



OPEN ACCESS

EDITED BY

Zhen Chen,
University of Strathclyde, United Kingdom

REVIEWED BY

Apurva Pamidimukkala,
University of Texas at Arlington, United States
Abinash Sahoo,
Odisha University of Technology and
Research, India

*CORRESPONDENCE

Yaser Gamil,
✉ yaser.gamil@ltu.se
Hassanean S. H. Jassim,
✉ hassanean.salam.eng@uobabylon.edu.iq

RECEIVED 21 December 2023

ACCEPTED 24 January 2025

PUBLISHED 17 February 2025

CITATION

Jassim HSH, Hasan MF, Altaee MJ and Gamil Y
(2025) A model for preliminary cost
estimation in buildings construction projects.
Front. Built Environ. 11:1359777.
doi: 10.3389/fbuil.2025.1359777

COPYRIGHT

© 2025 Jassim, Hasan, Altaee and Gamil. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A model for preliminary cost estimation in buildings construction projects

Hassanean S. H. Jassim ^{1*}, Musaab F. Hasan ²,
Mohammed J. Altaee ³ and Yaser Gamil ^{4*}

¹Faculty of Civil Engineering, College of Engineering, University of Babylon, Hilla, Iraq, ²General Directorate of Education Baghdad Rusafa First, Ministry of Education, Baghdad, Iraq, ³Environmental Research and Studies Center, University of Babylon, Hilla, Iraq, ⁴Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, Luleå, Sweden

Despite recent rapid developments in computer programs used to improve buildings construction sector cost estimation techniques, costing models utilizable in a building construction project's early planning stages remain scarce due to the lack of a simplified, integrated method that can operate with the limited data available. This research aimed to help buildings construction project stakeholders in Iraq to improve the accuracy of preliminary buildings project cost estimations based on the initial information available and historical buildings project data to evaluate a future project's feasibility. A literature review of topics related to preliminary cost estimation, particularly buildings projects and factors affecting their cost, informed development of an integrated mathematical model based on a support vector machine to facilitate preliminary cost estimation during early buildings construction project planning utilizing relevant factors from both old and new projects (i.e., total floor area of buildings, construction duration, total number of floors, average floor height, location index, project quality standards, project complexity, and facilities provision). To enhance accuracy, the final cost forecasting model was modified using an inflation rate impact scenario to account for any future economic effects on anticipated building costs. This study concluded that the dominant factors behind cost variations between similar projects over different periods, and thus the primary factors for building cost estimation, are building area, average floor height, and number of floors. Thus, this study contributes to enhancing best practice in cost estimating for building construction projects in the pre-design, early planning phase, enabling decision makers to plan alternative options to enhance decisions on feasibility within the project's constraints. Novelty arises from the features of the input factors used for the developed model to determine the preliminary budget required in the early planning stage taking into account construction sector inflation values for the following period.

KEYWORDS

buildings construction, early planning stage, preliminary cost, SVM model, influence factors

1 Introduction

The construction industry in developing countries is considered one of the national economy's main drivers for creating wealth and improving life quality in society through the provision of appropriate social and economic living conditions (Okereke, 2019). Buildings construction projects are an inherent part of the construction industry sector, and they are dependent on cost and time, with performance quality standards based on specific indicators (Rezaian, 2011; Shehatto, 2013; Tayeh et al., 2018). In particular, the cost forecasting for the construction of buildings project before detailed design preparations is crucial for project stakeholders (Akinsiku et al., 2011; Alumbuğu et al., 2014). This is because, for most developing countries, a key part of the construction industry sector's funding relies on resources from national budgets (Adeagbo, 2014). Furthermore, the primary function of approximate and preliminary cost estimating is to reveal whether a construction project is economically feasible, and whether budget support is required or not, which is essential for decision making regarding proceeding with future stages of a project (Sanni-Anibire et al., 2021; El-Sawalhi and Shehatto, 2014; Cheng et al., 2010). Thus, such costing methods, as an inherent part of a feasibility study in the early planning stage of a project, are a cornerstone for early understanding of the construction cost and for decision making (Mahamid et al., 2014; Liu et al., 2013). Moreover, cost estimation has a major effect on the construction project's degree of success, especially during the upstream stage of the project (Kim et al., 2004; Ahuja et al., 1994), to an extent that might even prevent the awarding of a project (Aibinu et al., 2011). The accuracy of the early cost estimation is, therefore, vital, but in construction projects it is largely dependent on the quality of historical cost data available and the level of professional expertise of cost estimators (Enshassi et al., 2013).

The various clients for buildings construction projects have a primary concern regarding initial project costs (Cunningham, 2015). Consequently, they need to understand their potential financial obligations before comprehensive design work is carried out (Mahamid and Bruland, 2010). Furthermore, the fact that preliminary cost estimation has a degree of high influence on the performance of both the client and the contractor of a project due to problems arising from limited budgets and resulting financial difficulties (Leung et al., 2008; Ling and Boo, 2001) has been acknowledged (Babalola and Adesanya, 2009). Therefore, the first challenge to satisfying clients' needs is achieving a more realistic and precise cost estimate for projects despite a lack of information and data especially in the preliminary stage of a project before detailed design work take place (Enshassi et al., 2013; Jarkas et al., 2014; Kadiri, 2014; Yu and Skibniewski, 2010). Thus, preliminary cost estimation should be built on a number of essential elements that would eliminate incorrect overestimates of these costs (Dagostino and Peterson, 2011), despite the great uncertainty that can surround construction projects (Alumbuğu et al., 2014).

Several studies have adopted different techniques to develop early cost estimation models that could maximize utilization of the limited construction project information at an early stage (Cheng et al., 2010). However, there are still a number of limitations with existing approaches to cost estimation in the early planning phase of buildings construction projects, especially given the need to make a number of assumptions regarding the construction project

costs and development (Dysert, 2007). Moreover, these models can be difficult to be apply and are unintegrated, so they are not easy to use due to the fact they cannot account for variations in local practices across different regions when making estimations for work involved in realizing a project (Choon and Ali, 2008). Furthermore, in this context, traditional methods of project cost forecasting, such as expert systems for estimating and unit cost or cost build-up, cannot be applied professionally (Arafa and Alqedra, 2011). Thus, there is a need for a systematic method for estimating the preliminary cost of construction projects based on the limited information and data provided in the early phase of a project (Mahamid and Bruland, 2010; Arafa and Alqedra, 2011; Meharie et al., 2019), so that it can play an important role in evaluating project feasibility and consider the details of the scope of the construction work (Okereke, 2019).

In the context of the Iraqi buildings construction industry, the matter of preliminary cost estimation plays a significant role for decision makers in their early considerations about future projects because the country's unstable national income and economic policies mainly depend on annual budgeting policies for investment in the buildings construction sector. Therefore, an important question for practitioners in the building construction sector to focus on is: What are the ways to support and facilitate the decision-making process for buildings owners regarding the project budget required in the early planning stage of a building construction project? This economic restriction, among other factors relevant to construction projects, demonstrates a crucial need for an applicable computer program-based model to facilitate early decision making regarding whether or not to proceed with a new construction project. In order to address this objective, an integrated mathematical model based on a support vector machine presented as a computer application with a graphical user interface (GUI) was developed to describe and facilitate preliminary costing for building projects during the early planning stage as a function of a project's quantitative and qualitative characteristics based on relevant factors from both old and new projects. Furthermore, the model includes a consideration of inflation rates to adjust for future economic variability since not many comprehensive studies of this factor are employed in preliminary cost estimating.

To describe this development process, first the important relevant works are reviewed in Section 2. The integrated research methodology, the factors selected and the model developed using an SVM algorithm hypothesis are explained in Section 3. In Section 4, the results are presented, and the proposed mathematical model and application is presented in detail regarding the impact of different inputs factors. Then, Section 5 contains discussions related to the SVM technique's validity, the verification of the developed model, how the input variables can affect the model's output, together with considerations on managerial implications. Finally, conclusions are drawn and recommendations made in Section 6.

2 Research background

Any project into construction sector is initiated in preliminary cost estimations in order to decide on the project's feasibility and the approximate budget required for it. Although this process is a key factor in decisions regarding future projects in developed countries

(Mahamid and Bruland, 2010), understanding of the issue is still in infancy due to data and information restrictions that only provide a minimum level of details in the early stage of a construction project. Over recent decades, various studies have been undertaken to overcome the difficulties of preliminary cost estimation in the early phases of the many types of buildings construction projects. While some of the resulting methodological facilitate the outputs of more recent studies, they also raise comparative difficulties between the studies due to a lack of conceptual consistency (Gunner and Skitmore, 1999). Typically, before detailed designs are drawn up, costs are forecast by using unit price estimating techniques, such as cost per unit, or cost per unit area relevant to the historical costs in comparable past projects (Cunningham, 2015). However, an investigation of such traditional cost estimation methodologies shows that the unit costs of items might be not easily applicable in a future project due to changes related to the marketplace's economy (Yu and Skibniewski, 2010). In this context, the preliminary cost estimating model needs to consider variances in the marketplace's economy as one of the necessary elements to develop new methodologies or models.

In fact, few of the studies include intrinsic information reflecting the elements that contribute to costing buildings construction projects (Akintoye, 2000). It is, however, the choice of suitable input factors for estimating models that is a primary procedure, playing a vital role in developing a powerful model by presenting precise prediction values (Mahalakshmi and Rajasekaran, 2018; Gransberg et al., 2017; Kim and Shim, 2014). Preliminary cost estimating is mainly undertaken based on the principle parameters known from earlier projects, without detailed designs available for the new project (Hegazy, 2013). For instance, the models proposed by Hakami and Hassan (2019); Latief et al. (2013) for cost prediction in the early design stage might be difficult to apply in the early planning stage of buildings projects where there is limited information (e.g., foundations type, interior details, and electrical and mechanical systems) for planning building construction projects. Whatever the shortage of information, the preliminary estimating is guided by outline plans, areas, and elevations before provision of a quick estimation for a project in a costing study (David et al., 2002). Overall, most of the preliminary cost estimating work relies on using the gross areas or volumes of the intended building project (Choon and Ali, 2008), whereas there may be a need to employ particular restrictions and definitions specific to the project's individual characteristics and specifications so as to arrive at a better cost estimation.

