Cost estimation and prediction in construction projects: a systematic review on machine learning techniques

Sanaz Tayefeh Hashemi¹ · Omid Mahdi Ebadati² · Harleen Kaur³

Received: 27 December 2019 / Accepted: 6 September 2020 / Published online: 15 September 2020
© Springer Nature Switzerland AG 2020

Abstract

Construction cost predictions to reduce time risk assessment are indispensable steps for process of decision-making of managers. Machine learning techniques need adequate dataset size to model and forecast the cost of projects. Therefore, this paper presents analysis and studied manuscripts that proposed for cost estimation with machine learning techniques for the last 30 years. The impact of this manuscript is deep studied of machine learning techniques and applied an analysis methodology in cost estimation based on direct cost and indirect cost of construction projects, which consists of two parts. In the first part, for study the proposals, we focus on collecting related studied from Google Scholar and Science Direct journals. The interested application areas for project cost estimation are building, highway, public, roadway, water-related constructions, road tunnel, railway, hydropower, power plant and power projects. The second part is regarded to the analysis of the proposals. For cost analysis, there are possibilities to consider two approaches as qualitative and quantitative. However, reflect to the machine learning techniques the quantitative approach is studied. In quantitative approach, we categorized the models in three parts, as statistical, analogues and analytical model and analyse them based on their features. Correspondingly, papers have been thoroughly investigated based on the application area, method applied, techniques implemented, journals, which have been published in, and the year of publication. The most important outcome of this study is to find out the different analytics methods and machine learning algorithms to predict the cost estimation of construction and related projects and aid to find out the suitable applied methods.

Keywords Cost estimation · Prediction · Construction project · Machine learning · Systematic review

1 Introduction

Cost prediction is a vital process for every business in that it is a predecessor for budget prices and resource allocation in a project life cycle. Actually, it is hard to obtain input data for cost estimation process, while the scope of work is barely known in that it might lead to poor and rough estimates. The more, the project scope is known there are more chances to generate estimates that are more accurate in that more specifications of the project are defined. However, it should be taken into account that, on the other

hand, by the progressive elaboration, the process of cost control becomes more difficult if the project is based on inaccurate cost estimates. Furthermore, construction industry due to its characteristics and large amounts of capital needed to initiate and continue the project, are the project types which need more attention because they are high-risk [1]. Either overestimating or underestimating the cost of these projects will lead to future deviations in budget vs. realized cost. Hence, the methods used in this realm, their respective accuracy, and even their gaps have shown growing interest. Methods with more consistent

✉ Omid Mahdi Ebadati, ebadati@khu.ac.ir | ¹Department of Information Technology Management, Kharazmi University, Tehran, Iran. ²Department of Mathematics and Computer Science, Kharazmi University, Tehran, Iran. ³Department of Computer Science and Engineering, Jamia Hamdard, New Delhi, India.

Fig. 1 Bid structure and analysis in projects

results can facilitate and smooth the path for cost estimators provided that their related gaps can be investigated and overcome in order to acquire better results. In conventional methods, by knowing work packages and their prices and how they are distributed along the project lifetime; the total project cost can be estimated. Which this will be an input for project resource allocation and further budget calculations. The conventional methods have shown that they are not merely enough. Thereby the lack of a systematic approach in order to reduce the error of the estimation process has entailed in studies that most of all have tried to take advantage of mathematical models, machine learning techniques, and so on to overcome inaccurate or may even erroneous predictions. What is estimated as project construction cost is different from tender price in that the tender price contains other amounts, including company profit and contingency reserve. Contingency reserve is the amount allocated to known risks during the project execution, which is an estimated amount of reserve. The components of project cost are depicted in Fig. 1 due to the contractor's viewpoint [2].

As shown in Fig. 1 [3], the project cost includes the project indirect cost and direct cost. The project direct cost itself is composed of costs directly spent in the project and the indirect part, which is mainly the overhead of the project, incurred either in the project itself or on the staff side.

This classification is described as follows:

- **Direct costs** Direct costs can be defined as costs that are directly spent in the project and its production activities, which can be well estimated, while adequate information is available about site condition, construction method used, and the resources utilized. In fact, direct costs are composed of several items such as cost of the labor assigned to the project, equipment used, materials and crews and the subcontractors, which the work packages are assigned to, on behalf of the general contractor.
- **Indirect costs** Indirect costs are classified into the following categories:
 - **Project overhead** These costs are mainly the costs, which are indirectly incurred in the project and are in charge of the project work packages, but cannot be directly assigned to them such as utilities, supervisory, etc.
 - **General overhead** These costs, in contrary to project overhead, cannot be attributed to each project individually and are mainly the staff side costs, such as an amount of money spent in the head office, personnel cost, and so on, which can be attributed to projects proportionate to their costs toward the total costs of the contractor's organization.
- **Markup** The company bid price is the summation of project's cost, and an amount regarded as markup

which itself is comprised of the following amounts of money:

- *Profit* The amount of money attributed to company's profit, which depends on the business objectives, the industry competition level, and also how much the contractor wills to win the project over its rivals.
- *Risk contingency* Usually known as identified risks or known unknown, which is also considered in markup and is the amount of money, set aside for uncertain situations, which can affect the project performance, including unexpected events, labor issues, etc. [3].

