value for each tree. The principle of the Random Forest algorithm is that the averaging process is the main idea; that is, the final prediction is the average of the predictions of individual base learners. The nonlinear relationships learned between the input predictor variables (construction material waste ratios) and the target variable (total construction waste cost) are shown in the individual decision trees of the ensemble (Figures 3–7).

$$WC = \frac{T_1(x)+T_2(x)+T_3(x)+T_4(x)+T_5(x)}{5} \tag{4}$$

$$T_1(x) = \{IF(BR \leq 0.2293, IF(SB \leq 0.10865, IF(GP \leq 0.0954, IF(GR \leq 0.3002\ 1856776.22455556, IF(GP \leq 0.0904, IF(SA \leq 0.1759, IF(SB \leq 0.0921, 2175257.91342857, 3073782.00333333), 3544716.26), \ 1676376.886))\ 3568520.2), 5410571.4), 13070120.5)\} \tag{5}$$

$$T_2(x) = \{IF(BR \leq 0.2319, IF(BR \leq 0.19295, IF(SA \leq 0.17905, IF(SB \leq 0.0756, 1296985.795, IF(SL \leq 0.0252, IF(MT \leq 0.0186, 3263195.7, 2184999.37686667), IF(RA \leq 143, 2110693.899, 3870336.865))), 3795747.51), 5785894.725), 13070120.5)\} \tag{6}$$

$$T_3(x) = \{IF(BR \leq 0.2319, IF(RA \leq 151.1, IF(SL \leq 0.02465, IF(MT \leq 0.02135, 1513966.53333333, 2316258.55675), IF(RA \leq 115.05, \ 1969252.32466667, IF(SA \leq 0.18285, IF(SL \leq 0.02645, \ 2776055.37028571, 3642280.52333333), 2405374.2275))), 5045057.12), 13070120.5)\} \tag{7}$$

$$T_4(x) = \{IF(RA \leq 160.3, IF(GP \leq 0.0803, IF(CT \leq 0.22935, IF(CT \leq 0.22265, 2153320.02966667, 3010321.0876), 2009428.07788889), IF(CP \leq 0.2317, 3943713, IF(GR \leq 0.3459, 2061157.77633333, 3339995.494))), 8483278.875)\} \tag{8}$$

$$T_5(x) = \{IF(BR \leq 0.19265, IF(RA \leq 133.05, IF(SA \leq 0.1796, IF(CE \leq 0.0473, 1250616.1084, IF(SB \leq 0.0913, IF(CE \leq 0.0538, 1903205.42577778, 1469042), 2314864.193)), 4192240.19), IF(BR \leq 0.18275, 3853498.08175, 3242795.29966667))\ , 8387605.2)\} \tag{9}$$