The number and type of significant factors can vary from project to project and from country to country. For instance, using actual data from 590 building projects constructed in South Korea, location, number of floors, building coverage ratio, floor area ratio, total area, project type, and duration were recognized as important for cost estimating (Kim and Shim, 2014). In Seoul, South Korea, the variable included in the datasets used for preliminary cost estimates in 498 residential buildings projects were the total floor area, number of storeys, total units, duration, type of roof, type of foundation, type of basement, grades of finish (Kim et al., 2005). In China, the foundation type, structural type, floor area, number of basement floors, and number of total floors were the elements adopted in the preliminary cost estimating of building construction projects based on datasets from 110 buildings projects distributed across

twenty-five cities (Yu and Skibniewski, 2010). Also in China, the datasets from only twenty-four buildings construction projects were collected from the website of company called "Glodon" to develop a cost prediction model based on ten input factors that were identified as having the most impact (Shutian et al., 2017). In Taiwan, site area conditions, geological properties of the construction site, earthquake impact, planned householder numbers, total area, floors above basement level, basement level floors, decoration category, and facility category were used to develop an early stage cost estimating model based on analyzing datasets from 29 housing construction projects (Cheng and Wu, 2005). In Ireland, the building size, shape, elevations, total height, type, level of specifications, facilities offered, the market situation, inflation rate, venue characteristics, quality level, architectural footprint and planning efficiency were dominant factors in the early cost estimating (Cunningham, 2013). In Turkey, the approximate cost, total construction area, number of floors, building elevations, and contract value were the main inputs for cost estimating based on datasets acquired from three cities covering a total of 232 public buildings projects (Bayram et al., 2016). In the United States, Sönmez (Sönmez, 2004) suggested the time index and total building area were more important factors than others on the project cost per unit area based on datasets analyzed from 30 specific types of construction projects (i.e., "retirement housing community") executed in 14 different states. Cho et al. (2013) used the type of financing, floor area, total building area, number of classrooms, number of basement levels, and number of upper floors to estimate cost in the early stages of elementary school projects using a total of 96 historical data sets. Although these techniques are significant, and models have been developed to fill a part of the existing knowledge gap, they are not capable of being adopted for all cases globally. In addition, there are still some important factors for developing an integrated model that are not being considered, such as the inflation rate and location change, to account for fluctuating construction costs for project items when estimating costs for a future project.

Other techniques have employed a different approach to improving the preliminary cost estimating procedures, using statistical methods/regression analysis (RA), simulation, support vector machines (SVM), artificial neural networks (ANN) and, recently, case based reasoning (CBR). For instance, the boosting approach was applied to solve the regression problem that was proposed as a model to predict the construction cost of school buildings in Korea based on a given budget without considering the number of floors (Shin, 2015). RA was adopted as the first method for conceptual cost estimating (Alshamrani, 2017; Ha and Lee, 2012; Nsofor, 2006; Li et al., 2005; Trost and Oberlender, 2003). Simulation has also been applied efficiently to enhance estimating with mentioned issues (Chou et al., 2009; Yang, 2005). Case based reasoning (CBR) works on the principle that how problems were solved in past projects can be a guide to solutions for similar issues early stage cost estimation in future projects. It can be applied using various indicators, either individually and/or in conjunction with other techniques (Kim and Shim, 2014; Leśniak and Zima, 2018; Kim and Kim, 2010). SVM has been a widely utilized technique applied in this field to estimate cost of various types of projects (Cheng and Wu, 2005; Chandanshive and Kambekar, 2021; Chen et al., 2019; Juszczak, 2018; An et al., 2007). Finally, with the development of modern computerized programs to

improve the performance of machine learning applications, ANN has been adopted among these cost estimation techniques (Arafa and Alqedra, 2011). ANN has been used in the preliminary cost estimating of public educational buildings projects (Son and Kim, 2006), apartment housing construction projects (Park et al., 2002), and it has also been used in collaboration with the evolutionary fuzzy hybrid (Cheng et al., 2010), and genetic algorithms (Kim et al., 2005). All these techniques have been used to develop various models for cost estimating in the early stages of construction projects, either applied separately or by combining two or more techniques; however, these methods have still proved to be inadequate for adoption across all regions. For instance, a hybrid model was proposed to predict the cost of residential buildings in the early stage based on the number of floors, floor area, and type of external and internal finishing (Badawy, 2020). However, despite the limited feasibility of adopting this model, the input parameters of the proposed model could not provide a comprehensive view of the characteristics of building projects in the studied region. Certainly, the building information model (BIM) recently applied to predict cost based on the architectural information could be utilized in the early planning stage of a construction project (Yang et al., 2022). However, a major effort is still required to push participants to adopt this model because the application of BIM is still a minority practice among practitioners in this sector.

Therefore, there is still a need for a comprehensive model, one that considers the influences of several factors, to help decision makers to identify the cost of buildings construction projects at an early stage. To meet this need, this research develops a cost estimation application based on SVM and collected data for completed building construction projects (institutional and universities) in Iraq. As this cost estimate is required early in the project, consideration was given to the fact that the input data for the required mathematical model should be easily available from just the limited scope and definition of projects. Furthermore, based on the author's research of the existing literature and other relevant investigations, this is the first study to propose a comprehensive model for the preliminary cost estimating of building construction projects, especially in Iraq, that incorporates essential construction elements in the forecast.

3 Research methodology

In order to achieve the research objective, this section proposes a methodology organized into three interconnected stages, with the outputs of the upper stage being used as input for the next stage, as shown in Figure 1. The first stage aims to identify elements and factors affecting the preliminary costs of buildings construction projects based on an investigation of the previous theoretical and practical concepts related to the objective of this research. The output of this stage is a list of the preliminary cost parameters that are critical to determining the cost of building construction projects and that can be provided in the early planning stage of a project. This list serves as the input for initiating the second stage, which is the collection of information and historical data of building construction projects in order to arrange and standardize the final data sets for a proposed model consisting of eight input parameters for each project against the target output, which is preliminary cost.

In addition, inflation rate datasets are used to adjust the model's future cost values. Therefore, the datasets for both the exploratory and quantitative analysis methods employed involved supervised learning models with associated learning algorithms to analyze historical datasets and to perform regression analysis. By using the support vector machines of the Weka Program package (WEKA-3-9), the relationship between the final cost of building construction projects and the independent variables were tested to find the most conclusive mathematical expressions of their regression. This technique was used because it is an appropriate method to develop and upgrade a mathematical model for the study variables that are used later to build a final application that includes other indicators. Most importantly, the final product is presented in form of a GUI application, which is more convenient and practical for construction projects decision makers to use in order to solve problems or allocate budgets. The proposed method takes two sets of inputs: first, the total actual cost of building construction projects and, second, independent elements that generate these costs for each project. The method and procedures for collecting, extracting and processing datasets, selecting factors, developing and presenting the SVM model, and adjusting the final cost for future value are provided and explained in the following subsections of the research methodology.

3.1 Sources of data collection

The research was conducted in accordance with the ethical rules and regulations of the Iraqi Ministry of Higher Education and Scientific Research. Permission to access and collect the required documentary data from the various official consulting engineering offices, companies, and Iraqi Government Offices for building construction projects was obtained, supported by a document issued by the authors' institution declaring their need for such documentary data for the research work. Prior to collecting and reading the data, the data source managers and administrators requested that the researchers ensure that all participants must keep all data and information in an inventory using a coding method, must avoid mentioning the names of projects or buildings, and must also consent in writing to respect the confidentiality of the project contract documents. Moreover, the researchers were also permitted to ask managers or administrators any type of study related questions, especially for old building projects documents where there was a lack of detailed information.

In this context, and with respect to the limited data and information available and, in recognizing the importance of how studies have enhanced this sector and the vital role it plays in future developments, the scope of this research is limited to the examination of construction costs of educational buildings in Iraq. In order to obtain datasets and information with the best quality and accuracy, they were acquired mainly from original sources such as the construction project documents archives of the parties concerned. Thus, data from 90 building construction projects completed during the period 2011 to 2023 were obtained directly from a number of government engineering departments, consultants engineering offices, and construction companies in several governorates (Baghdad, Babylon, Karbala, Al-Najaf, Al-Qadisiyah, Wasit, Al-Muthanna, and Al-Basra) in the middle and

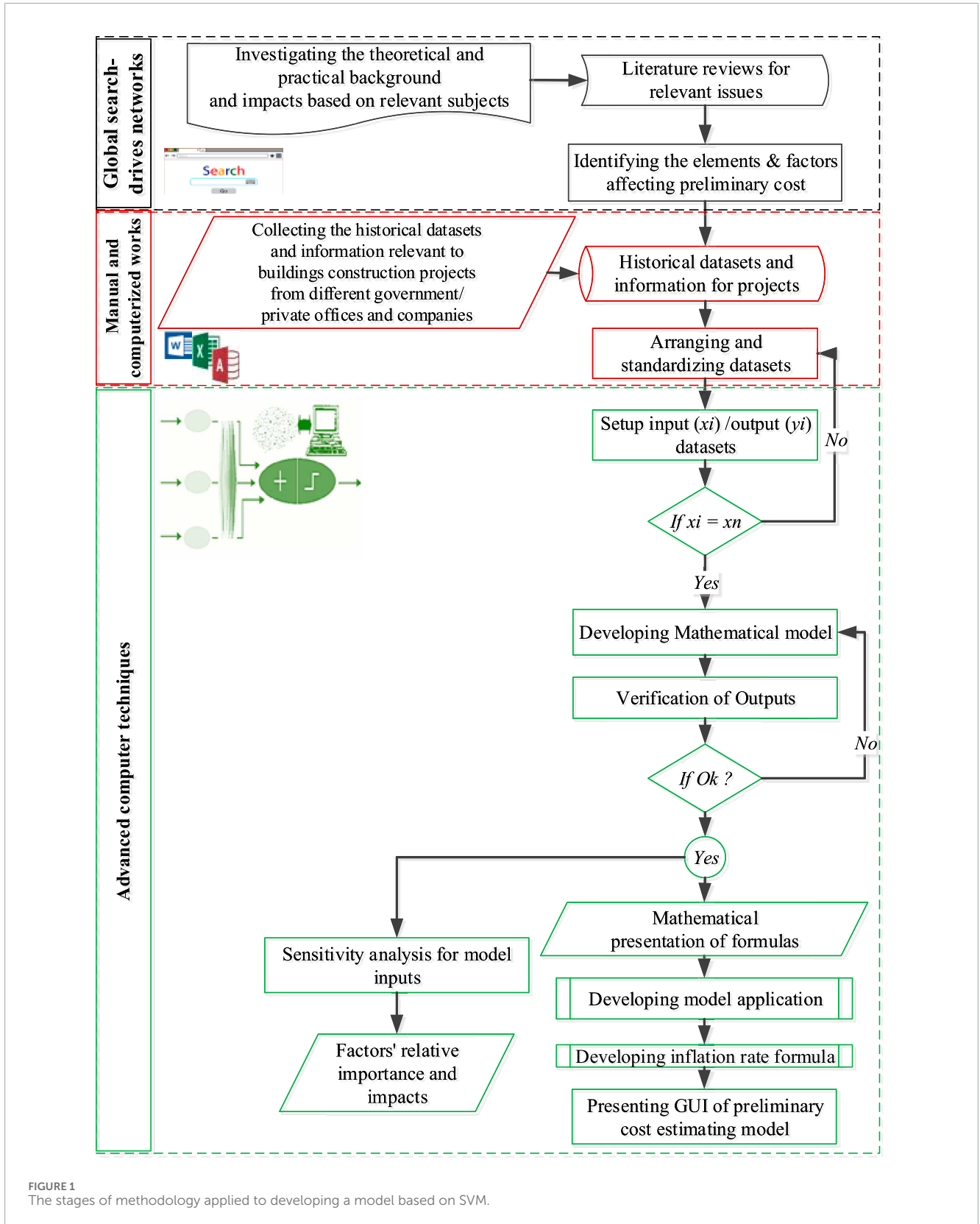


FIGURE 1 The stages of methodology applied to developing a model based on SVM.

south of Iraq. The focus was primarily on educational buildings projects or educational office building projects with costs ranged between 216,700,000 and 4,120,000,000 Iraqi Dinars. The data

include detailed information on the final construction cost of building projects, number of floors, total floor area, average area of similar floors, average area of other floors, average height of

TABLE 1 The characteristics of possible modes of data presentation in scientific publications.