Aims and objectives

The objectives of this systematic review include:

1. Investigating the criteria for construction projects cost estimation.
2. Determine the criteria of construction projects based on application area, method applied, techniques implemented, journals, and the year of publication.
3. Reviewing the existing models of machine learning techniques in cost estimation of construction projects.
4. Assessing the methods, techniques and criteria for construction project cost estimation.

The rest of the paper is structured as follows; Sect. 2, explores the research methodology, the way to retrieve data, cost estimation techniques and analytics models. Section 3, concisely deliberates about the results and related discussion and distribution methods, and the paper is concluded in Sect. 4, and brief about the final results and methods, limitations and future work.

2 Research methodology

2.1 Types of studies

This research investigates the available models and criteria in the field of the smart-grid project for cost estimation from the past 30 years. This is to emphasize that the present review paper does not include all the articles done in this scope and just the ones with the defined keywords and in the domain of construction projects. This study will impose no restriction on the type of proposal work conducted on the subject and no limitations on the date of publication of the documents as well.

2.2 Information sources and search strategy

Databases such as Google Scholar and Science Direct will be searched to access the relevant documents. These two main sources of academic database are totally included more than 400 million documents. Database will be searched using following keywords to obtain relevant papers: "Construction", "Cost estimation", "Cost Prediction", "Regression Analysis", "Case Based Reasoning", "Analogy", "Artificial Intelligence Techniques". The used keywords in this study are the most important guidelines in this area, which can help to reach to relevant papers. For such, no limitations impose on the publication status of the extracted studies.

Google Scholars

All fields from 1985–2020 (((Cost Estimation AND Construction) OR (Cost Prediction AND Construction) OR (Cost Estimation AND Regression Analysis) OR (Construction AND Regression Analysis) OR (Case Based Reasoning) OR (Analogy) OR (Construction AND Artificial Intelligence Techniques) OR (Cost Prediction AND Artificial Intelligence Techniques) OR (Cost Prediction AND Analogy) OR (Regression Analysis)))) AND ((Machine Learning Techniques OR forecasting)).

Science Direct

All Article Types in journals or books, years 1985–2020 (((Cost Estimation AND Construction) OR (Cost Prediction AND Construction) OR (Cost Estimation AND Regression Analysis) OR (Construction AND Regression Analysis) OR (Case Based Reasoning) OR (Analogy) OR (Construction AND Artificial Intelligence Techniques) OR (Cost Prediction AND Artificial Intelligence Techniques) OR (Cost Prediction AND Analogy) OR (Regression Analysis)))) AND ((Machine Learning Techniques OR forecasting)).

2.3 Selection process

We briefly investigate the papers to identify relevant manuscripts based on the title and abstract. The results entered into the EndNote and remove the duplicates. Following determining the relevant headlines used to consider eligibility of the manuscript criteria to study the full-texts of any potentially related discussions identified so far. To find other theoretically relevant articles, the references of the extracted papers also are examined. Using the comments made by specialists of this field, key journals of the field identified and the relevant articles have been reviewed in details in terms of the following sections, which are embedded in a form used to retrieve data from paper (Table 1).

Table 1 Form used to retrieve data from studied articles

Data retrieved	Description	Detail
Title	Title of the main study	
Year	Publication year of the study	
Journal	Name of the Journal publishing the article	
Application area	The area within which the case has been studied	
Approach type	The main domain of cost estimation techniques	Qualitative, Quantitative
Approach	The subdomain of cost estimation techniques	Parametric, Analogous, Intuitive, Analytical
Method	The final tool implemented to conduct the study	ANN, Fuzzy NN, SVM, RA, PSO, RBFNN, Cost Model, Expert System, AHP, CBR, Monte Carlo, Decision Tree, Reinforcement Learning, Online Machine Learning

2.4 Cost estimation techniques

The total number of 92 papers have been studied thoroughly, in terms of application area, applied methods, techniques implemented, journal published in, and the year of publication.

In cost estimation scope, many methods and techniques are used, out of which Artificial Neural Networks (ANNs), hybrid models of ANN with secondary artificial intelligence or meta-heuristic methods, Radial Basis Function Neural Network (RBFNN); Case-Based Reasoning (CBR), Regression Analysis (RA), Particle Swarm Optimization (PSO), Decision Tree (DT), and Expert Systems are investigated here.

Artificial neural networks are one of the many algorithms, which are modelling biological learning processes by computers. They are classified under a major classification named machine learning. In fact, machine learning is the process of programming the computers to optimize a performance based on a past available data or experience [4].