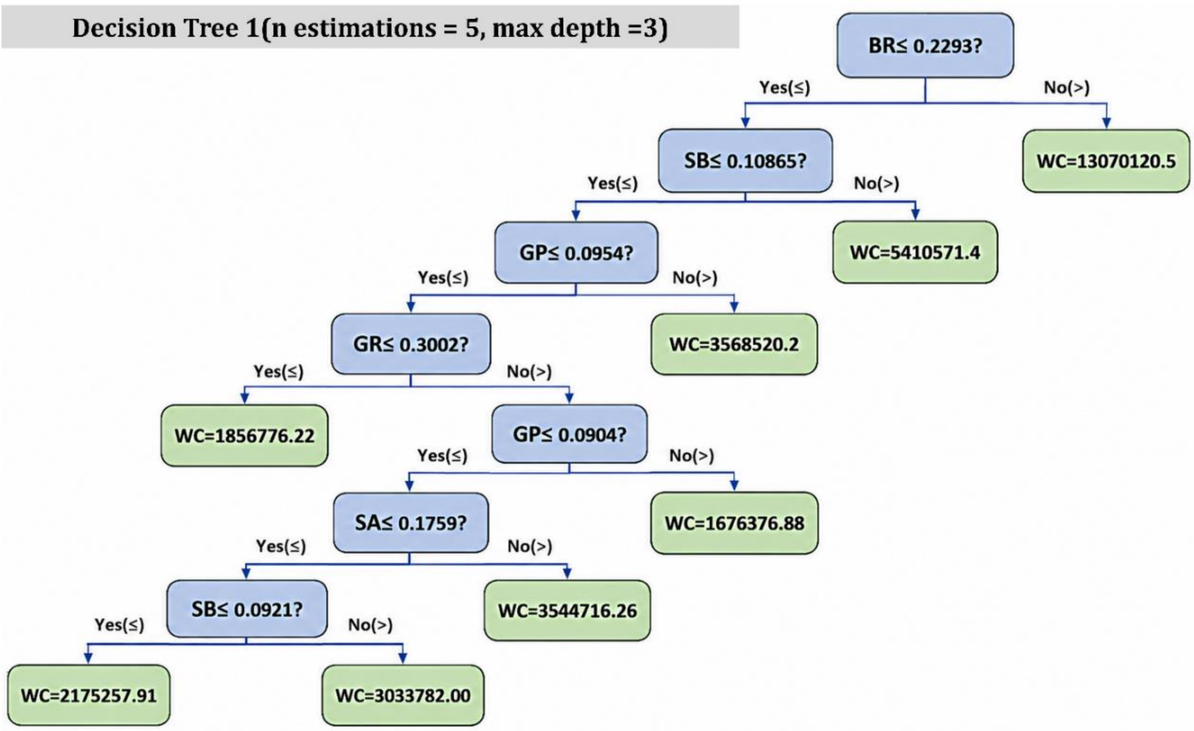

Fig. 3. Decision tree 1 for waste cost with n_estimator 5 and max_depth 3

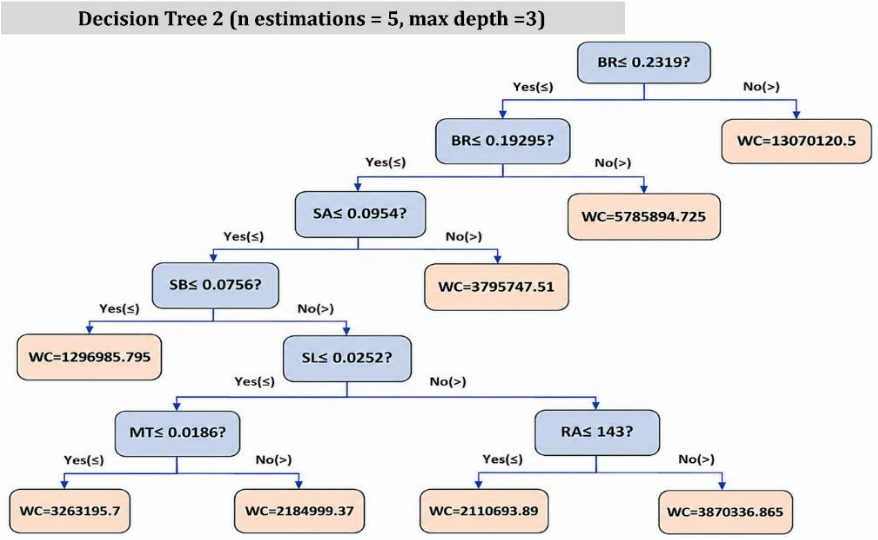

Fig. 4. Decision tree 2 for waste cost with n_estimator 5 and max_depth 3

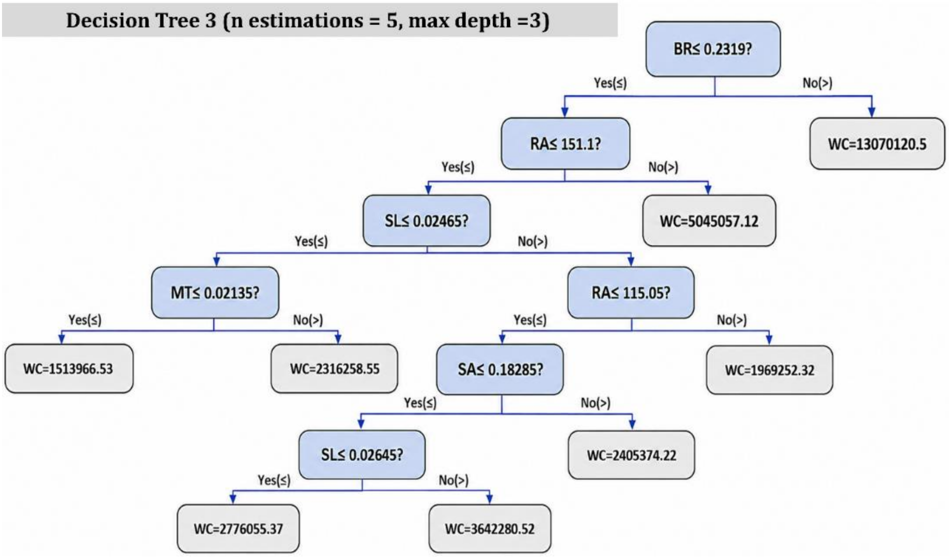

Fig. 5. Decision tree 3 for waste cost with n_estimator 5 and max_depth 3

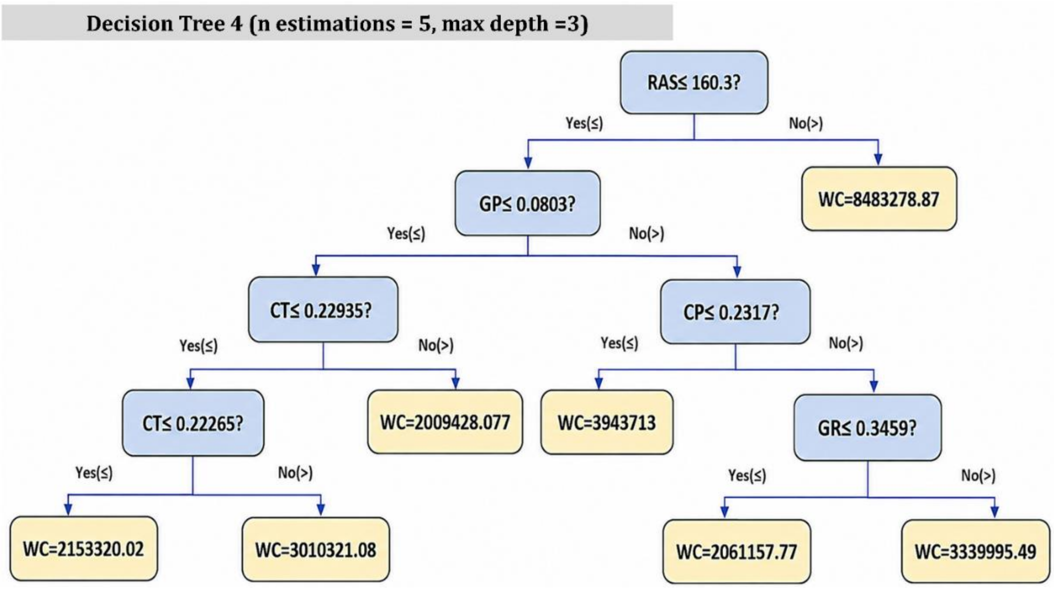

Fig. 6. Decision tree 4 for waste cost with n_estimator 5 and max_depth 3

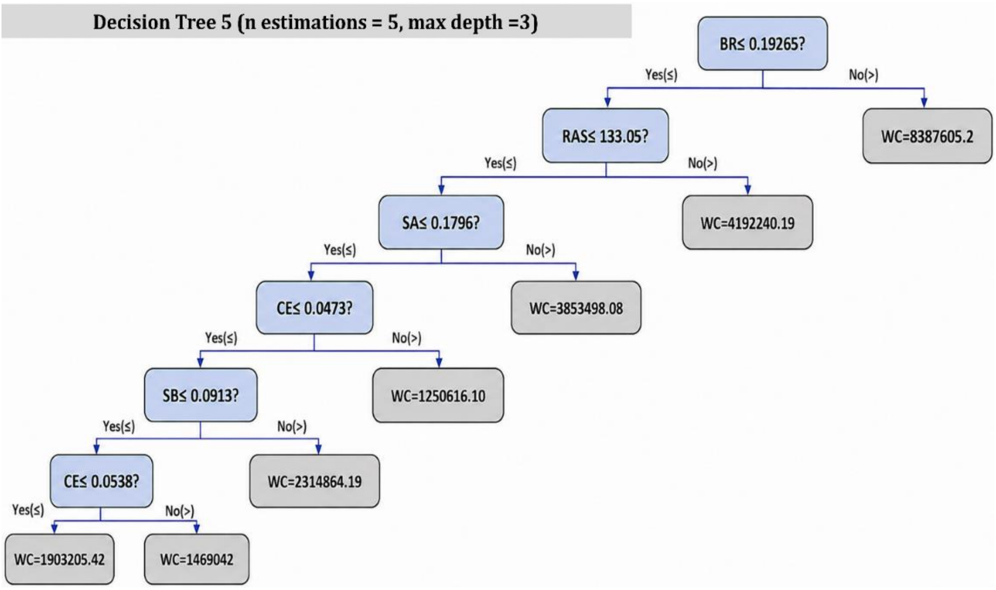

Fig. 7. Decision tree 5 for waste cost with n_estimator 5 and max_depth 3

3.4.1 Benchmark Models Comparison

The predictive performance of the proposed Random Forest (RF) model was compared with three commonly-used machine learning models namely Linear Regression (LR), Decision Tree Regression (DT) and Support Vector Regression (SVR). All models were implemented with the use of the Scikit-learn library and the models were trained by splitting the data into the train and test sets in the same proportion (86–14) and with the same number of folds (5) for a fair and unbiased comparison. To guarantee a full reproduction of the modelling process, the final hyperparameter value used for all machine learning models are summarized in Table 7.