Variables category	Variables abbreviation	Variables description	Variables unit	Variables type
Input	T_{FA}	Total floors area of buildings	m^2	Quantitative values
	D_C	Construction duration	day	Quantitative values
	N_F	Total number of floors	Number	Quantitative values
	A_{HF}	Average height of floors	m	Quantitative values
	L_I	Location index	Scale	Qualitative values
	S_Q	Standard quality of project	Scale	Qualitative values
	P_C	Project complexity	Scale	Qualitative values
	F_P	Facilities provision	Scale	Qualitative values
Output	C_p	The cost of project per m^2	Money (\$)	Quantitative values

similar floors, average height of other floors, total building height, number of storeys above ground, location description, construction specification and requirement, degree of project complexity, and facilities provision.

3.2 Factor selection for conceptual cost estimates

In general, the common goal of early cost predicting in the buildings construction industry is improved planning in work efficiency for the project’s construction phase due to the influence of budget allocation on the project’s progress status, the avoidance of deficits in project cash flow within annual budgets, and the strategic planning of future projects costs. It should be noted that, because of uncertainties and lack sufficient data, it is not possible to perform the entire process for making decisions without the existence of integrated models that can overcome these early stage difficulties. Therefore, developing an integrated model that includes the more critical factors relevant to preliminary cost estimates can help planners and estimators to reduce subjective assumptions and judgments during the estimation process. The essential factors in early cost estimating that were selected to develop the estimation model consisted of a total of eight key independent input factors. Tables 1 and 2 list and describe all 8 most key influencing factors (i.e., Inputs) and the preliminary construction cost (i.e., Target output) for the development of SVM cost prediction models:

To increase users’ understanding of the developed model’s components and provide clear definitions for the selected input and output parameters for the selected model factors shown in Tables 1 and 2, a brief description of the parameters is given as follows: Total floor area of buildings (T_{FA}): the sum of the area of each floor of the building measured from the edges of the outer surface of the outer walls, including the area of floor cantilevers, basements, elevator room-shafts, and all the common spaces in multi-dwelling buildings. Construction duration (D_C): the total construction time

TABLE 2 Description of qualitative factors for the SVM prediction model.

Variables description	Scale	Qualitative values
	Uncrowded area	1
Location index	Normal density area	2
	Crowded area	3
	Standard	1
Standard quality of project	Good	2
	Very good	3
	Non complexity	1
Project complexity	Mid complexity	2
	High complexity	3
	Not provision	1
Facilities provision	Some provision	2
	Full provision	3

in real-work days required to finish a building in order to become ready for delivery and operational use. Total number of floors (N_F): counts the number of floors from ground floor up to the highest floor. Average height of floors (A_{HF}): the average height of the floors for building, calculated as the sum of heights from ground floor to the highest floor, divided by the number of floors. Location index (L_I): refers to a project’s place characteristics distributed in a range from crowded-urban areas to open rural areas, where the impact cost is associated with location of the construction site either in the countryside or in the centers of crowded cities and towns that

might reflect some relative difficulties like the problems of resource access and executing works (Ashworth and Perera, 2015). Work in urban areas is generally more expensive than in rural areas due to higher costs related to access constraints, labor costs, limited space for staff facilities and material storage, as well as the safety and security requirements for a project site (Cunningham, 2013). Standard quality of project (S_Q): relates to resources used in the form, type, and soundness of the construction work such that the buildings meet the quality standards required. Various quality standards may be expected depending the type of project and owner requirements as translated by the designer within the known, default quality standards for the industry. Project complexity (P_C): generally refers to the geometry and design of building projects, categorized in terms of shape complexity, size and detail complexity, and necessary techniques used for construction and assembly, for instance where the specific size required for such items requires mechanical support for construction, which of course has an impact on costs. Facilities provision (F_p): relates to where the owner or client of a project may be able to provide a more competitive price for supplying some of the construction materials to a building company or contractor, for example, as often happens in case of executing construction projects for the government sector, or when an owner also owns a relevant materials factory or other manufacturing site. Actual cost of project (C_p): this relates to the datasets of final cost values for various types of constructed buildings projects, including a building's scope and definition.

3.3 Support vector machine theory and mechanism

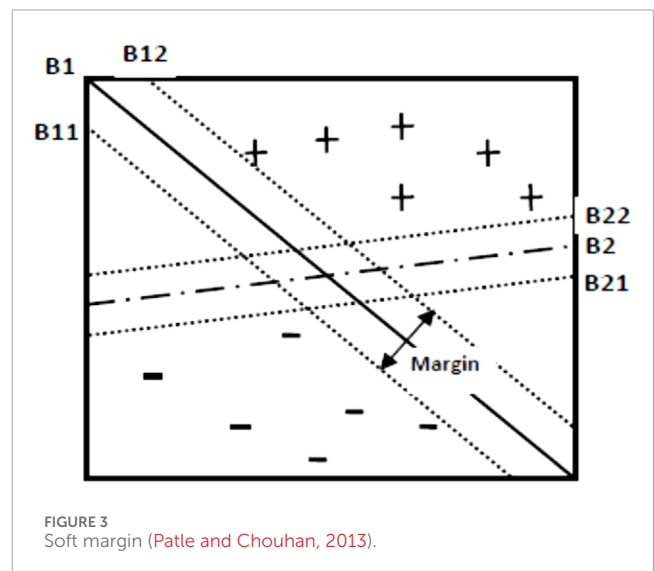
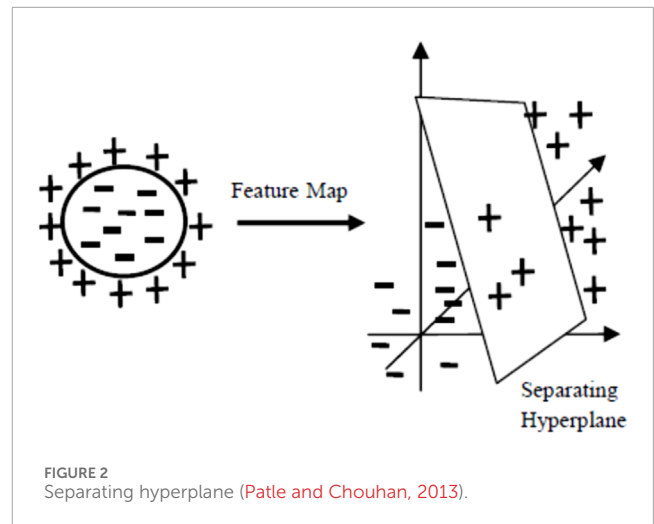
A mathematical structure known as an SVM is a potent modern algorithm for maximizing a particular mathematical function to a specific data group. SVMs are a group of connected supervised learning techniques utilized for regression and classification (Patle and Chouhan, 2013) based on the Vapnik-Chervonenkis theory. The popularization of the model to new data is referred to as regularization, and strong regularization properties exist for SVM. The four foundational ideas of SVM are separating hyperplane, maximum margin hyperplane, soft margin, and kernel function (Gunn, 1998). An SVM performs classification by creating an N-dimensional hyperplane.

3.3.1 Separating hyperplane

Dividing space into halves in three dimensions requires the use of a plane, called the separating hyperplane, which is a straight line. The line that divides the several data group samples is known as a separating hyperplane because it is the comprehensive expression for a rectum line in a high-dimensional area as seen in Figure 2 (Patle and Chouhan, 2013).

3.3.2 Maximum margin hyperplane

Choosing a middle line is the maximum-margin hyperplane. Put another way, one can choose a line that divides the two categories while maintaining the greatest possible dimension from any provided term profiles (Chang and Lin, 2011).



3.3.3 Soft margin

SVM can deal with data errors by permitting some anomalous term profiles to lie in the 'incorrect direction' of the separating hyperplane. A 'soft margin' must be added to the SVM algorithm to handle situations like the one in Figure 3. Generally, this permits several data points to exceed the separating hyperplane's margin without impacting the outcome. The hyperplanes in Figure 3 are B1 and B2. B1 maximizes the margin, making it superior to B2 (Patle and Chouhan, 2013).

Given a training group of N data points $\{(x_i, y_i)\}$ $N_i = 1$, with input data x_i and symmetric binary class labels $y_i \in \{-1, +1\}$, the SVM classifier, depending on Vapnik's original formulation, meets the following conditions (Cortes and Vapnik, 1995):

The SVM classifier, by Vapnik's original formulation (Equations 1–3), meets the following requirements, specifically a training group of N data points $\{(x_i, y_i)\}$ $N_i = 1$, input data x_i and symmetric binary class labels $y_i \in \{-1, +1\}$:

$$W^T \varphi(X_i) + b \geq +1, \text{ if } Y_i = +1 \tag{1}$$

TABLE 3 Behavior of the model concerning data division.

Split of data%			RMSE training	RMSE testing	RMSE validation	Correlation coefficient(R)	Determination Coefficient(R ²)
Training	Testing	Validation					
85	5	10	0.0829	0.1058	0.5046	85	5
80	5	15	0.0749	0.1153	0.396	80	5
80	10	10	0.0773	0.1023	0.375	80	10
75	5	20	0.0749	0.1153	0.308	75	5
75	10	15	0.0895	0.1542	0.472	75	10
70	5	25	0.0801	0.1682	0.526	70	5
70	10	20	0.0952	0.1425	0.348	70	10
70	15	15	0.0951	0.1577	0.490	70	15
65	5	30	0.0998	0.1782	0.456	65	5
65	10	25	0.0982	0.1282	0.522	65	10
65	15	20	0.0845	0.1356	0.391	65	15
60	5	35	0.0658	0.1852	0.432	60	5
60	10	30	0.0998	0.1782	0.471	60	10
60	15	25	0.0928	0.1488	0.505	60	15
60	20	20	0.0942	0.1825	0.475	60	20

Optimal data size partitioning for all projects dataset.

TABLE 4 The min-max boundary values of parameters for the SVM developed model.

Items	T _{FA}	D _C	N _F	A _{HF}	L _I	S _Q	P _C	F _P
Min	100	180	1	2.83	1	1	1	1
Max	3,500	950	3	5	3	3	3	3

$$W^T \varphi(X_i) + b \leq -1, \text{ if } Y_i = -1 \tag{2}$$

Which is equivalent to:

$$Y_i [W^T \varphi(X_i) + b] \geq 1, i = 1, \dots, N \tag{3}$$

Where: W^T is the weight vector, “b” is the constant threshold, and the high-dimensional feature spaces (x_i) are represented by φ(X_i).