The first mathematical model of an artificial neural network model was formulated by McCulloch and Pitts in 1943 [5]. Artificial neural networks known as neural networks are analogy-based, non-parametric information-processing systems that have inspired their functionality and structure from the brain's biological neural networks [6]. The most challenging problems, which neural networks are used for, are pattern recognition, clustering/categorization, and prediction/forecasting [7]. In forecasting problems, neural networks are trained based upon past data and depending upon their generalization ability; they can provide forecasting for novel cases.

Neural networks have several advantages, including their capability to perform predictions with less required developed statistical trainings, ability to detect intricate nonlinear relationships among variables, ability to discover all possible interrelations between variables, and the capacity to be developed through the use of numerous

training algorithms. However, like any other subject, there remain some disadvantages, including their “black box” mechanism leading to discouragement in finding the origin of the results, their difficult applicability to some problems, their need for high computational resources, and their vulnerability in overfitting and experimental construction, which are highly in need of resolving several matters such as their topology and other methodological matters [8].

On the other hand, ANN is extremely data driven and will show low prediction performance, while being fed with a small number of data, leading to over specification, which means that they can perform well with the available data, but are incapable of predicting novel cases [9]. In their point of view, the application of heuristic rules such as preventing the model from being further trained, while there seems to be no more improvement in the network MSE and also using fewer numbers of nodes in hidden layers can mitigate this possibility.

Despite the black box mechanism of neural networks, they have been widely used in prediction problems demonstrating reasonable results as scrutinized in the literature. Developing hybrid model of back propagation neural networks and genetic algorithm will lead to more accurate predictions and prevent the model from presenting erroneous performance hence can overcome the encapsulated shortcomings [10]. The use of hybrid models of ANN with secondary artificial intelligence or meta-heuristic methods such as genetic algorithms, bee colony algorithm, and artificial immune systems have been proposed in numerous articles in order to cover the drawbacks of ANNs and thus enable them to be applied in diverse problems [11]. Genetic Algorithm (GA), one of these meta-heuristic methods and a family of evolutionary computation models, was first invented by John Holland in 1960s [12]. As the optimization problems are occurring in dynamic settings, they require a kind of feedback from the environment, which the problem is taking place regarding the success or even failure of the current applied strategy, that will exploit the

earned knowledge in order to evolve the applied strategies and recombine the best pieces of competing strategies to reproduce much more fitting individuals [13].

Furthermore, CBR is a data mining technique, which remembers the information and also uses the solutions implemented for the similar past cases in solving new problems [2]. The main source of this information and knowledge is the case, which is reused though through matching by some kind of tolerance [14].

On the other hand, RA can be classified as a data oriented technique that deals with just the data in hand and not the characteristics behind them and is divided to two linear and nonlinear models [15]. In addition, decision trees are widely used for solving classification problems. Decision tree is constructed continuously based on the feature that best satisfies the branching rule. This process is then performed iteratively for each branch [16]. Classification and regression decision trees deal with predicting a dependent variable based upon a predictor variable. The response variable in the former includes a finite set of values, while in the latter contains continuous or discrete set of variables [17]. Regression trees are good substitution for basic regression methods. Decision tree is mainly constructed based on those attributes in the dataset that are pertinent to the classification case, thus it can be mostly regarded as a feature selection problem [18].

Besides expert systems are well known by their application of knowledge, facts and methods elicited from human experts that have been affirmed to be effective in solving the cases of the similar domain [19].

Furthermore, the papers are categorized by the year of publication and the journals within, which papers have been published. In addition, the papers are studied in terms of the area within, which the cost estimation method has been applied. The current fields are as followings: building projects, highway projects, public projects, road way projects water-related construction projects, road tunnel projects, railway projects, hydropower projects, and power plant projects. Besides, the cost estimation methods in these papers are investigated from the applied technique's viewpoint.

Cost estimating methods are classified into two main categories: qualitative and quantitative methods, which will be described in detail later. The total view of cost estimation modelling techniques is depicted in Fig. 2 (Modification of [20]).

2.4.1 Qualitative approaches

Qualitative approaches are based on estimator's knowledge of the project, the scope of work, and influencing factors and are divided into two classes: expert judgment and heuristic rules. Expert judgment depends on the good

or bad results of the past estimations based on judgment. According to [21], expert judgment technique is mainly taking advice from the more experienced experts and peers to check the validity of the estimating results. This technique is, in fact, intuition-based and mainly relies on unspoken yet not well documented extrapolation techniques, which are the power in hands of experienced experts that can professionally assure the reliability of estimations [21].

On the other hand, the heuristic rules in cost estimating are due to intuitive judgments and are done as a rule of thumb to ease the process of estimating and are extracted from relative similar projects.

2.4.2 Quantitative approaches

Quantitative approaches can be defined as methods relying on the process of collecting and analysing historical data and applying quantitative models, techniques, and tools to estimate the project's cost. Quantitative cost estimating approaches are classified into three main categories: statistical, analogous, analytical ones.