Table 7. Hyperparameter settings of the evaluated machine learning models

Model	Hyperparameters
Random Forest (RF)	n_estimators = 5, max_depth = 3, random_state = 42, bootstrap = True
Decision Tree (DT)	max_depth = 3, random_state = 42
Support Vector Regression (SVR)	kernel = RBF, C = 1.0, epsilon = 0.1, gamma = scale
Linear Regression (LR)	Default Scikit-learn parameters

The predictive performance of all models on the independent testing dataset is presented in Table 8. The proposed Random Forest model achieved the highest predictive performance and lowest predictive error as compared to all the benchmark models. The ensemble-learning mechanism of this superior performance allows combining multiple decision trees to capture nonlinear relationships, to decrease the variance of prediction and to increase the generalization of the model.

Table 8. Performance comparison of benchmark models on the independent testing dataset

Model	MAE	RMSE	MAPE (%)	R^2
Linear Regression (LR)	947231	1131275	33.35	0.674
Decision Tree (DT)	765407	1062420	23.50	0.713
Support Vector Regression (SVR)	635501	838454	20.07	0.7278
Random Forest (RF)	237949	282380	11.14	0.9441

The proposed model showed very good predictive abilities, although it is important to approach the results with the necessary precautions due to the small sample size of 50 completed school renovation projects. To ensure reliability and to minimize potential overfitting, 5-fold cross validation was used as a part of the model development and the final model performance values were reported in the independent testing set only (Table 8). The cross-validation procedure was only used to check the robustness and stability of the model and was not used to report the test performance.

3.5 Model Performance

The final predictive performances of the proposed Random Forest model are provided only in the testing (or independent) dataset, in Tables 8 and 9. The proposed Random Forest model was tested for its predictive performance by following the validation procedure outlined in Section 3.4. Performance metrics shown in Table 9 are only for the independent testing dataset and the 5-fold cross validation approach was used during the model development to check for robustness and stability. Such a validation strategy guarantees unbiased performance assessment of the predictive capability and minimizes overfitting. In regression based machine learning applications, the assessment of the predictive accuracy and generalization ability of a model is an important part of the model development and validation process [44].

The size of construction waste cost was predicted using Eq. 3 and comparison between the predicted and observed values is shown in Table 9. The results show good consistency with the waste costs predicted and the waste costs in the project.

The percentage error in predicting the accuracy of the prediction was also calculated for 7 school renovation projects of 10 waste material categories with Eq. (10). The prediction error was between −38.94% and +9.44%, which means that the actual construction waste cost was slightly over or underestimated by the prediction. The very low error range observed is due to the high predictive capacity of the proposed model.

$$Prediction\ Error\ \% = \left(\frac{Actual\ WC - Predicted\ WC}{Actual\ WC}\right) \times 100\% \tag{10}$$

In practice, a deviation within this range is deemed acceptable for preliminary budgeting and site estimation of construction waste during the early stages of a project, when uncertainties regarding material quantities and site conditions are accepted. Hence, the suggested model is able to assist in the decision-making process in the budgeting, procurement planning and construction waste management process.

For prospective application, the ten waste-ratio inputs do not represent observed waste from the project being planned. Instead, they are estimated from historical waste-generation records of comparable completed school renovation projects and, where available, from material quantities reported in the preliminary BOQ and the corresponding historical waste rates. The renovation area (RA) is obtained directly from the project design or BOQ. These planning-stage estimates are then entered into the trained Random Forest model to obtain an early estimate of the total construction waste cost before renovation activities begin. Thus, the model does not require actual waste quantities or post-construction consumption records from the new project at the prediction stage.

Besides project level evaluation, several statistical performance measures such as MAE, RMSE, MAPE, and R^2, were calculated to assess the performance of the regression model as recommended in the assessment of regression models [45]. The results are presented in Table 10 which shows the high predictive accuracy with R2 value of 0.944 on the independent testing dataset.

In spite of the good prediction performance of the model, it is necessary to consider that there are only 50 completed renovation projects in the data set, which might affect the applicability of the proposed application. To overcome this drawback, the model was tested using 5-fold cross validation and compared with other machine learning algorithms. It is important to note that the performance reported in this paper describes only the behavior of the proposed model when tested on the projects from the studied data set of 50 projects and should not be considered as proof of general applicability outside the studied field. Therefore, future studies should consider larger and more diverse datasets to further enhance model robustness, external validity, and generalizability.

Table 9. Performance of RF model

BR	MT	SB	CE	SA	GR	SL	CT	GP	CP	RA
0.1843	0.0323	0.0818	0.0485	0.1742	0.3246	0.0271	0.2124	0.08	0.2304	92
0.1974	0.0433	0.0557	0.0419	0.1857	0.3697	0.0309	0.1904	0.0993	0.265	146.5
0.1814	0.0323	0.081	0.0489	0.1747	0.3306	0.027	0.2133	0.0838	0.2305	106.1
0.1952	0.0413	0.0574	0.0433	0.1847	0.3665	0.0302	0.1933	0.097	0.2584	135.9
0.1712	0.025	0.0994	0.0536	0.1655	0.3031	0.0248	0.2271	0.0708	0.2103	102.7
0.1801	0.0316	0.0812	0.0495	0.1742	0.3306	0.0268	0.2143	0.083	0.2279	117.2
0.1848	0.036	0.0704	0.0479	0.1778	0.3421	0.0288	0.2048	0.0875	0.239	131.3

Actual WC (IQD)	Predicted WC (IQD)	Prediction Error %
1484380	2062346	-38.94
4746787	4697478	1.04
2192117	2420425	-10.41
4858716	4697478	3.32
2772325	2510644	9.44
2957416	2755030	6.84
2304910	2489669	-8.02

Table 10. Statistical evaluation for RF model

Performance Metric	Result
Mean absolute error (MAE)	237949
Root means square error (RMSE)	282380
Mean absolute percentage error (MAPE %)	11.14
Coefficient of Determination (R^2)	0.944

3.6 Feature Relative Importance Analysis of Predictor Variables

The relative contribution of each predictor variable to the Random Forest model was quantified using feature importance analysis. Furthermore, a number of algorithms have been used to achieve this goal previously; for example, importance values were obtained directly from the trained model from the Random Forest model using the feature importance scores obtained from the scikit-learn's implementation [46]. Hence, the feature relative importance was calculated using a heterogeneity-based feature relative importance metric, also known as the mean decrease in impurity (MDI) that has implemented in the Random Forest Regressor model from the Scikit-learn library. Therefore, the relative importance of each feature is calculated as the average weighted reduction in node impurity, i.e., the variance of the regression trees resulting from splits on that feature across all trees in the forest model. The yielding importance values are relative, specific to the analyzed dataset and the trained model, and should be interpreted as indicators of the model's reliance on each feature rather than as direct measures of prediction error reduction or evidence of causal relationships.