3.3.4 Kernel function

The kernel function enables the SVM to carry out a two-dimensional classification on a group of data that was primarily only one dimension. Using a kernel function, data is typically projected from a lower to a higher-dimensional space. Here, Equations 4–6 give some examples of kernel equations.

- Polynomial kernel:

$$k(x, y) = (x \cdot y + 1)^p \tag{4}$$

- Radial basis function kernel:

$$k(x, y) = \exp(-\gamma \|x - y\|^2) \tag{5}$$

- Sigmoid kernel:

$$k(x, y) = \tanh(kx \cdot y - \delta) \tag{6}$$

Where: p, γ and δ are the kernel parameters.

3.4 Data extracting, pre-processing, dividing and scaling

The total of 90 sample datasets consists of the bidding construction cost dataset of eight crucial independent factors governing the project cost as mentioned above in Section 3.1. The dataset and imperative basic information of building construction projects in Iraq were gathered from the different sources as mentioned, from which a uniform excel spreadsheet was developed for the arrangement of the database to avoid duplicating and mistakes of reading all input-output datasets when importing

TABLE 5 Statistical analysis for input and output of the adopted preliminary cost estimation model.

Data group	Statistical Parameters	Input								Output
		T_{FA}	C_D	N_F	A_{HF}	L_I	S_Q	P_C	F_P	$LN(C_p)$
Training n = 72	Range	3,075.000	770.000	2.000	2.170	2.000	2.000	2.000	2.000	2.915
	Minimum	300.000	180.000	1.000	2.830	1.000	1.000	1.000	1.000	18.725
	Maximum	3,375.000	950.000	3.000	5.000	3.000	3.000	3.000	3.000	21.640
	Mean	1,212.417	436.264	1.639	3.548	2.125	2.111	1.986	1.403	20.310
	Std. Deviation	760.775	168.360	0.737	0.440	0.749	0.723	0.778	0.664	0.709
Testing n = 9	Range	2,900.000	495.000	2.000	1.500	2.000	1.000	2.000	1.000	2.688
	Minimum	100.000	280.000	1.000	3.000	1.000	2.000	1.000	1.000	18.873
	Maximum	3,000.000	775.000	3.000	4.500	3.000	3.000	3.000	2.000	21.561
	Mean	958.222	464.444	2.111	3.472	2.556	2.556	2.333	1.111	20.220
	Std. Deviation	898.320	211.445	0.782	0.461	0.726	0.527	0.707	0.333	0.862
Validation n = 9	Range	1,250.000	315.000	2.000	1.080	2.000	1.000	2.000	1.000	0.973
	Minimum	750.000	290.000	1.000	3.420	1.000	2.000	1.000	1.000	20.333
	Maximum	2000.000	605.000	3.000	4.500	3.000	3.000	3.000	2.000	21.306
	Mean	1,221.111	418.889	2.667	3.918	2.000	2.333	2.444	1.111	20.802
	Std. Deviation	435.221	104.844	0.707	0.396	0.866	0.500	0.726	0.333	0.292

TABLE 6 Effects of the kernel function on SVM model developed.

Kernel type	MAE	RMSE	Correlation coefficient (%)
Normalized poly-kernel	0.3535	0.4739	74.04
Poly-kernel	0.2941	0.3817	84.05
RBF kernel	0.374	0.4698	77.45

database to the work environment of a machine learning technique (SVM). In order to build the SVM model, subsets of the available data were also created. The best number of statistical consistency groups and default Weka Program parameters were chosen using the statistical data division method, which offers one of the best (training, testing, and validation) alternatives. SPSS (version 23) was used to perform these division procedures to ensure the best random data division, and the subsets were examined using the Weka program. In fact, there is no specific rule for how best to divide data into training and validation sets when adopting machine learning techniques, so a rule of thumb approach based on experience was adopted here to produce the highest performance (i.e., minimum error in model outcoming) in processing the datasets. Here, after preprocessing and arranging the datasets, the subsample random

selections of data was achieved by a computer coding that read the data from project 1 to project 90 as series, and then selecting the first eight subsets read for training, and the next two subsets for testing and validating respectively, and this partitioning process continued until the datasets were completed. The optimal size of data division for all dataset of projects was determined to be (80%) of the total projects datasets for training, (10%) for the validation dataset, and (10%) for the testing dataset, as shown in Table 3. This selection depends on the minimum error produced for testing subsets and the greatest correlation coefficients (R) and determination coefficient (R^2) during data processing to figure out their behavior. According to this rule, datasets from (72) projects were used for the training stage, (9) for the validation stage, and the other (9) for the testing and verification stage. It is obvious that, when using more or less of a number of projects in the training subsets shown in Table 3, the Root Mean Squared Error (RMSE) of the testing set increases and (R) and (R^2) decrease because the model depends on the number of data used in the training group; that is, the greater the number of training data, the more accurate the model becomes. The model was developed based on the real and normalized scales for the input and output data. The input-output datasets of the selected parameters with boundary values for each one is shown in Table 4. Preparing and scaling all the input-output values of datasets was used to overcome problems of overfitting and outliers during the training and testing of the developed model in order to normalize

TABLE 7 Effect of the parameter C in the SVM model performance of preliminary cost estimation.

Parameter(C)	MAE	RMSE	Coefficient correlation (%)
1	0.2941	0.3817	84.05
2	0.294	0.3809	84.1
3	0.2937	0.3812	84.09
4	0.2937	0.3808	84.12
5	0.2938	0.3808	84.12
6	0.2938	0.3811	84.1
7	0.2939	0.3807	84.13
8	0.2938	0.3807	84.14
9	0.2938	0.3806	84.14
10	0.2938	0.3808	84.13

the data sets utilized in the model so that they fall in the specified interval [0, 1] that is produced by Equation 7, so as to calculate scaled values for all variables with a minimum and maximum of (xmin/xmax):

$$\text{Scaling of input value} = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (7)$$

In addition, the analysis of some statistical parameters was conducted to ensure that the data split into training, testing, and validation groups demonstrated the same statistical population. Table 5 shows these parameters with the mean, standard deviation, minimum value, maximum value, and ranges. The training, testing, and validation groups are mostly statistically proportionate.

3.5 Development of the SVM model

The kernel function is used to develop the SVM equations. Table 6 shows the best kernel function model according to certain values of the equation parameters and this was checked by the R test and RMSE. The poly kernel function was selected in this model as having the minimum RMSE (0.3817) and maximum R (84.05%). On the other hand, Table 7 shows the best value parameter C which is (9) with the minimum RMSE (0.2938) and maximum R (84.14%). In addition, the best value parameter Epsilon is shown in Table 8, which is (0.04) with the minimum RMSE (0.2982) and maximum R (84.86%).

3.6 Presentation of SVM prediction equation

Depending on the aforementioned analysis, the forecasting equation for the total cost of the project based on the SVM algorithm

for the developed model is shown in Equations 8–10, which show the optimal connection weights value for the SVM equation as presented in Table 9.

$$C_{P_{nor}} = 0.1914 + 0.7041T_{FA} + 0.0254D_C + 0.0885N_F - 0.0232A_{HF} + 0.0113L_I + 0.0958S_Q + 0.0935P_C - 0.0920F_P \quad (8)$$

$$C_{P_{act}} = \text{InvLn}[C_{P_{nor}} \times \text{range} + \text{min}] \quad (9)$$

$$C_{P_{act}} = \text{InvLn}[C_{P_{nor}} \times 2.915 + 18.725] \quad (10)$$

Before using Equation 8, it must be noted that all input variables must be changed to values between (0–1) because Equation 8 depends on Equation 7. To return the output values to their actual value, Equation 10 is used.

3.7 Cost adjustment for inflation rate

Cost escalation is an inherent feature of the construction industry, whether the percentage change value is small or large. The inflation rate reflects this changeable cost value for building similar items in the construction sector in different future time periods by using a specific base period/year datum. The inflation rate can be defined in economic construction terms as the increase in the construction cost amount due to the passage of time when compared to the same quantity of work as previously done. Thus, it can be calculated as the percentage rate of change in cost over a specific period, usually measured per year (Mankiw, 2012). To address inflation rate trends in the construction sector, the data for construction materials and labor costs for construction firms with similar projects were analyzed in order to generate a cost indices history for tracking inflation values. Based on this concept, here the inflation rate is estimated for the reinforced concrete cost data analysis, which comprises data on costs for a wide specification range of construction materials (i.e., cement, sand, gravel, reinforcement bars), frames, skilled/un-skilled labor, technicians and engineering staff, as shown in Table 10, as an example of their statistical analysis indicators. Furthermore, the cost data used to calculate the inflation rate were extracted from the same buildings projects data that was applied in order to develop the SVM model. Thus, the preliminary cost prediction for buildings can be adjusted by formulating an Equation 11 that is based on the principle rule of inflation rates taking 2011 as a base year, and the final adjusted prediction cost is arrived at by formulating Equation 12:

$$i_{rate} = 0.0043CON - 0.9059 \quad (11)$$

$$C_{P_{adjusted}} = C_{P_{act}} \times \left(\frac{i_{rate}}{100} + 1 \right) \quad (12)$$

Where: i_{rate} is the inflation rate value for the target year of building cost estimations, CON is the average reinforced concrete cost per cubic meter for the target year of building cost estimations, $C_{P_{act}}$ is the prediction cost of Equation 10, and $C_{P_{adjusted}}$ is a final adjustment of the cost prediction.

TABLE 8 Effect of the parameter Epsilon in the SVM model performance of preliminary cost estimation.

Parameter Epsilon	MAE	RMSE	Coefficient correlation (%)
0.001	0.2938	0.3806	84.14
0.002	0.2939	0.3805	84.15
0.003	0.0528	0.0746	84.15
0.004	0.2944	0.3792	84.28
0.005	0.2946	0.3784	84.36
0.006	0.2949	0.3772	84.45
0.007	0.2951	0.3766	84.51
0.008	0.295	0.3778	84.4
0.009	0.2954	0.3785	84.34
0.01	0.2958	0.3781	84.37
0.02	0.2976	0.3741	84.69
0.03	0.299	0.3731	84.74
0.04	0.2982	0.3717	84.86
0.05	0.3015	0.3717	84.86

Specific value parameter Epsilon.

TABLE 9 Weighting factors values for developed SVM model.

Items	$Bias$	T_{FA}	D_C	N_F	A_{HF}	L_I	S_Q	P_C	F_P
Weights	0.1914	0.7041	0.0254	0.0885	-0.0232	0.0113	0.0958	0.0935	-0.0920

3.8 Graphical user interface for the SVM predicting model

In order to make the SVM preliminary cost predicting model simple to use for cost estimators and those planning buildings construction projects, a graphical user interface (GUI) was created by editing the required code within the MATLAB program work environment interface. Figure 4 shows the GUI for the SVM model before and after the parameters' values have been entered into the application.