2.4.3 Statistical methods in cost estimation

Statistical methods, on the other hand, are based on formulas or other alternative approaches to establish a causal correlation between final costs and its corresponding characteristics [20].

Parametric cost estimating methods evaluate the cost through regarding characterizing parameters like mass, volume, and cost without considering little details [22]. In fact, in this way, the project cost is estimated based on defining its causal link with these parameters, and the result will be a mathematical function of the corresponding variables. This approach is efficient at early stage of a project, where there is little information available about the project [23], however, it suffers from the minimal necessary result justification [22]. There are three types of parametric cost estimation methods as follows [24]:

- *The method of scales* This method is applicable in prevailing technologies, which simple products of different sizes are produced. Evaluating the most influencing technical parameters is the prerequisite of this method. Thereafter, this evaluation is compared with those of finished projects, which makes this method a combination of analogous and parametric approaches. The main disadvantage of this method is assuming that the cost and considered parameters are interrelated through a linear function [24].
- *Statistical models* In this method, the activities are divided into major different scopes through, which the

Fig. 2 Cost estimation modelling techniques. (Modification of [20])

final mathematical formulae is constructed. This model is composed of three main data types [24]:

- Technical specifications
- Relationships between the data and final variables
- Constants
- *Cost estimation formulae (CEF)* CEF is a mathematical relationship between the final cost and a limited set of technical parameters. The major parameter categories are as follows [24]:
 - *Physical values* According to functional description
 - *Dimensioning values* According to solution description

The most probably prevailing parametric methods are regression analysis and optimization techniques [20].

Parametric cost estimation methods are faced with different drawbacks, which some of them are described as follows; through application of these methods, different results are the sole issue without giving a vision about

the origin of them. On the other hand, lack of necessary parameters during early stages will result in uncertainty of the results. In addition, the designer should be aware of the influence of each parameter on the final cost. CEFs in particular, are incapable of solving specific cases. Sometimes also, there is a need to obtain results of regression analysis in four or five similar cases to reach to the most reliable cost. Despite all these disadvantages, they are considered as useful cost estimation methods due to their rapidity of execution [24].

Analogous models

Analogous models are based on similar past cases, which are reused and adjusted in different cases [25]. This similarity is due to functional or geometrical homogeneity between cost structures, which are alike [20].

According to [21], analogous methods are known to be the simplest method of estimating through. The cost of projects is estimated in compared to their similar completed projects that are available as a historical database. Thus, project managers have to consider the most

available parameters to include in their process of estimating to reach better results; however, this method is a kind of rough estimate, which is easy to use, but with lower levels of complexity and accuracy as well [21].

Analytical models

Analytical models instead, are the process of estimating costs by accurately defining the cost corresponding to each processing phase attribute in details, and afterwards using a bottom-up approach for aggregating the project total cost, thus this approach is leading to a more accurate result [25].

3 Results and discussion

This section discusses the findings of this study. Initially, an overview of the data analysis describes. Then, it presents the report and discussion of the study findings according to the research methodology in the separate subsections. Furthermore, illustrate the result of comparison of different models within the context.

3.1 Data analysis

The present study explores the existing methods and techniques for the cost estimation of projects and extracts approaches components. A classify analysis is conducted using the existing methods and tools and comparison made for different models. The components extracted from all the studied papers classified in terms of application area, methods, techniques, journals and year of publications. The results are discussed in the following sub-sections.

The total studied proposal papers are 92, which based on the considered features, they categorized for different approaches. The sum of 69 of articles are directly reviewed in the field of cost estimation in construction projects and 48 of them have focused on machine learning techniques. Elfaki et al. [26] have also reviewed the application of intelligent techniques in the construction cost estimation field. All these results have been summarized in "Appendix 1". This Appendix shows; the total view of the present reviewed papers, in terms of the reference, year of publication, first author, area within which the method(s) has/have been applied and the method(s) in order of superiority of performance.

3.2 Application area

Table 2 shows an overall view of the reviewed papers applied in different areas. As it is shown in this table, most of the articles have studied building projects in general

Table 2 Reviewed articles by areas

Reviewed articles by areas	
Building projects (40)	[1, 27–65]
Highway projects (8)	[23, 66–72]
Public projects (5)	[73–77]
Road way projects (6)	[78–83]
Water-related construction projects (3)	[84–86]
Road tunnel projects (2)	[87, 88]
Railway projects (2)	[89, 90]
Hydropower projects (2)	[15, 91]
Power plant and power projects (1)	[92]

and less than half have scrutinized specific construction projects.

Cost estimation in building projects has been studied in a wide range of studies. In fact, building projects in this paper is meant the projects related to constructing buildings and such cases. The aforementioned projects' distributions are shown in the time horizon in Fig. 3. As it is shown in this figure, the most studies are done in the year 2011 and 2017 with building project standing on the top; on the other hand, hydropower projects, own the least number of studies in this spectrum.