Therefore, the significance of changing the validation data was then calculated by altering each predictive variable individually and measuring the resulting deterioration in prediction performance. This process was repeated across several resampling iterations. Importance scores were then normalized to sum up to 100% to compare the relative contribution of each predictor variable towards the prediction of the total cost of construction waste. These are not causal values, but rather show the relative importance of each predictor to the model output for the data set analyzed.

The result shown in Table 11 and Fig. 8 indicates that the most influential predictor was brick waste ratio (BR) (67.20 %), renovation area (RA) (20.53 %), subbase waste ratio (SB) (3.69%), sand waste ratio (SA) (2.71%), gypsum plastering waste ratio (GP) (1.68 %) and gravel waste ratio (GR) (1.16 %). Additionally, soil layer waste ratio (SL) (1.07%), cement plastering waste ratio (CP) was included in the model but with a much smaller importance (0.94%). The other variables had a smaller contribution to the predictive model. These results were validated for each predictive variable based on the mean switching significance and standard deviation, and were calculated across all redistributions, to record the frequency with which each predictive variable appeared among the highest-ranking variables, which is considered to have relatively stable significance when it maintains a high mean significance with a relatively low variance and appears consistently among the highest-ranking predictive variables across repeated redistributions. Figures 9 to 11 show good statistical indicators for predictive variables in the overall datasets.

Table 11. Feature relative importance of predictor variables with others stability analysis indicators in the Random Forest model

Variable	Best_Model_RF_Relative_Importance_Percent	Best_Model_Rank	CV_RF_Mean_Relative_Importance_Percent	CV_RF_SD_Relative_Importance_Percent	CV_Mean_Importance_Rank	CV_RF_Mean_Rank	CV_RF_Median_Rank	Permutation_Top1_Frequency_Percent	Permutation_Top3_Frequency_Percent	Permutation_Top5_Frequency_Percent
RA	20.53	2	23.636	15.949	1	2.62	2	39	69	79
BR	67.20	1	19.647	16.683	2	3.57	3	12	37	53
SB	3.69	3	18.361	18.007	3	4.25	3	17	40	50
GP	1.68	5	12.031	10.366	4	4.28	4	13	42	61
CT	0.41	9	6.6496	11.084	5	6.86	7	1	15	36
SL	1.07	7	5.5491	7.1104	6	6.51	6	4	20	41
GR	1.16	6	3.7484	5.4445	7	7.23	8	5	23	38
MT	0.36	10	3.3289	5.4172	8	7.47	8	5	16	39
SA	2.71	4	2.8925	3.4439	9	7.23	7	1	20	45
CE	0.26	11	2.4699	2.8263	10	7.38	7	1	8	30
CP	0.94	8	1.6877	2.8857	11	8.6	9	2	10	28

In summary, the results indicate that waste cost is influenced by both project scale and the material waste characteristics within the investigated dataset. The inclusion of renovation area (RA) provides an important project-scale descriptor that complements the material waste ratios. The relatively high contribution of RA (20.53%) indicates that project size contributes substantially to the prediction of total waste cost. Thus, projects with comparable waste ratios may still have different total waste costs when their renovation areas differ. Understanding the most influential predictors can help project managers and engineers focus on the material categories and project-scale characteristics that are most relevant to planning, procurement, budgeting, and waste management operations.

In the visual display of the resulting values, Figures 8 to 13 summarize these relative importance values of the input variables across all iterations and folds using their mean and standard deviation. For example, the stability of each predictive variable was assessed based on its mean importance in switching with the standard deviation of its importance, as well as the frequency with which it was ranked among the most influential predictive variables across repeated regroupings.

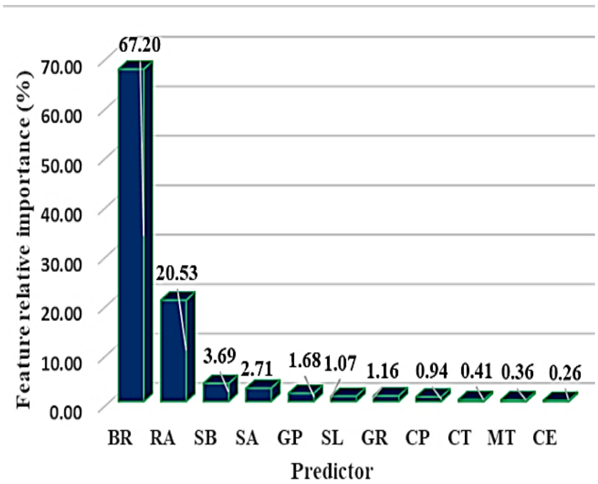

Fig. 8. Feature importance of predictor variables in the Random Forest model

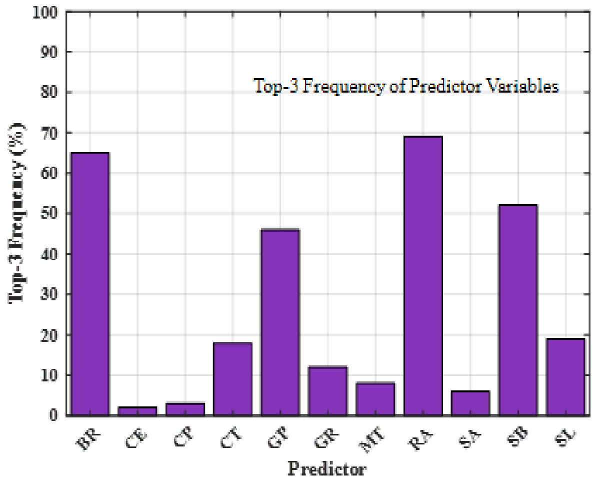

Fig. 9. Top-3 frequency of predictor variables in the Random Forest model

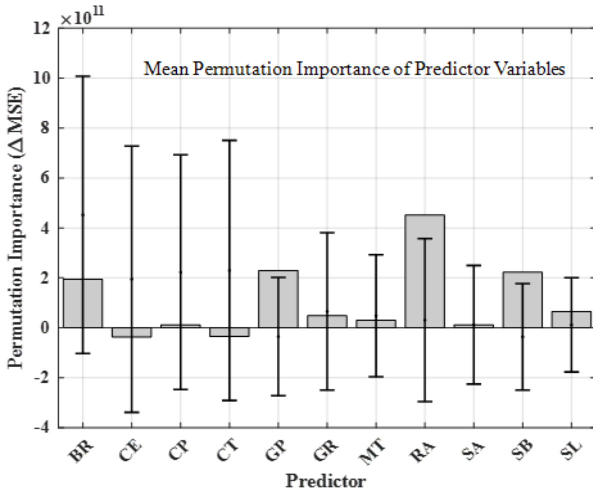

Fig. 10. Mean permutation importance of predictor variables in the Random Forest model