4 Results

Based on the aforementioned inputs and outputs for defined parameters that were used to develop a model in each iteration, a number of SVM models were formed, and one SVM model with the smallest assessment error for R and RMSE values was adopted. The summary analysis of descriptive statistics of the developed SVM model during the training, testing, and validating stages are shown in subsections below. In general, within systematic

processing for the training and verifying stages of all SVM models, the Poly-kernel function is based on the high value of Epsilon for the support vector regression model (Son et al., 2012) as it presents the SVM prediction model as being better than other proposed models.

4.1 Model accuracy, validity, and evaluation

Testing a model's accuracy and validity is one of the key steps in model development. It entails using test or validation data to test and assess the developed model. Some representative data from the intended population that were not used in the model's development should be included in the validation data. Equations 8–10 are used to project the project's anticipated total cost Table 11 displays the results. The residual values displayed in this table make it clear that the model performs well.

The determination coefficient is used by plotting the predicted against actual validation values as shown in Figure 5 using the natural logarithm (Ln). This is done to determine the validity of

TABLE 10 Statistical analysis of data sets for the adopted inflation rate equation.

Statistical parameters	Gravel (IQD/m ³)	Sand (IQD/m ³)	Cement (IQD/ton)	Reinforcemer bars (IQD/ton)	Frames (IQD/m ³)	Labor cost (IQD/m ³)	Tech. And Eng. Staff (IQD/m ³)	Concrete (IQD/m ³)	Inflation rate
Range	13,000	9,000	70,000	550,000	35,000	30,000	10,000	185,000	0.792621
Minimum	25,000	19,000	105,000	875,000	20,000	15,000	10,000	210,000	0.000000
Maximum	38,000	28,000	175,000	1,425,000	55,000	45,000	20,000	395,000	0.792621
Statistical index	Mean	23,111	132,189	1,168,889	36,056	2,676	16,557	259,541	0.210000
	Std. Deviation	2,857	23,382	194,997	12,051	10,121	3,123	40,673	0.174667
	Coefficient of variation	12.787	17.689	16.682	33.423	37.827	18.867	15.671	83.046007

the SVM model's total predicted project cost. Plotting the predicted versus the actual values for thirteen extra data points that were not used in this model is shown in Figure 6. These Figures make it abundantly clear that SVM techniques can be generalized to work with this data sample.

4.2 Model evaluation

The average accuracy percentage (AA = 98.483%), mean absolute percentage error (MAPE = 1.517), root mean square error (RMSE = 0.3309), and mean percentage error (MPE = 1.517) produced by the SVM model is among the statistical measures used to assess the performance of prediction models. Here, R = 0.937, R² = 0.878. Besides this, overall verification for the developed SVM model was done by emphasizing the variation in the actual and predicted values for preliminary buildings cost for different buildings projects that were built in different building construction site conditions. Figure 7 shows good agreement of overall buildings data between the actual and predicted values for the SVM model developed.

4.3 Relative importance of model parameters

A sensitivity analysis approach looks at how small changes to input values (i.e., an independent variable) affect output value (i.e., a dependent variable) for any model. This is achieved through model simulation, which runs a model through unlikely, extreme, and real-range cases for each input parameter in order to determine how much the final output value changes, and whether the model outcomes can survive all scenarios over a range of real datasets used to build up a model. In other words, the analysis process was achieved through a variety of changes one by one (i.e., for each parameter) to better understand how each independent variable impacts on the target outcome, which can be called the relative importance of each independent variable among other variables. In this study, investigating the relative importance of the various input parameters in the developed SVM model is significant for an understanding of the effects of each parameter on the model's targeted output. Furthermore, knowledge of the impact of each input parameter on the final output of the designed SVM model is an essential procedure when choosing or comparing different options for parameters used as an aid to making an optimal choice within an available arrange of design values in the upstream stage of the building's projects. For instance, if we have more than one option for the detailed design of a building related to decisions about floor height, degree of building design complexity, and the opportunity to have facilities support during the construction process, the model can compare among various design options in terms of operational characteristics so as to mitigate the total cost of the building project. In this study, the percentage values of importance for each parameter was based on the partitioning weights method to assess the effects of different input parameters on the target output. This method has been used here to determine the impact of the aforementioned collected data from completed building projects in order to assess the relative importance of the various input parameters used in

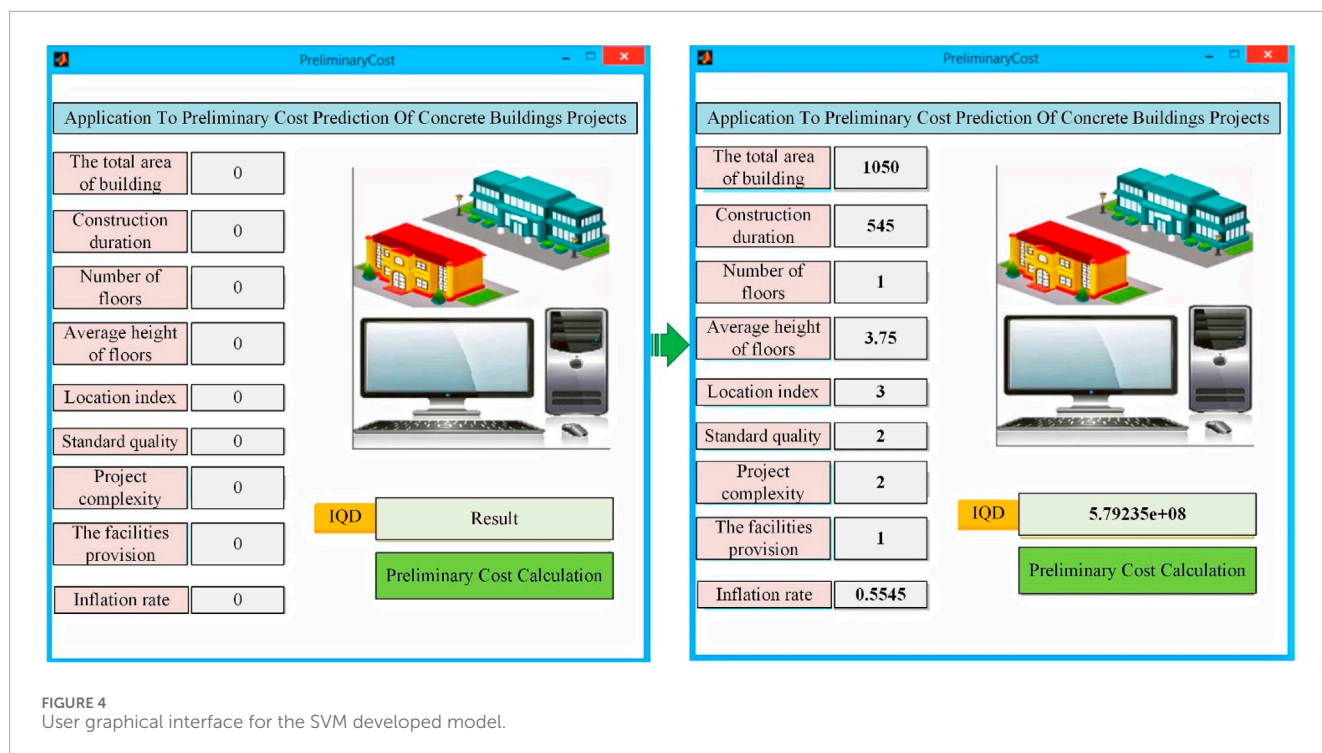


FIGURE 4 User graphical interface for the SVM developed model.

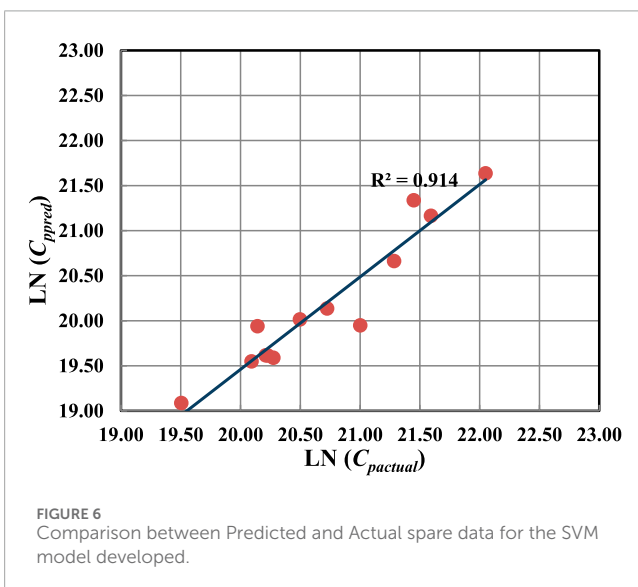
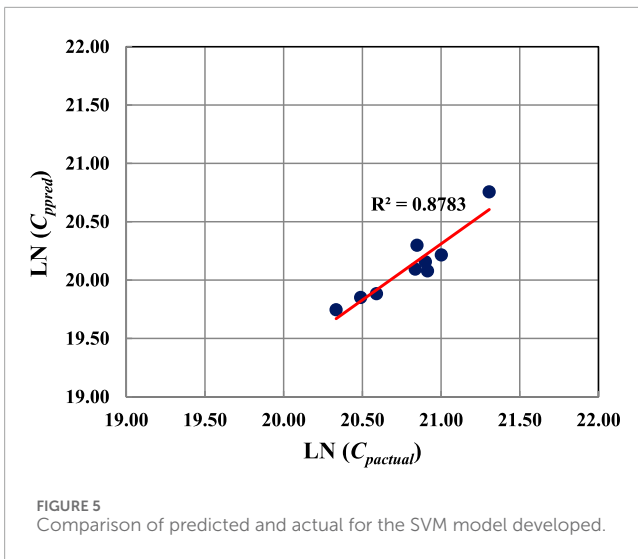
TABLE 11 Verification of the SVM model developed.

Case no.	T_{FA}	D_C	N_F	A_{HF}	L_I	S_Q	P_C	F_P	$LN(C_{Pact})$	$LN(C_{PPred})$	Residual
82	1,000	500	3	4.17	3	3	3	1	20.215	20.582	0.420
83	825	350	3	4.33	3	3	3	1	20.079	20.446	0.468
84	915	410	3	3.42	2	2	3	2	19.883	20.250	0.340
85	1,350	510	3	3.67	1	2	2	1	20.157	20.523	0.377
86	2000	415	2	4.25	1	2	1	1	20.298	20.665	0.182
87	1800	605	3	3.75	2	3	3	1	20.756	21.123	0.184
88	1,300	290	1	4.5	3	2	2	1	19.851	20.218	0.272
89	1,050	290	3	3.5	2	2	3	1	20.093	20.460	0.377
90	750	400	3	3.67	1	2	2	1	19.746	20.112	0.221

developing the SVM model and their effect on the total construction cost within various operational buildings' characteristics. As shown in Figure 8, the most important parameter influencing buildings cost in the proposed SVM model was found to be the total floor area of buildings (T_{FA}), at 27%. Of second highest importance was average height of floors (H_F) (18%), followed by the total number of floors (N_F) (17%), the project complexity (P_C) (13.5%), facilities provision (F_P) (8%), the construction duration (D_C) (7%), the location index (L_I) (5%), and the quality standard required for the project (S_Q) (4.5%). This result is shown in Figure 8, where the total floor area of building is a demonstrated factor on the output of a model.