3.3 Methods

Machine learning techniques have been defined as a system that can learn from data. In general, the main strong point of machine learning techniques can be identified as: the ability of handle uncertainty in methods, the ability to manage and perform with incomplete data, and the ability to decide and conclude the new cases based on experiences from analogous cases.

Khalaf et al. [27] have applied PSO in estimating cost and duration of 60 construction projects at the early stage. What has been inferred from this study is that PSO has been well performed with high accurate results, while it is encountering parameters with a wide range of variability. The other strength of this model is that it is based on existing projects and is more reliable than the projects based on judgement and experimental cases. However, this paper tries to examine the model with a wider range of parameters and also apply it to green buildings. On the other hand, [28] have studied the application of ANN in cost estimation of building projects, and it compared the results with RBFNN paper methods, and showed the ANN outperforms. Then, the study followed by optimizing the model accuracy, and applying it to other types of projects, and using other methods for cost factors' screening. In addition, [1] have proposed a cost model, which is a quantity based one, through which the results will be

Fig. 3 Distributions of different projects studies in years

finally multiplied by the desired prices. Although the recommended model outperforms the CBR method is compared to it, there is a need to conduct more researches to compare the results with further parametric methods to validate the reliability of the current model. This study also, takes advantage of a parameter making process, which its role is to summarize many effective cost factors into a package of influential parameters. On the other hand [29] have investigated the capability of multilayer feed forward neural network model with a backpropagation learning algorithm for estimating the cost of 78 building projects in India, along with testing the effectiveness of either the early stopping or Bayesian regularization approach on the generalization competency of the network and on the overfitting error as well; where the later approach surpasses. Furthermore, [30] have implemented fuzzy logic to predict the cost of building projects. As their model is not dynamic in response to market prices, the need for more agile model is felt. Furthermore, [31] have used an integration of BP neural network and genetic algorithm to estimate the cost of residential buildings. The role of GA is to improve the ANN performance by preventing it from falling into local maximum point and increasing the convergence speed. Besides, [32] it takes the advantage of multiple regression analysis to estimate the cost of residential buildings. In the research point of view, 92% of the cost of residential building is affected by the land

area and building area, and the remaining 8% is stemmed from other factors.

Cost estimation of residential buildings with the use of multifactor linear regression has been considered in [33], which has reached an accuracy around 92% in the end. The research has recommended to compare the results with those researches that implemented neural network technique to see the differences. Actually, the study is highly advocated the use of cost estimation models in construction projects instead of conventional methods. In [34], application of Back-Propagation Artificial Neural Network (BPANN) in order to predict the cost of building projects in Nigeria can be seen, however, the model can only be implemented in institutional type of buildings and no other types of buildings or any other projects cannot be estimated by this method. Also, the criterion for the model performance is the prediction errors and other means of evaluations have not been taken into account. Furthermore, [35] have conducted a survey to investigate the most influencing factors on the cost estimating process, then developed the ANN model, and eventually conducted a sensitivity analysis. They have reached remarkable results with MLP neural network, while applying it at the very early stage of the project. Furthermore, [36] have implemented ANN for cost prediction of building projects in Philippines. They have concluded that ANN oftentimes can show an acceptable performance

despite the incomplete available datasets; however, the enriched datasets is highly recommended. Besides, [37] have implemented a hybrid model of ANN and GA in order to overcome some drawbacks of ANN, including the slow convergence and being trapped in local minimums. Also, [38] have applied a hybrid method of CBR and GA in early stages of high-rise building projects to estimate the cost, in a less erroneous way. The application of GA has successfully improved the process of the estimation model by defining the weights of cost factors, though, they recommend to include other cost categories for these types of buildings such as engineering fees and contingencies, considering indexes for different locations, applying other algorithms, instead of GA in order to improve the weights, implementing the model with larger projects dataset, and determining other different cost factors that are effective on the cost estimation process.

On [39], has comprehensively studied different possible ANN architectures with different learning rates and eventually has compared them, and it is concluded that the best one is an MLP neural network with two hidden layers. It has reached to key findings in the research such that, the number of neurons in the hidden layer, and the learning parameters have more effects on the network generalization ability rather than on its accuracy ability. In addition, the number of hidden neurons is more effective than the learning parameters. On the other hand, the network is highly sensitive to the number of inputs, so that the more inputs; the more the possibility of overlearning in the network. Finally, the study suggests to implement the model in other types of buildings and to compare the current results with other cost estimation methods such as multiple linear regression. Moreover, the development of ANN and Support Vector Machine (SVM) for predicting the cost of building projects and schedule is presented in [40], out of which, SVM has shown superior performance; therefore, ANN is more applicable in nonlinear sample data. The paper also recommends using an ensemble of ANN and SVM, while it should be taken into account that early planning is considered a key factor in project success. In addition, [41] have conducted a survey and implemented data analysis in order to extract the main influencing input parameters of their fuzzy model. They have mentioned that the use of two-sided membership function has shown better results than other studied models. They also suggest that comparing the result with other single or combined methods can also be useful. Besides, [42] have taken advantage of Multiple Regression Analysis (MRA) capabilities to revise CBR in order to enhance the prediction accuracy. They suggest considering also nominal variables and investigating the origins of the increase in the error rate. Son et al. [43] have also applied a hybrid model of principal component analysis and Support Vector Regression (SVR)