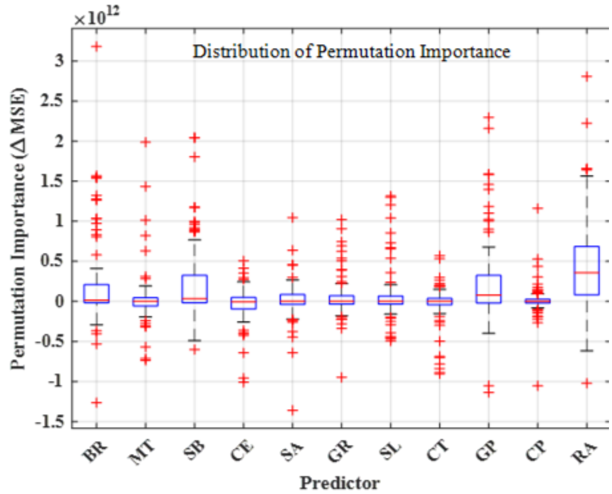

Fig. 11. Distribution of permutation importance of predictor variables in the Random Forest model

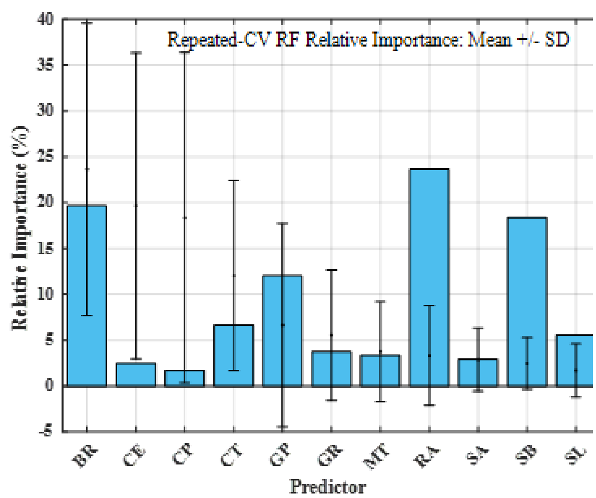

Fig. 12. Repeated-CV (Mean +/- SD) for features relative importance in the Random Forest model

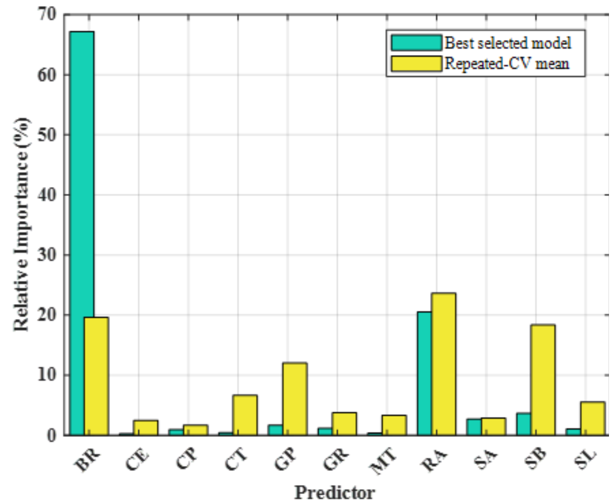

Fig. 13. Best Random Forest Model vs Repeated- Mean Random Forest Importance

The results showed that the main predictive variables identified in the original random forest analysis remained consistently among the most influential variables during repeated regroupings, although some variation was observed in their relative order. This suggests that the overall significance pattern is reasonably stable, while the precise ranking of individual predictive variables should not be interpreted as definitive.

3.6 Modifying Random Forest Model for Cost Inflation

Changes in building material prices are a key indicator of inflation in the construction sector, particularly in economic studies related to the built environment and construction, which examine the price range of a variety of raw materials and products over different time periods. To measure the rate of change in product prices, prices in the base period are compared with statistical data from the subsequent period, reflecting the percentage increase in the price of each material. Therefore, processing material price data involves more than just calculating the absolute price difference; it primarily relies on converting this difference into a comparable percentage across different materials and time periods. Furthermore, using a weighted index improves the accuracy of estimating the impact of inflation on total costs, particularly in studies involving a large number of materials of varying relative importance. In this study, when analyzing a group of ten construction waste materials, the inflation rate for each material is calculated individually. The overall inflation rate for the group can then be calculated to obtain a preliminary index of the overall price change, which is calculated using equation (11), and then the adjusted cost of construction waste by equation (12).

$$i_T = \sum_{j=1}^{n} PRI_j \times (\frac{P_2 - P_1}{P_1})_j \tag{11}$$

$$WC_M = WC \times (1 + i_T) \tag{12}$$

Where, $\boldsymbol{i_T}$: The overall weighted inflation rate for all individual building materials (decimal value), PRI: Feature relative importance value of each construction materials (decimal value), $\mathbf{P_1}$: Reference price (i.e., Table 5) for waste materials used in building a Random Forest model, $\mathbf{P_2}$: New price for waste materials at forecast time, $\mathbf{WC_M}$: Cost of construction waste (IQD) adjusted based on inflation rate, WC: Cost of construction waste (IQD) calculated by Eq. (4).

4. Results and Discussion

The results show that the various school renovation projects that have been investigated have a significant amount of variation in their construction waste, due to differences in types of materials used, the type of renovation work, and site conditions. The largest volume of construction waste among the surveyed projects was found in the structural and finishing materials.

To develop the proposed Random Forest model for the total cost estimation of construction waste using waste ratios of construction materials, a detailed dataset was collected from some real building renovation projects. Project-level waste related predictor variables, based on the construction material waste ratios for various renovation activities, are used in the model. Therefore, the results indicate that waste estimation costs should be considered from the initial budget cycle as these costs can help to manage the cost of renovation, which will benefit sustainable asset management throughout the building life cycle, both economically and environmentally [47].

A breakdown of waste volume on each project is indicative of the importance of checking and adapting waste management techniques in the field. Specifically, a recycling and re-use strategy can significantly lower the waste management and disposal costs, especially for high volume materials such as structural materials and finishes found in educational buildings [48]. These are reduction levels reported in the construction waste management literature where optimum material recovery and recycling systems have been shown to markedly alleviate disposal pressures and lifecycle costs.

The obtained results were further assessed through comparison with previously reported empirical studies on construction and demolition waste generation. For instance, brick waste generation in China has been reported at approximately 12% of total construction waste during building phases [49], while values around 10% have been observed in India [50] and between 10–13% in the United Kingdom [51]. Similarly, studies conducted in Iraq reported cement waste at approximately 4%, sand at 12.5%, and brick waste at 15% in construction projects, while governmental projects show higher waste levels, including 28% for concrete and 24% for finishing materials [52]. Other results show that the waste from the gypsum plastering is about 16% and flooring tile waste is from 10–12% [53].

These differences are due to the variations in construction practices, skills of the workforce, material handling systems, and modern waste management practices between countries [54]. Similar ratios of concrete to brick to gypsum waste have been reported in China (45, 25, 5)%, Turkey (55, 35, 5)%, and the United Kingdom (65, 30, 3.5)%, respectively [54–56].