5 Discussion

Because buildings construction project costs are affected by several complex factors that can be directly or indirectly accounted for, there is a widespread effort to produce different effective models to predict project costs at an early stage, even though project data is limited at this stage in a project. However, there is still a shortage of viable models, and thus the easy, effective formula for the prediction model based on SVM has been designed in this study to tackle this gap. SVM can basically be seen as part of the process to find a favorable solution of preliminary forecasting of costs in building construction projects. The intention is to offer an option based on



a solution using information available at an early stage of planning of building projects that could help in making decisions about the project's feasibility and/or budget allocation based on the project's preliminary predicted costs for any building project, all main critical parameters considered. It is important to note that certain changes in the characteristics of buildings construction projects will influence a project's preliminary costs, and this is definitely the case for the actual cost of buildings. Therefore, a proper SVM technique should provide an efficient algorithm model for cost prediction in building construction projects that can positively improve costs estimation within the upstream stage despite limited information in the early planning stage of a building project.

Among options for machine learning applications, the SVM technique is to the fore in terms of developing cost estimate models, and it has been yielding robust results through developing several models to solve various problems like prediction with best regression (Petrușeva et al., 2017). This has been actively adopted for developing solutions for several practical problems

in the construction field due to its capability to generate and predict unknown values such that, nowadays, it can be considered as a standard method of machine learning for dealing with the aforementioned issues. Moreover, it offers a simple representation of the final mathematical expression used to compute the objective function when compared to the other multi regression algorithms of the same efficiency level (for example, an artificial neural network needs to be represented by a number of matrix values and biases). To enhance SVM performance, input data may be transformed into some common forms (such as ln, log, exponential, etc.), and can also present output as the total cost of the building construction project by being transformed using the same kind of natural logarithm. The SVM can effectively deal with independent parameters for the sample data with multiple types of variables, either quantitative and/or qualitative, and does not demand any assumption on the data distribution used. Thus, SVM is a powerful technique for use in predicting a building project's preliminary costs but, to successfully develop SVM, data pre-processing is crucial. The choice regarding what data is made available to build the model is made through the training and testing steps that might take the form of data transformation, normalization, and scaling (Beale et al., 2010).

Several verification statistical indexes like R, R^2 , RMSE are calculated to ensure the developed model's outputs fall into an acceptable range of validity. Almost all of the building projects data that was used to develop the SVM model was then tested using the developed SVM model to emphasize the high degree of compatibility among the samples. Furthermore, this study does not make any attempt to test the cost estimation processing efficiency of the developed model against national and/or globally conducted research because a prediction model with the same input/output factors has not yet been published, and so the input data to run the model has not been published either. Therefore, it is impossible to compare the cost value obtained from the model with previous studies. Through the analysis of the above inputs and output parameters for the developed SVM prediction model, it has been proved that the total construction area of a building has a higher impact on the building project's costs. This finding confirms the result of previous studies that investigated the impact of specific factors on building construction costs (Chandanshive and Kambekar, 2021; Lin et al., 2019). The important factor revealed in this study is the average height of floors. The importance of this factor and its impact on the building costs can be emphasized by the considerable amount of discussion of it in many of the studies that have investigated the cost of building projects in connection with building design elements (Saidu et al., 2015; Xu et al., 2015; Kim et al., 2013; Lau and Yam, 2007). The findings of this study also indicate that the number of floors is the third highest impact factor for building construction cost variation. Therefore, this result has confirmed the outputs of previous studies done by Gulcicek et al. (2013).

Another significant finding of this study is that project complexity is ranked as the 4th most important factor. This has also been extensively investigated in studies concerned with buildings project costs or construction duration time. For instance, it was found that project complexity was associated with adding cost to building construction projects (Nady et al., 2022; Qazi et al., 2016), and that it had a negative impact on building construction performance (Tran et al., 2016), and thus that it needs to be identified

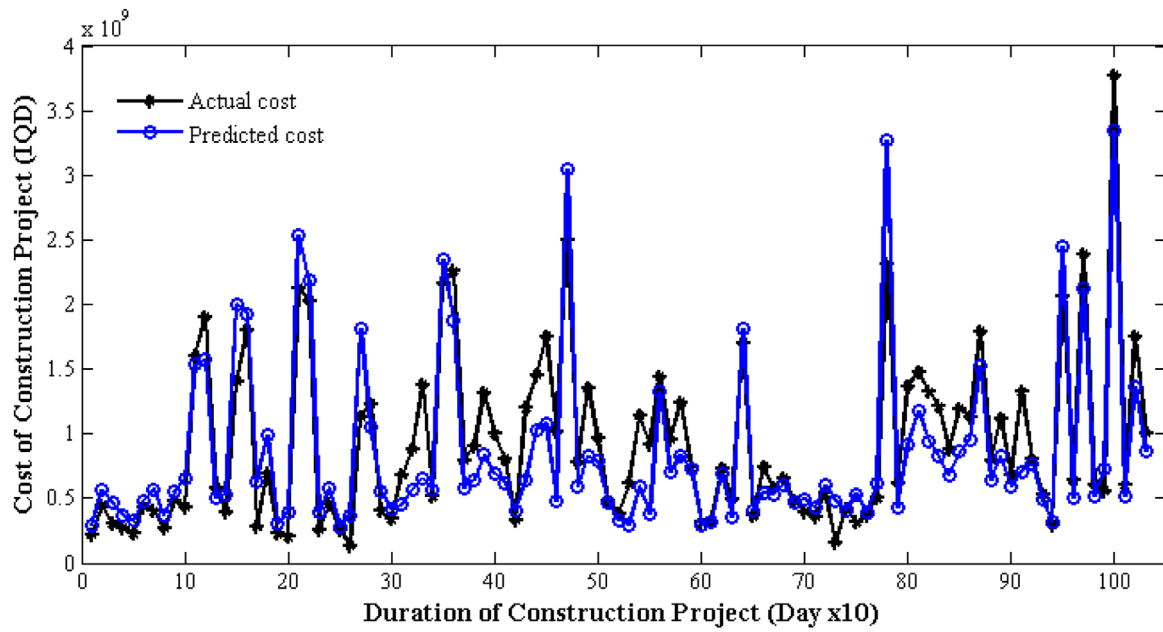


FIGURE 7 Comparison between predicted and actual spare data for the SVM model developed.

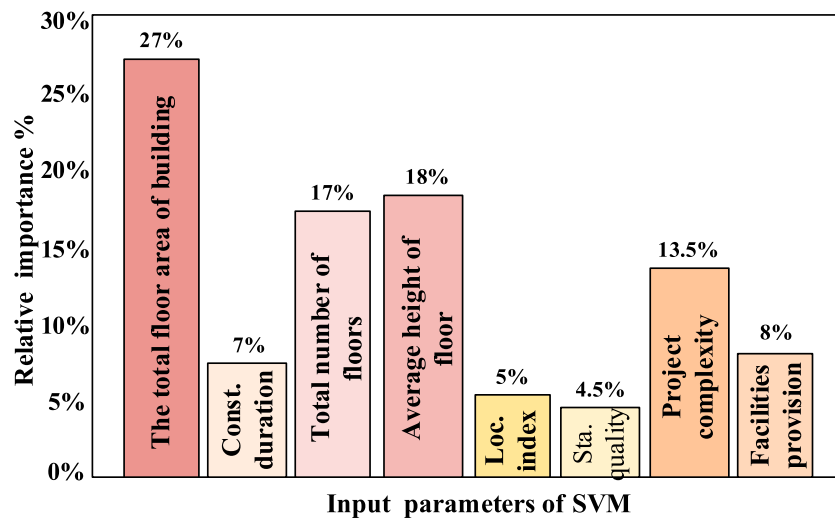


FIGURE 8 Relative importance of input parameters for the preliminary cost prediction model.

at an early stage in building construction projects (Eriksson et al., 2017). The next important factor is facilities provision, which reflects the level of aid and facilities provided to contractors or construction companies for executing construction contract work. Importantly, these facilities should be written clearly into bidding contracts and documents due to their impact on the cost and performance of construction projects. In the Iraq construction sector, some government projects and infrastructure projects provide locally manufactured construction materials at a specially mitigated price or with a tax discount for other construction materials required for

the construction project. Location is the 5th most important factor to affect building construction costs, and it involves a number of elements that, overall, have a final impact on building costs. For instance, it is seen as a characteristic that significantly influences the construction project costs due to the construction delays that might occur because of difficulties in reaching the construction site and the varying availability of labor in certain project locations. In other words, this factor reflects the nature of a specific construction project location based on whether it is inside/outside cities and urban areas, or within an open or closed allocated area, and this factor definitely

has a cost impact from a workflow management perspective. This factor does not, however, represent the entire variation in surface topography in the region because the data was acquired from buildings projects in the middle and south of Iraq, both of which have broadly similar topographic profiles. However, this surface topography profile for construction sites can be extended for some areas north of the middle areas up to the northern region of the country. Ultimately, the result of the sensitivity analysis for the inputs/output model parameters that have the biggest effects were compared with similar studies conducted in this field containing one or of these parameters to show higher consistency and confidence in the presented findings based on the existing research.

The process for adjusting the predicted cost based on the inflation rate has a significant impact on the accuracy of predicted outputs, and this can be used to decide the profitability and/or feasibility of construction projects. The inflation rate formula was developed using a linear regression method, and it gave a predicted cost close to the actual project costs. The coefficient of the determination was ($R^2 = 0.9877$) based on the cost of concrete per cubic meter, including all concrete ingredients costs (i.e., work and labor), and this concept was adopted based on cost data analysis of other studied buildings which shows that the cost of concrete work reaches more than half of a building's total cost. Moreover, it provides an easy way to forecast the inflation rate for the current year based on a base year for the model data. The impact of this result in the field of building construction can be expected to be strong for estimating the preliminary cost of two general decision-making tasks: first, the budget required to establish a project and feasibility study is investigated, and second, the feasibility of obtaining a reliable estimation in this way for characteristics of the buildings and construction sites using SVMs model is proven, adjusting for future inflation, thereby also demonstrating a knowledge of the building's cost components in order to analyze them and compare their effects on the building construction cost.

Given all of the above details, the implication of this research is to encourage buildings construction industry actors (i.e., engineers and stakeholders) to adopt an estimation model orientated towards critical relevant information to identify the preliminary cost and/or budget needs that would produce a more effective feasibility study by integrating technology into the project planning process. From another perspective, this would also promote knowledge and innovation for future research to improving cost estimation performance in the early planning of construction projects in line with the needs of buildings construction sector clients regarding what type of development estimation model is suitable for projects in the early planning stage. The scope of development of this model is necessarily limited to national-level practice for institutional and educational buildings projects. It also has some limitations due to the lack of complete documentation datasets for buildings where there are other factors not included in the model as values for the qualitative scale, such as type of exterior and interior finishing, and the topography profile of building construction site. Therefore, further study might be done by applying this theoretical framework for developing a cost estimation model in various other developing countries. The study would also have applications in different types of residential buildings projects.