and compared them with SVR, ANN, Decision Tree, and Multiple Linear Regression (MLR) out of which eventually, they presented that SVR algorithm is outperformed.

In another research, [44], the authors have successfully applied case adaptation in order to enhance CBR performance. They suggest that implementing this model in other types of projects as a future research. Also, they proposed the uses of qualitative factors are effective on the model and highly recommend considering the bias resulted from data originated from different regions. In addition, [45] in their study, have studied BPANN model and compared it with regression in cost estimation of building projects. The best architecture of the neural network is chosen after a process of trial and error out of which eventually, the neural network showed a better performance in compared to regression analysis. In this research, it has been recommended that larger dataset with more accurate information can be used in the future researches. Application of a hybrid model (Modified PSO and fuzzy neural network) in cost estimation of construction projects has also been scrutinized in [46] as a novel approach within, which the model is capable of being applied to other new cases. Further, [47] have investigated a BP-ANN model to predict cost of building projects. The best promising architecture is generated after several trials. They also called for larger dataset as an input for the network, in order to improve its performance. Cheng et al. [48] have integrated neural network with fuzzy logic in order to handle uncertainties as a novel approach. They claim that the hybrid neural network is more effective than the mere neural network in predicting cost of construction projects at very early stage of the project. In addition, there is a concrete evidence that the hybrid neural network is able to address both linear and nonlinear connections in the hidden layer. Cheng et al. [49] have also taken simultaneous advantage of GA, fuzzy logic, and ANN for global optimization, approximate reasoning, and input–output mapping, respectively. Their cost estimation method is applicable to early stages of the project for project manager's decision-making process.

A combination of the AHP-based and simulation-based cost model can be seen in [50] for a single project. They have reached to reasonable results, and they suggest that a wider range of data be fed to their model will be better results. In addition, they recommend a cash allocation system for multiple projects can be developed with a user interface to work around. Their model can be applied to other construction projects as well, and they provided that a modification in weights and evaluation criteria are considered. On the other hand, a combination of rough set (RS) theory and artificial neural network (ANN) is implemented in [51]. In fact, rough set theory is used to filter the main effective factors in a cost estimation process. They

recommend that this hybrid network be implemented in construction projects in that it surpasses the mere ANN results. In their point of view, the less input data can cause, less overfitted network. They recommended combining their model with cost control methods, dealing with data and project cost index in a more scientific way as their future work. An et al. [52] have used CBR to estimate construction cost of residential buildings. What makes their method worth of use is the application of AHP method in order to interfere with expert's knowledge in the estimation process. In addition, [53] have compared three models of NN to predict project's cost, including BPANN, BPANN adjusted with GA, and NN modified with GA, where the second one outperforms the others. The future of this research is needed to more adjustment of the GA parameters rather than determining them manually.

In [54], the researchers have used regression analysis to estimate cost of building projects in Singapore and have selected principal components, while being encountered with a large amount of independent variables towards dependent variables. Thus, this will render a regression model with few uncorrelated principals that will eventually produce a better performance. Li et al. [55] also have investigated the application of regression analysis to estimate the cost of building projects, while incorporating a step-wise variable selection in order to scrutinize the relationship between the available independent variables and the cost of a project as a dependent one. This wise is noted by the authors that the accuracy of the model has improved towards the classic model. Comparison between MRA, ANN and CBR is delicately performed in [56] out of which, ANN outperforms in terms of accuracy, while CBR outperforms in terms of time spent for cost estimation process. In fact, in this study, three approaches for cost estimation consisting of multiple regression analysis, CBR and ANN have been compared, which finally CBR and ANN outperform MRA, and error associated with ANN is smaller. Also, CBR is the most appropriate model, due to fewer time-consuming features. The use of a BP ANN can be seen in [57], which is delicately applied to estimate the cost of structural systems of buildings and eventually the results have been compared with regression-based estimations, where the BPANN outweighs the other. Kim et al. [58] have also implemented BP-ANN, which has been improved through the application of GA algorithm. They have also compared the results of applying GA in order to omit the trial-and-error process of selecting the best ANN architecture with those of the model in the absence of GA, out of which, GA has shown an effective role in improving the model results.