The distribution patterns extracted in this research are similar to the ones given by other international studies and the results obtained are within the range of those given by other studies. This agreement does not demonstrate external validity of the model, but rather demonstrates its applicability on the level of a practice within the context of the project discussed. Before generalizing the model to other geographical regions, further testing with larger and more diverse datasets is needed. The results of the feature importance analysis suggest that brick waste ratio (BR), renovation area (RA), subbase waste ratio (SB), sand waste ratio (SA), gypsum plastering (GP), and gravel waste ratio (GR) are the most influential predictor variables. These variables had the greatest influence on the prediction of the total cost of construction waste within the data set investigated. The relatively high importance of these waste ratios indicates that the differences between those waste ratios have a higher impact on the model's prediction of the total construction waste cost for the investigated set of data. This could be linked to high demand for these materials in school renovation works and the possibility of their loss due to handling and transport, cutting, storage and installation on site. The stability analysis results confirmed these findings, showing that the main predictive variables identified by the original Random Forest model generally remained among the most influential variables during repeated regroupings. However, some slight variation was observed in the relative ranking of the predictive variables, particularly among variables with correlated material-to-waste ratios. Therefore, the relative importance results for the characteristics are interpreted as indicators of their relative predictive contribution within the current range of the dataset, and not as evidence of causation in all cases.

Project management issues that have been identified from the findings indicate the need for enhancing procurement planning, stock management and site management, particularly for the finishing work and flooring work that can create a significant waste volume. The quantity estimation process during the planning stage with more estimation performance will reduce unnecessary purchases of material and minimize surplus materials. Likewise, a more advanced system for material storage and a more rigorous management of material handling could result in significant reduction in the amount of waste that can be avoided and in the associated costs of disposal. The results also support the use of waste minimization strategies for re-use and the recycling of recoverable construction waste materials, and the continued tracking of project waste during the project lifecycle.

Within the investigated project context, the proposed predictive model has the potential to serve as a practical decision-support tool for public authorities and policymakers involved in planning and budgeting school renovation projects. It may help identify high-risk waste categories before construction activities begin and support more informed decisions regarding waste management, resource allocation, and sustainable renovation practices. In addition, predictive waste-cost estimation may contribute to reducing budget overruns, improving the accountability of public spending, and promoting sustainable construction practices.

In addition, the high coefficient of determination ($R^2 = 0.944$) together with the low prediction error values indicate that the proposed Random Forest model demonstrated good predictive capability within the investigated dataset. The close agreement between the predicted and observed total construction waste costs indicates the potential of the proposed model to reflect the general situation and application in Iraq in order to capture the relationships between the waste-related predictor variables and the project-level waste cost. Furthermore, taking into account inflation rates in future years based on the weighted percentage importance of variable-priced materials is a key advantage of adopting this model now, and the possibility of updating it in the future. However, these findings should not be interpreted as evidence of general applicability beyond similar school renovation projects without further external validation using larger and more diverse datasets. In other words, confirmation using a larger and independent dataset is needed before generalizing the relative importance of these indicators to other renovation projects that can be considered a clear path for developing future studies in this field.

The small range of variations shows that the model predictions are very close to the actual values, and the predicted values are in close agreement with the observed values. Some variation may be attributed to prices of materials, occurrence of site conditions left unanticipated, differences in the quality of workmanship, and differences in practice in the way the projects were executed at renovation sites, which are all factors that create a degree of uncertainty not fully represented in historical data.

The results from an engineering and project management viewpoint show that the waste production in renovation works is not random but is highly dependent on material type, construction technique and project specific situations. Therefore, incorporating a predictive construction waste cost model during the early stages of a project may improve budgeting accuracy, procurement planning, and contingency allocation within similar project contexts. Within the investigated project context, this predictive capability may help project managers mitigate budget uncertainty, improve resource utilization, and reduce unplanned waste-related expenditures. Such benefits are especially important when schools are undergoing renovation as budgets are often set and budget overruns can have a significant impact on the feasibility of a project.

In recent years, construction waste management has been considered a vital element of sustainable construction in public educational buildings where environmental compliance, safety and economic efficiency are all essential requirements. Concrete debris, bricks, gypsum, tiles, packaging wastes, plastics and glass are just a few of the waste streams created by school renovation projects. Unless managed correctly, these materials can cause degradation of the environment, high disposal expenses and poor project sustainability performance. Hence, it is crucial to estimate the amount of waste and its costs precisely not only to optimize the budget but also to facilitate the application of the principles of a circular economy in the construction industry [48].

This study has yielded promising results, but there are some restrictions to be taken into account. The information presented is applicable only to renovation works in the city of Baghdad, Iraq, and may not be representative of the generality of the model for other regions with different construction methods, regulations and climate. Secondly, the projects under analysis were mainly related to finishing and maintenance works, and not to major structural works, so that it may be assumed that the applicability of the models will not be affected by complex structural works. Thirdly, the long-term reliability of cost forecasts can depend on the long-term evolution of materials costs and labor costs, which may require the need for periodic re-calibration of the model.

5. Conclusions

- The present study developed and internally validated a practical Random Forest-based machine learning model for predicting the total construction waste cost of school renovation projects. The model was developed using data from 50 completed public-school renovation projects and incorporated ten material waste ratios together with renovation area (RA) as project-level predictors. The inclusion of RA provided an explicit descriptor of project scale, while the model target was the total direct construction waste cost.
- The proposed RF model achieved prediction errors ranging from −38.94% to +9.44%, with a MAPE of 11.14% and an R^2 of 0.944 within the investigated dataset. These results indicate promising predictive performance for the studied school renovation projects. However, the reported performance applies only to the investigated dataset and should be interpreted cautiously until externally validated using larger and more geographically diverse datasets.

- Renovation area showed a substantial predictive contribution within the investigated dataset, together with the most influential material waste variables. This finding indicates that both project scale and material waste characteristics contribute to the prediction of total construction waste cost. The identified influential predictors can assist project managers, engineers, and contractors in prioritizing material categories and project characteristics during planning, budgeting, procurement, and waste-management activities.
- The proposed framework provides a transparent and potentially useful decision-support approach for estimating construction waste costs during the planning stage of comparable school renovation projects. By providing an early estimate of waste-related costs, the model can support budgeting, procurement planning, resource allocation, and waste-reduction strategies.
- Although renovation area was incorporated to represent project scale, the current model does not explicitly include other scale- and market-related variables, such as contract value, detailed material quantities, unit-price indices, renovation scope, and price-year adjustments, as independent predictors. Therefore, the model should be interpreted as a project-specific waste-cost prediction model for school renovation projects represented by the investigated dataset, rather than as a universal estimator of absolute construction waste cost across projects with substantially different scales, scopes, or market conditions. Changes in material and labor prices may also require periodic model recalibration.
- The main limitations of the study are related to the relatively small dataset of 50 completed school renovation projects located in Baghdad, Iraq, with most projects involving maintenance and finishing works rather than major structural rehabilitation. Accordingly, the model's applicability to other building types, geographical regions, renovation scopes, and market conditions remains to be established. External validation using larger and more diverse datasets is therefore recommended before wider practical implementation.
- Future studies should evaluate the proposed approach using larger datasets covering different building types, geographical locations, renovation scopes, and economic conditions. Additional project-scale and market-related variables, such as contract value, detailed material quantities, unit-price indices, and price year, could be incorporated to improve the robustness and generalizability of waste-cost prediction. Future research could also investigate advanced and hybrid machine-learning methods, real-time project data, and the integration of environmental indicators and life-cycle assessment to develop more comprehensive economic and environmental decision-support systems for sustainable construction-waste management.