6 Conclusion

This study of the Iraqi buildings construction industry addressed the main challenges of conducting a feasibility study for a building construction project for the preliminary cost estimation practice, as it has a vital role in ensuring the successful performance of the initial phase of a project. The building project preliminary cost can have a major influence on planning procedures in the buildings construction sector, especially in countries like Iraq, as in many cases, because the required feasibility study documents for construction projects are extensive, while the timeline for allocating a budget is short. Moreover, the estimation process for building construction projects is affected by many complex and changeable factors that have direct and indirect impacts on costs. Despite the importance of preliminary cost estimation in the buildings construction sector, it has not been highlighted enough by researchers in developing countries, and especially in Iraq.

The eight parameters for the SVM model that were defined as input factors were sorted into two groups, quantitative parameters and qualitative parameters, and they have major influences on building project costs in developing countries, especially in Iraq, where qualitative factors need to be more carefully built into the model. Thus, the major results from sensitivity analysis for the parameters of the developed SVM model can provide planners and cost estimators with more understanding about the attributes of buildings construction costs that included some more commonly used than other, indicating that independent factors are given different levels of significance in the practice of building construction projects.

The inflation rate needs to be considered when preparing the preliminary cost estimates while planning future buildings so as to avoid planners/estimators ending up with a budget deficit and/or lower profit margins or, in extreme cases, even a loss of competitive advantage in the next stage of the project's lifecycle. The value of this element should be considered when estimating preliminary cost regarding project feasibility in the early planning process in the construction industry sector in each region. It is recommended that, for the implementation of this developed model, participation by all project stakeholders (i.e., owners, clients, and engineering planning consultants) is required to provide a clear vision of the initial information that will be entered into the model. The framework adopted to develop this model could also enable other countries to shape their preliminary cost estimation models based on their buildings assessment criteria, thereby overcoming restrictions so as to enhance the performance of the building feasibility study in terms of cost estimation issues. For future research, the qualitative factors need to be analyzed and described with detailed steps to investigate the implicit elements for each factor so as to better aid representation of the cost prediction model with optimum estimation values. Furthermore, the inflation rate formula can be modified to involve various construction materials that have a cost impact within changeable situations.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

HJ: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. MH: Formal Analysis, Methodology, Software, Validation, Writing—review and editing. MA: Writing—review and editing. YG: Writing—review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

First author would like to thank civil engineers, construction managers, and construction experts at different Iraqi Government Offices and Companies for their support by providing the bidding