Sonmez [59] have implemented RA and ANN in cost estimation of building care retirement community projects. They believe that there is not a distinct line between

these two methods, and none of them can be called superior to the other; however, they have investigated the for and against of both methods in their case study. Again, Multivariate regression analysis has been implemented in [60], while accompanied with factor analysis in order to select the best promising factors in the cost estimation process. They believe that the factors effective on cost model accuracy should be more explored. Besides, additional analysis is needed for circumstances, where new projects with new specifications and technologies are added to the project portfolio. In their point of view, different project factors can be taken into consideration, such as regional factors, project categorization, and so on to improve the model performance. Future research shall be conducted to study the cost factor's behaviour throughout the project lifetime. Emsley et al. [61] have also implemented ANN in addition to MRA and again factor analysis is implemented to help the process of retaining the best influencing factors in predicting construction cost. Setyawati et al. [62] have fully compared different situations under which, an ANN may perform better by including different inputs, different structures, data transformation, data preparation, size of dataset. Eventually, ANOVA¹ test has been implemented to investigate significant difference among four different input sets.

Besides, [63] have implemented BPANN for predicting the construction cost of school buildings by considering two proposed architectures, where the one, with more inputs outperforms the other. They claim that results that are more accurate stem from more data fed to the network in that neural networks are highly data driven. Boussabaine et al. [64] have presented an ANN approach developing 6 networks for different n (1 to 6) intervals of the project cash flow as completed intervals throughout the project and m (2 to 7 up to the end of the project) outputs as the remaining m intervals of the project, up to the completion of the project. Khosrowshahi et al. [65], have also implemented pure MRA to predict cost and time of housing projects in U.K. In this regard, they hope to generate a model, which is more general and can be applied to more diverse cases in terms of type, location, and so forth.

Cost estimation in highway projects has also been the main concern of some studies, which are scrutinized as follows. Mahalakshmi et al. [66] have estimated the cost of highway projects with the application of a multi perceptron neural network with the back-propagation learning algorithm. The model is composed of significant common cost factors such as topological index and project duration. In [67], a hybrid model of CBR and AHP is investigated in order to enhance the capabilities of CBR in many aspects,

¹ Analysis of variance.

such as improving the accuracy of the results, saving the time, and improving the performance of the model. They claim that, the use of more comprehensive dataset will lead to higher accuracy in results. In addition, the application of indexes related to geographical locations and cost factors should be taken into consideration. On the other hand, [68] have considered an expert system based upon a regression model in order to facilitate the process of transportation cost estimation. The novel approach in this study is the process of separating quantity from price for removing the need for considering regional factors. Thereafter, when the quantity is estimated, it'll be applied to unit price retrieved from an up-to-date database. Furthermore, [69] have implemented ANN in order to improve the estimation accuracy over conventional methods such as EVM.² They have considered several factors out of which, traffic volume, topography, weather conditions, evaluating date, contract duration, construction budget, percent of planned completion, and percent of actual completion, are assumed as the most effective parameters in the project cost. Wilmot and Mei [70] have also implemented and compared two models, including ANN model and Regression based model for forecasting highway construction cost and the associated escalation in a future, which finally shows the out-performance of the ANN model. In their point of view, factors such as facility (i.e. labour and price of equipment and material), the contract's characteristics (i.e. terms of payment, duration, geographical location), and overall contract terms (i.e. changes in specifications, amendments and so forth) are the most influencing factors on a cost estimation process.

Sodikov [71] have successfully implemented ANN in forecasting cost of highway construction and strongly advocate the ANN capabilities in being applied in uncertain circumstances and is used in early stages in projects. Further research is also needed to apply a hybrid of ANN with fuzzy logic, case-based reasoning, and so forth. Hegazy and Ayed [23] have developed an ANN model in this scope and optimized the corresponding weight through three different methods, including back-propagation training, simplex optimization, and applying genetic algorithm, out of which, simplex optimization surpasses the others. Their model is adaptive to new cases and can be compatible based on new circumstances. Adeli and Wu [72] have taken into consideration a regularization neural network, while a cost function composed of a standard error is applied and regularization error in order to simultaneously improve the network performance and prevent the network from being overfitted.

They defined public projects for their model as, whatever projects that are related to public sector, such as, schools, warehouses, hospitals, highways, bridges, water-related projects, and so on. In fact, the projects with such cases have been considered in this category.

Alshamrani [73] have considered cost estimation in building projects by taking advantage of regression analysis. Hyari et al. [74] in their work, they have developed an ANN model for cost estimation of engineering services through which, the influencing factors on a cost estimation process *are selected* via interview and literature review and further the best architecture of network is chosen after a process of trial and error. They desired to expand their model by feeding it with diverse datasets from different places worldwide; and also applying it to specific projects like bridges and schools that may increase its accuracy by confining the inherent variance in the input variables. Besides, [75] present a new method called Principal Item Ratios Estimating Method (PIREM), including parametric estimating, ratio's estimating, and cost significant model, which is capable of estimating costs under high fluctuations in prices, and it even can predict with least data available equal to only 20% of all cost factors.