References

[1] Jain MS. A mini review on generation, handling, and initiatives to tackle construction and demolition waste in India. Environmental Technology & Innovation, 2021; 22: 101490. https://doi.org/10.1016/j.eti.2021.101490
[2] Mohajan HK. Impacts of construction and demolition waste on environment: An overview. Studies in Social Science & Humanities, 2025; 4(3): 1–5. https://doi.org/10.63593/SSSH.2709-7862.2025.05.001.
[3] Antoniou F, Aretoulis G, Giannoulakis D, Konstantinidis D. Cost and Material Quantities Prediction Models for the Construction of Underground Metro Stations. Buildings, 2023; 13(2): 382. https://doi.org/10.3390/buildings13020382
[4] Arsenos P, Giannadakis G. Construction projects' waste prevention and expected minimization of cost and environmental impacts through adopting a comprehensive system for document management. Environmental Research, Engineering and Management, 2023; 79(2): 77–87. https://doi.org/10.5755/j01.erem.79.2.33532
[5] Bekr GA. Study of the Causes and Magnitude of Wastage of Materials on Construction Sites in Jordan. Journal of Construction Engineering, 2014; 2014(1): 283298. https://doi.org/10.1155/2014/283298
[6] Oko John A, Emmanuel Itodo D. Professionals' views of material wastage on construction sites and cost overruns. Organization, technology & management in construction: an international journal, 2013; 5(1): 747–757. https://doi.org/10.5592/otmcj.2013.1.11
[7] Osmani M. Construction Waste Minimization in the UK: Current Pressures for Change and Approaches. Procedia - Social and Behavioral Sciences, 2012; 40: 37–40. https://doi.org/10.1016/j.sbspro.2012.03.158

[8] Domingo N, Batty T. Construction waste modelling for residential construction projects in New Zealand to enhance design outcomes. Waste Management, 2021; 120: 484–493. https://doi.org/10.1016/j.wasman.2020.10.010
[9] Ajayi SO, Oyedele LO. Critical design factors for minimising waste in construction projects: A structural equation modelling approach. Resources, Conservation and Recycling, 2018; 137: 302–313. https://doi.org/10.1016/j.resconrec.2018.06.005
[10] Ding Z, Zhu M, Tam VWY, Yi G, Tran CNN. A system dynamics-based environmental benefit assessment model of construction waste reduction management at the design and construction stages. Journal of Cleaner Production, 2018; 176: 676–692. https://doi.org/10.1016/j.jclepro.2017.12.101
[11] EPA. Recycling Economic Information (REI) Report. US Environmental Protection Agency (EPA), 2020.
[12] Sagan J, Mach A. Construction waste management: Impact on society and strategies for reduction. Journal of Cleaner Production, 2025; 486: 144363. https://doi.org/10.1016/j.jclepro.2024.144363
[13] Carney S. Reimagining our futures together: a new social contract for education. Comparative Education, 2022; 58(4): 568–569. https://doi.org/10.1080/03050068.2022.2102326
[14] Lafortune G, Ubaldi B. OECD 2017 OURdata Index: methodology and results. OECD Working Papers on Public Governance, 2018; (30): 0_1. https://doi.org/10.1787/2807d3c8-en
[15] Koskela L. Application of the new production philosophy to construction. Stanford university Stanford, 1992.
[16] Zhou W, Zhao L, Huang H, Chen Y, Xu S, Wang C. Automatic waste detection with few annotated samples: Improving waste management efficiency. Engineering Applications of Artificial Intelligence, 2023; 120: 105865. https://doi.org/10.1016/j.engappai.2023.105865
[17] T. FC, Lucio S, Claudia DC, L. IE. Material Waste in Building Industry: Main Causes and Prevention. Journal of Construction Engineering and Management, 2002; 128(4): 316–325. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:4(316
[18] Formoso CT, Isatto EL, Hirota EH. Method for waste control in the building industry. In Proceedings International Group for Lean Construction, vol 7, 1999; 325.
[19] Radhi AR, Nassar YS, Alhilli HK. A bidding assessment framework using CRITIC and MABAC methods. Engineering, Technology & Applied Science Research, 2025; 15(3): 23048–23053. https://doi.org/10.48084/etasr.10665
[20] T. P. C. Department. Study of the Guideline for Construction and Demolition Waste Management in Thailand. 2018.
[21] Sawant Surendra B, Manoj H, Madhav K. Impact of the construction waste on the cost of the project. International Journal of Engineering Research, Volume, 2016; (5): 126–128. https://doi.org/10.17950/ijer/v5i1/028
[22] Prairie Village KS. Characterization of building-related construction and demolition debris in the United States., 1998.
[23] Atabi KL, Reig-Cerdá L, Colomer-Mendoza FJ. Categorization and Quantification of Construction Waste of Commercial Buildings in Iraq. In International Conference on Geotechnical Engineering and Energetic, vol 427. EDP Sciences, 2023; 2006. https://doi.org/10.1051/e3sconf/202342702006
[24] Li J, Ding Z, Mi X, Wang J. A model for estimating construction waste generation index for building project in China. Resources, Conservation and Recycling, 2013; 74: 20–26. https://doi.org/10.1016/j.resconrec.2013.02.015
[25] Kofoworola OF, Gheewala SH. Estimation of construction waste generation and management in Thailand. Waste Management, 2009; 29(2): 731–738. https://doi.org/10.1016/j.wasman.2008.07.004
[26] Radosavljevic J, Vukadinovic A. CONSTRUCTION, DEMOLITION, AND RENOVATION WASTE. Facta Universitatis, Series: Working and Living Environmental Protection, 2024; 21(2): 133–141. https://doi.org/10.22190/FUWLEP240610012R
[27] Zunjarrao AR. Construction Waste Management. International Journal of Engineering Research & Technology (IJERT), 2018; 7(4). https://doi.org/10.17577/IJERTV7IS040137
[28] Hittini B. Construction waste management in UAE: An exploratory study. WIT Transactions on Ecology and the Environment, 2019; 238(8): 679–686. https://doi.org/10.2495/SC190581
[29] Yuan H. A model for evaluating the social performance of construction waste management. Waste Management, 2012; 32(6): 1218–1228. https://doi.org/10.1016/j.wasman.2012.01.028
[30] Agha M, Mohd Fairullazi Ayob. BUILDING CONSTRUCTION WASTE MANAGEMENT: EVALUATION IN KURDISTAN REGION OF IRAQ. Journal of Architecture, Planning and Construction Management, 2022; 12(1 SE-Articles). https://doi.org/10.31436/japcm.v12i1.661
[31] Kareem IM, Ahmed AM, Abdulateef LA. REVIEW OF SOLID WASTE MANAGEMENT CHALLENGES AND SOLUTIONS IN IRAQ. International Journal of Trendy Research in Engineering and Technology, 2024; 8(4): 23–26. https://doi.org/10.54473/IJTRET.2024.8404
[32] Khaled ZSM, Alshathr BS, Hadi AH. Development of construction material waste management system. Engineering and Technology Journal, 2015; 33(7A). https://doi.org/10.30684/etj.2015.106878