References

- Adeagbo, A. (2014). Overview of the building and construction sector in Nigerian Economy. *JORIND* 12 (2), 349–365.
- Ahuja, H. N., Dozzi, S. P., and Abourizk, S. M. (1994). *Project management: techniques in planning and controlling construction projects*. John Wiley and Sons.
- Aibinu, A. A., Dassanayake, D. H. A. R. M. A., and Thien, V. C. (2011). “Use of artificial intelligence to predict the accuracy of pre-tender building cost estimate,” in *Management and innovation for a sustainable built environment MISBE 2011*. Available at: <http://resolver.tudelft.nl/uuid:2a68de15-27b3-45ad-9c20-26b2b75e5c43> (CIB, Working Commissions W55, W65, W89, W112; ENHR and AESP. Amsterdam, Netherlands.
- Akinsiku, E. O., Babatunde, S. O., and Opawole, A. (2011). Comparative accuracy of floor area, storey enclosure and cubic methods in preparing preliminary estimate in Nigeria. *J. Build. Apprais.* 6 (3), 315–322. doi:10.1057/jba.2011.9
- Akintoye, A. (2000). Analysis of factors influencing project cost estimating practice. *Constr. Manag. Econ.* 18 (1), 77–89. doi:10.1080/014461900370979
- Alshamrani, O. S. (2017). Construction cost prediction model for conventional and sustainable college buildings in North America. *J. Taibah Univ. Sci.* 11 (2), 315–323. doi:10.1016/j.jtusci.2016.01.004
- Alumbugu, P. O., Ola-Awo, W., Saidu, I., Abdullahi, M., and Abdulmumin, A. (2014). Assessment of the factors affecting accuracy of pre-tender cost estimate in Kaduna state, Nigeria. *IOSR-JESTFT* 8 (5), 19–27. doi:10.9790/2402-08541927
- An, S. H., Park, U. Y., Kang, K. I., Cho, M. Y., and Cho, H. H. (2007). Application of support vector machines in assessing conceptual cost estimates. *J. Comput. Civ. Eng.* 21 (4), 259–264. doi:10.1061/(ASCE)0887-3801(2007)21:4(259)
- Arafa, M., and Alqedra, M. (2011). Early stage cost estimation of buildings construction projects using artificial neural networks. *J. Artif. Intell.* 4 (1), 63–75. doi:10.3923/jai.2011.63.75
- Ashworth, A., and Perera, S. (2015). *Cost studies of buildings*. London and New York: Routledge.
- Babalola, O., and Adesanya, D. A. (2009). An evaluation of the level of accuracy of mechanical service cost estimates in Nigeria. *Const. Res. J.* 2, 12–19.
- Badawy, M. (2020). A hybrid approach for a cost estimate of residential buildings in Egypt at the early stage. *Asian J. Civ. Eng.* 21 (5), 763–774. doi:10.1007/s42107-020-00237-z
- Bayram, S., Ocal, M. E., Laptali Oral, E., and Atis, C. D. (2016). Comparison of multi layer perceptron (MLP) and radial basis function (RBF) for construction cost estimation: the case of Turkey. *J. Civ. Eng. Manag.* 22 (4), 480–490. doi:10.3846/13923730.2014.897988
- Beale, M. H., Hagan, M. T., and Demuth, H. B. (2010). Neural network toolbox 7. Available at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=59c12012aa5481818f98d3c04a8b8f65f1f1d185>.
- Chandanshive, V. B., and Kambekar, A. R. (2021). Prediction of building construction project cost using support vector machine. *Ind. Eng. Strateg. Manag.* 1 (1), 31–42. doi:10.22115/IESM.2021.297399.1015
- Chang, C. C., and Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2 (3), 1–27. doi:10.1145/1961189.1961199
- Chen, J. H., Su, Y. M., Hayati, D. W., Wijatmiko, I., and Purnamasari, R. (2019). Improving preliminary cost estimation in Indonesia using support vector regression. *Proc. Institution Civ. Engineers-Management, Procure. Law* 172 (1), 25–33. doi:10.1680/jmapl.18.00040
- Cheng, M. Y., Tsai, H. C., and Sudjono, E. (2010). Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Syst. Appl.* 37 (6), 4224–4231. doi:10.1016/j.eswa.2009.11.080
- Cheng, M. Y., and Wu, Y. W. (2005). “Construction conceptual cost estimates using support vector machine,” in *22nd international symposium on automation and robotics in construction ISARC. Ferrara, Italy*, 5, 1–5.
- Cho, H. G., Kim, K. G., Kim, J. Y., and Kim, G. H. (2013). A comparison of construction cost estimation using multiple regression analysis and neural network in elementary school project. *J. Korea Inst. Build. Constr.* 13 (1), 66–74. doi:10.5345/JKIBC.2013.13.1.066
- Choon, T. T., and Ali, K. N. (2008). A review of potential areas of construction cost estimating and identification of research gaps. *J. Alm. Bina.* 11 (2), 61–72. Available at: http://eprints.utm.my/id/eprint/8238/1/AlIKherunNita2008_A_review_of_potential_areas.pdf.
- Chou, J. S., Yang, I. T., and Chong, W. K. (2009). Probabilistic simulation for developing likelihood distribution of engineering project cost. *Autom. Constr.* 18 (5), 570–577. doi:10.1016/j.autcon.2008.12.001
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi:10.1007/bf00994018
- Cunningham, T. (2013). *Factors affecting the cost of building work-an overview*. Dublin, Ireland: Dublin Institute of Technology. Available at: <https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1028&context=beschreoth>.
- Cunningham, T. (2015). *Cost control during the pre-contract stage of a building project – an introduction*. Dublin, Ireland: Report prepared for Dublin Institute of Technology. doi:10.21427/83w4-r689
- Dagostino, F. R., and Peterson, S. J. (2011). *Estimating in building construction*. 7th Edn. New Jersey: Pearson Education, Inc.
- documents and data for different buildings construction projects. In particular, we would like to thank Babylon Reconstruction Projects Department for their valuable data and information to buildings construction projects.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- David, J., Ross, A., Smith, J., and Love, P. (2002). *Building design cost management*. Oxford, UK: Blackwell Science. Available at: <https://research.bond.edu.au/en/publications/building-design-cost-management>.
- Dysert, L. (2007). Is estimate accuracy an oxymoron? *Cost. Eng.* 49 (1), 32–36. Available at: <https://www.proquest.com/openview/fd9ddcdcc28550e181dd3b3cb7a906d/1?pq-origsite=gscholar&cb=49080>.
- El-Sawalhi, N. I., and Shehatto, O. (2014). A neural network model for building construction projects cost estimating. *J. Const. Eng. and Proj. Manag.* 4 (4), 9–16. doi:10.6106/JCEPM.2014.4.4.009
- Enshassi, A., Mohamed, S., and Abdel-Hadi, M. (2013). Factors affecting the accuracy of pre-tender cost estimates in the Gaza Strip. *J. Constr. Dev. Ctries.* 18 (1), 73–94. Available at: <http://hdl.handle.net/10072/57331>.
- Eriksson, P. E., Larsson, J., and Pesämaa, O. (2017). Managing complex projects in the infrastructure sector—a structural equation model for flexibility-focused project management. *Int. J. Proj. Manag.* 35 (8), 1512–1523. doi:10.1016/j.ijproman.2017.08.015
- Gransberg, D., Jeong, H. D., Karaca, I., and Gardner, B. (2017). “Top-down construction cost estimating model using an artificial neural network,” in *Final report FHWA/MT-17-007/8232-001, Montana department of transportation* (Helena, MT, USA). Available at: <https://drive.google.com/file/d/1Z1-gzSidOTYuQFA4gQAdtasofj13-GC4/view>.
- Gulcicek, U., Ozkan, O., Gunduz, M., and Demir, I. H. (2013). Cost assessment of construction projects through neural networks. *Can. J. Civ. Eng.* 40 (6), 574–579. doi:10.1139/cjce-2012-0442
- Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS Tech. Rep.* 14 (1), 5–16. Available at: https://see.xidian.edu.cn/faculty/chzheng/bishe/indexfiles/new_folder/svm.pdf.
- Gunner, J., and Skitmore, M. (1999). Comparative analysis of pre-bid forecasting of building prices based on Singapore data. *Constr. Manag. Econ.* 17 (5), 635–646. doi:10.1080/014461999371240
- Ha, K. S., and Lee, J. K. (2012). A study on the prediction of civil construction cost on apartment housing projects at the early stage. *J. Korea Acad.-Indus. Coop. Soc.* 13 (9), 4284–4293. doi:10.5762/kais.2012.13.9.4284
- Hakami, W., and Hassan, A. (2019). Preliminary construction cost estimate in Yemen by artificial neural network. *Baltic J. Real Estate Econ. and Cons. Manag.* 7 (1), 110–122. doi:10.2478/bjreem-2019-0007
- Hegazy, T. (2013). *Computer-based construction project management: pearson*. new international edition. London, England: Pearson Higher Ed.
- Jarkas, A. M., Mubarak, S. A., and Kadri, C. Y. (2014). Critical factors determining bid/no bid decisions of contractors in Qatar. *J. Manage. Eng.* 30 (4), 05014007. doi:10.1061/(ASCE)ME.1943-5479.0000223
- Juszczyk, M. (2018). Residential buildings conceptual cost estimates with the use of support vector regression. In *MATEC Web Conf., MATEC Web Conf. Theor. Found. Civ. Eng.* 196, 04090. doi:10.1051/mateconf/201819604090
- Kadiri, D. S. (2014). An assessment of current preliminary cost estimating practice in Nigeria. *J. Environ. Des. Manag.* 6 (1and2), 97–111. Available at: https://scholar.oauife.edu.ng/sites/default/files/dskadiri/files/assessment_of_current_prelim_cost_estg.pdf.
- Kim, G. H., An, S., and Hand Kang, K. I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Build. Environ.* 39, 1235–1242. doi:10.1016/j.buildenv.2004.02.013
- Kim, G. H., Seo, D. S., and Kang, K. I. (2005). Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates. *J. Comput. Civ. Eng.* 19 (2), 208–211. doi:10.1061/(ASCE)0887-3801(2005)19:2(208)
- Kim, G. H., Shin, J. M., Kim, S., and Shin, Y. (2013). Comparison of school building construction costs estimation methods using regression analysis, neural network, and support vector machine. *J. Build. Constr. Plan. Res.* 01, 1–7. doi:10.4236/jbcrp.2013.11001
- Kim, K. J., and Kim, K. (2010). Preliminary cost estimation model using case-based reasoning and genetic algorithms. *J. Comput. Civ. Eng.* 24 (6), 499–505. doi:10.1061/(ASCE)1943-5487.0000054
- Kim, S., and Shim, J. H. (2014). Combining case-based reasoning with genetic algorithm optimization for preliminary cost estimation in construction industry. *Can. J. Civ. Eng.* 41 (1), 65–73. doi:10.1139/cjce-2013-0223
- Latief, Y., Wibowo, A., and Isvara, W. (2013). Preliminary cost estimation using regression analysis incorporated with adaptive neuro fuzzy inference system. *Int. J. Technol.* 4 (1), 63–72. doi:10.14716/ijtech.v4i1.1218
- Lau, E., and Yam, K. S. (2007). A study of the economic value of high-rise office buildings, strategic integration of surveying services. 616.
- Leśniak, A., and Zima, K. (2018). Cost calculation of construction projects including sustainability factors using the Case Based Reasoning (CBR) method. *Sustainability* 10 (5), 1608. doi:10.3390/su10051608
- Leung, M. Y., Zhang, H., and Skitmore, M. (2008). Effects of organizational supports on the stress of construction estimation participants. *J. Cons. Eng. Manag.* 134 (2), 84–93. doi:10.1061/(ASCE)0733-9364(2008)134:2(84)
- Li, H., Shen, Q. P., and Love, P. E. (2005). Cost modelling of office buildings in Hong Kong: an exploratory study. *Facilities* 23 (9/10), 438–452. doi:10.1108/02632770510602379
- Lin, T., Yi, T., Zhang, C., and Liu, J. (2019). Intelligent prediction of the construction cost of substation projects using support vector machine optimized by particle swarm optimization. *Math. Probl. Eng.* 2019, 1–10. doi:10.1155/2019/7631362
- Ling, Y. Y., and Boo, J. H. S. (2001). Improving the accuracy estimates of building of approximate projects. *Build. Res. Inf.* 29 (4), 312–318. doi:10.1080/09613210122440
- Liu, X. L., Yin, T., and Wu, G. D. (2013). Practical application study of Gaussian process model in construction project cost estimation. *Adv. Mat. Res.* 671, 3100–3106. doi:10.4028/www.scientific.net/AMR.671-674.3100
- Mahalakshmi, G., and Rajasekaran, C. (2018). “Early cost estimation of highway projects in India using artificial neural network,” in *Sustainable construction and building materials: select proceedings of ICSCBM* (Springer Singapore), 25, 659–672. Lect. Notes Civ. Eng. doi:10.1007/978-981-13-3317-0_59
- Mahamid, I., Al-Ghonamy, A., and Aichouni, M. (2014). Factors affecting accuracy of pretender cost estimate: studies of Saudi Arabia. *nt. J. Appl. Eng. Res.* 9 (1), 21–36. Available at: <https://www.researchgate.net/profile/Mohamed-Aichouni/publication/265377554>.
- Mahamid, I., and Bruland, A. (2010). “Preliminary cost estimating models for road construction activities,” in *Proceedings of the FIG congress*.
- Mankiw, N. G. (2012). *Ten principles of economics*. Melbourne: Cengage Learning. Available at: https://homepage.ntu.edu.tw/~josephw/Principles_19F_lecture1a.pdf.
- Meharie, M. G., Gariy, Z. C. A., Ndisya Mutuku, R. N., and Mengesha, W. J. (2019). An effective approach to input variable selection for preliminary cost estimation of construction projects. *Adv. Civ. Eng.* 2019. doi:10.1155/2019/4092549
- Nady, A. E., Ibrahim, A. H., and Hosny, H. (2022). Factors affecting construction project complexity. *Int. J. Eng. Sci. Technol.* 37, 24–33. doi:10.21608/eijest.2021.96807.1100
- Nsofor, G. C. (2006). *Comparative analysis of predictive data-mining techniques*. MSc Thesis, The University of Tennessee, Knoxville, USA. Available at: https://trace.tennessee.edu/utk_gradthes/4495
- Okereke, R. A. (2019). Factors for preliminary cost estimation of building project in the Nigerian construction industry. *J. Sci. Eng. and Tech.* 2019 6 (1), 97–109. Available at: <https://d1wxtqst1xzl7.cloudfront.net/63965961/>.
- Park, W. Y., Cha, J. H., and Kang, K. I. (2002). A neural network cost model for apartment housing projects in the initial stage. *J. Archit. Inst. Korea.* 18 (7), 155–162.
- Patle, A., and Chouhan, D. S. (2013). “SVM kernel functions for classification,” in *2013 int. In conf. Adv. Technol. Eng. ICATE* (Mumbai, India), 1–9. doi:10.1109/ICADTE.2013.6524743
- Petruseva, S., Zileska-Pancovska, V., Žujo, V., and Brkan-Vejzović, A. (2017). Construction costs forecasting: comparison of the accuracy of linear regression and support vector machine models. *Tech. Gaz.* 24 (5), 1431–1438. doi:10.17559/TV-20150116001543
- Qazi, A., Quigley, J., Dickson, A., and Kirytopoulos, K. (2016). Project Complexity and Risk Management (ProCRiM): towards modelling project complexity driven risk paths in construction projects. *Int. J. Proj. Manag.* 34 (7), 1183–1198. doi:10.1016/j.ijproman.2016.05.008
- Rezaian, A. (2011). Time-cost-quality-Risk of construction and development projects or investment. *Middle East J. Sci. Res.* 10 (2), 218–223. Available at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=bc877b815a62c4a0bfaae911093144f2c789e301>.
- Saidu, I., Polycarp, A., Abdulazeez, A., and Wasiu, O. (2015). Assessment of the effect of plan shapes on cost of institutional buildings in Nigeria. 4 (3), 39–50. Available at: <http://repository.futminna.edu.ng:8080/jspui/bitstream/123456789/7266/1/My%20plan%20shape.pdf>.
- Sanni-Anibire, M. O., Mohamad Zin, R., and Olatunji, S. O. (2021). Developing a preliminary cost estimation model for tall buildings based on machine learning. *Int. J. Manag. Sci. Eng. Manag.* 16 (2), 134–142. doi:10.1080/17509653.2021.1905568
- Shehatto, O. M. (2013). Cost estimation for building construction projects in gaza strip using artificial neural network (ANN). *CE-Unit Cost.*, 102.
- Shin, Y. (2015). Application of boosting regression trees to preliminary cost estimation in building construction projects. *Comput. Intel. NEUROSC* 2015 (1), 1–9. doi:10.1155/2015/149702
- Shutian, F., Tianyi, Z., and Ying, Z. (2017). Prediction of construction projects’ costs based on fusion method. *Eng. Comput.* 34 (7), 2396–2408. doi:10.1108/EC-02-2017-0065
- Son, H., Kim, C., and Kim, C. (2012). Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables. *Automation Constr.* 27, 60–66. doi:10.1016/j.autcon.2012.05.013
- Son, J. H., and Kim, C. Y. (2006). A study on the model of artificial neural network for construction cost estimation of educational facilities at conceptual stage. *Korean J. Constr. Eng. Manag.* 7 (4), 91–99.

- Sönmez, R. (2004). Conceptual cost estimation of building projects with regression analysis and neural networks. *Can. J. Civ. Eng.* 31, 677–683. doi:10.1139/104-029
- Tayeh, B. A., Al Hallaq, K., Alaloul, W. S., and Kuhail, A. R. (2018). Factors affecting the success of construction projects in Gaza Strip. *Open J. Civ. Eng.* 12 (1), 301–315. doi:10.2174/1874149501812010301
- Tran, D. Q., Molenaar, K. R., and Alarcón, L. F. (2016). A hybrid cross-impact approach to predicting cost variance of project delivery decisions for highways. *J. Infrastruct. Syst.* 22 (1), 04015017. doi:10.1061/(ASCE)IS.1943-555X.0000270
- Trost, S. M., and Oberlender, G. D. (2003). Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *J. Constr. Eng. Manag.* 129 (2), 198–204. doi:10.1061/(ASCE)0733-9364(2003)129:2(198)
- Xu, M., Xu, B., Zhou, L., and Wu, L. N. (2015). “Construction project cost prediction based on genetic algorithm and least squares support vector machine,” in *5th international conference on civil engineering and transportation* (Guangzhou, China: Atlantis Press), 1004–1009. Available at: <https://www.atlantis-press.com/article/25845336.pdf>.
- Yang, I. T. (2005). Simulation-based estimation for correlated cost elements. *Int. J. Proj. Manag.* 23 (4), 275–282. doi:10.1016/j.ijproman.2004.12.002
- Yang, S. W., Moon, S. W., Jang, H., Choo, S., and Kim, S. A. (2022). Parametric method and building information modeling-based cost estimation model for construction cost prediction in architectural planning. *Appl. Sci.* 12 (19), 9553. doi:10.3390/app12199553
- Yu, W. D., and Skibniewski, M. J. (2010). Integrating neurofuzzy system with conceptual cost estimation to discover cost-related knowledge from residential construction projects. *J. Comput. Civ. Eng.* 24 (1), 35–44. doi:10.1061/(ASCE)0887-3801(2010)24:1(35)