[33] Khaleel T, Al-Zubaidy A. Major factors contributing to the construction waste generation in building projects of Iraq. 3rd International Conference on Buildings, Construction and Environmental Engineering , Sharm el-Sheikh, Egypt, 2018; 162. https://doi.org/10.1051/matecconf/201816202034
[34] Grigoroiu MC, Țurcanu C, Constantin CP, Tecău AS, Tache I. Determinant Factors of the Subjective Perception of Educational Projects with European Funding. Sustainability, 2025; 17(19): 8637. https://doi.org/10.3390/su17198637
[35] Ministry of Construction, Housing M and PW. Iraqi Standard Cost Analysis Manual., 2014. https://iqmuqawalat.com/books
[36] Dahuri M, Che-Ani AI, Johar S, et al. Optimizing maintenance budget allocation in higher educational institutions: a systematic review of building condition assessment and service quality. Journal of Facilities Management, 2025; 24(2): 245–270. https://doi.org/10.1108/JFM-09-2024-0112
[37] Huang Z, Liu K, Duan J, Wang Q. A review of waste-containing building materials: Characterization of the heavy metal. Construction and Building Materials, 2021; 309: 125107. https://doi.org/10.1016/j.conbuildmat.2021.125107
[38] Rahmat MN, Saleh AM. Strength and environmental evaluation of building bricks using industrial waste for liveable environments. Construction and Building Materials, 2023; 403: 132864. https://doi.org/10.1016/j.conbuildmat.2023.132864
[39] Bumanis G, Zorica J, Korjakins A, Bajare D. Processing of Gypsum Construction and Demolition Waste and Properties of Secondary Gypsum Binder. Recycling, 2022; 7(3): 30. https://doi.org/10.3390/recycling7030030
[40] Trihadiningrum Y, Anandita FD, Nadira A. Electronic waste management in schools: a case of Surabaya City, Indonesia. Journal of Material Cycles and Waste Management, 2023; 25(1): 597–611. https://doi.org/10.1007/s10163-022-01540-4
[41] Rubbens P, Brodie S, Cordier T, et al. Machine learning in marine ecology: an overview of techniques and applications. ICES Journal of Marine Science, 2023; 80(7): 1829–1853. https://doi.org/10.1093/icesjms/fsad100
[42] Carrizosa E, Molero-Río C, Romero Morales D. Mathematical optimization in classification and regression trees. Top, 2021; 29(1): 5–33. https://doi.org/10.1007/s11750-021-00594-1
[43] Stagnitta TJ, Woda JC, Graziano AP. Random forest regression models for estimating low-streamflow statistics at ungaged locations in New York, excluding Long Island. US Geological Survey, 2025. https://doi.org/10.3133/sir20255060
[44] Varoquaux G, Colliot O. Evaluating machine learning models and their diagnostic value. Machine learning for brain disorders, 2023; 197: 601–630. https://doi.org/10.1007/978-1-0716-3195-9_20
[45] Akselrud CIA. Random forest regression models in ecology: Accounting for messy biological data and producing predictions with uncertainty. Fisheries Research, 2024; 280: 107161. https://doi.org/10.1016/j.fishres.2024.107161
[46] Saveleva E, Svitelman V, Blinov P, Valetov D. Sensitivity analysis and model calibration as a part of the model development process in radioactive waste disposal safety assessment. Reliability Engineering & System Safety, 2021; 210: 107521. https://doi.org/10.1016/j.ress.2021.107521
[47] Eze EC, Aghimien DO, Aigbavboa CO, Sofolahan O. Building information modelling adoption for construction waste reduction in the construction industry of a developing country. Engineering, Construction and Architectural Management, 2022; 31(6): 2205–2223. https://doi.org/10.1108/ECAM-03-2022-0241
[48] Benachio GLF, Freitas M do CD, Tavares SF. Circular economy in the construction industry: A systematic literature review. Journal of Cleaner Production, 2020; 260: 121046. https://doi.org/10.1016/j.jclepro.2020.121046
[49] Wang Qiankun, Chen Lei, Hu Ruibo, et al. An empirical study on waste generation rates at different stages of construction projects in China. Waste Management & Research, 2020; 38(4): 433–443. https://doi.org/10.1177/0734242X19886635
[50] S. Joshi DM and AS. Resource Efficiency and Recycling of Construction and Demolition Waste. In Anant National University, Ahmedabad, 2024.
[51] Silva R V, de Brito J, Dhir RK. Use of recycled aggregates arising from construction and demolition waste in new construction applications. Journal of Cleaner Production, 2019; 236: 117629. https://doi.org/10.1016/j.jclepro.2019.117629
[52] Hussain A. Evaluating waste material among construction project: identify reason waste and suggestion improvement designer, contractor and client practice. Journal of Korean Society of Environmental Engineers, 2021; 43(6): 453–460. https://doi.org/10.4491/KSEE.202143.6.453
[53] Alajeeli HKB, Al Kaabi SAM. A study of waste management reality in construction projects in Iraq. Wasit Journal of Engineering Science, 2016; 4(1): 75–92

[54] Huang B, Wang X, Kua H, Geng Y, Bleischwitz R, Ren J. Construction and demolition waste management in China through the 3R principle. Resources, conservation and recycling, 2018; 129: 36–44. https://doi.org/10.1016/j.resconrec.2017.09.029

[55] Öztaş SK, Yüksek İ. LIFE CYCLE ENVIRONMENTAL ANALYSIS OF BRICK PRODUCTION: TURKEY AS A CASE STUDY. Journal of Green Building, 2022; 17(2): 125–142. https://doi.org/10.3992/jgb.17.2.125

[56] Blay-Armah A, Bahadori-Jahromi A, Mylona A, Barthorpe M. A life cycle assessment of building demolition waste: a comparison study. Proceedings of the Institution of Civil Engineers-Waste and Resource Management, 2024; 177(3): 100–113. https://doi.org/10.1680/jwarm.22.00012