Skitmore and Ng [76] have used a forward cross validation regression analysis to estimate time and cost in construction projects. They claim that, when the cost estimation model needs data such as the total amount of a contract, the accuracy of the cost estimation stems is derived from the accuracy of the total contract. Despite these limitations, the model can surpass the current risks and provide a practical tool in this scope. Bowen and Edwards [77] can also be regarded as a move from black-box mechanisms toward more logical and understandable methods like expert systems in the late twentieth century. They remind that the importance of historical data and expert's knowledge in cost estimation scope should not be disregarded. In addition, they desired to integrate a resource allocation system with the current cost model in the future.

Roadway projects are related to projects in the scope of paving roads, asphalt, and road-related works such as constructing bridges over roads, and mainly earth works. Few studies are done in this realm, which are as follows. In [78] the comparison of applying three different ANNs, including Multi-Layer Perceptron (MLP), General Regression Neural Network (GRNN), and RBFNN, has shown that GRNN is capable of estimating the cost of roadway projects with higher accuracy towards the two others. This type of neural network has shown outstanding performance, while encountering with incomplete datasets. They believe that a homogeneity in data set will also lead to better results in future researches in which, they have considered roadway projects with diverse specifications.

² Earned value management.

Swei et al. [79] have applied an integration of a Maximum Likelihood (ML) and Least Angle Regression (LAR) to estimate the cost of road pavement. They suggest that more cost inputs can be taken into consideration, in the model for the future. They also recommend that their model can be implemented by using actual cost rather than a bid price for further studies. Besides, the use of other methods such as regression analysis is also proposed.

Peško et al. [80] have considered comparing ANN and SVM capabilities in cost estimation for construction of urban roads out of which, SVM has shown superior result compared to ANN. They claim for more expanded database in the future researches. Also, they raise the need for a cost model that is capable of estimating at very early stage of the project for management purposes. [81], for instance, have taken advantage of simulation in management with the use of stochastic models and Monte Carlo simulation. On the other hand, [82], have applied CBR and GA for cost estimation of bridge construction projects. On the other hand, [83] have probed the application of ANN in cost estimation of bridge repair and maintenance projects compared to work package methods, which finally the results of ANN are more outstanding. A conclusion is drawn that the model performs well at the early stages of the project, and a hybrid of the current method with up-to-date techniques in general, and fuzzy logic in particular are recommended.

Water-related projects, here are referred to whatever project beyond the scope of water, sewer installation services, and so on. Cost estimation in this type of projects is less investigated, which are studied as follows. Marzouk and Elkadi [84] have determined variables effective on cost estimation process and conducted a survey to implement a factor variable reduction through Exploratory Factor Analysis (EFA). Eventually, the best ANN is selected from different architectures with an error almost equal to 22%. ANN has also been the main concern for cost estimation in [85], since it is capable of tackling non-linearity in early stages of projects. Furthermore, [86] have used ANN model to predict water and sewer services construction cost and selected the best network architecture based on trial and error through, which have reached an accuracy of 80%. They claim that one of the drawbacks of their model is the lack of regional factors, which can be effective in improving the performance and accuracy of the current model.

In roads, tunnels project area, two types of neural networks have been implemented in [87], and the results have been compared with those of multiple regression analysis, out of which neural networks show better performance. Their model can be implemented in other types of buildings as well. However, as they claim, the model needs to be updated to be compatible to newly complete and added

projects to their database. Petroutsatou and Lambropoulos [88], on the other hand, have approached the construction cost estimation via application of a Structural Equation Model (SEM) and compared the results with ANN and RA models, which SEM performs better. They try to follow their research in future to be able to predict the project profit and schedule programming as well as project cost.

On the other hand, in railway project's scope, comparison between MRA and ANN in estimating the cost of light rail transit and metro track works can be seen in [89] out of which, the MRA result is superior to the other due to the small number of available instances. This shows that the higher the number of the cases the higher will be the ANN results' accuracy. Besides, the work of [90] has depicted the effect of GA on optimizing CBR attributes weights for estimating the cost of railway bridge projects.

Recently, [15] have investigated forecasting hydroelectric power plant project's cost via ANN through which, three different architectures have been generated and examined, while seeking the best performance. The results have been compared with those of RA and concluded that the ANN shows better promising results. They set forth that, the model can be used for different parts of hydroelectric power plant projects as well. Singal et al. [91] probe RA in their study and compared the results with actual cases to validate their model.

Hashemi et al. [92] have thoroughly investigated effective parameters in cost estimation of power plant projects, while simultaneously considering risk in these projects by embedding PERT technique. The sensitivity analysis conducted in this research shows that the type of power plant is the most influencing factor in the model inputs. This data is finally fed into the hybrid of ANN and GA, to estimate the cost of these types of projects with an accuracy equal to 94.71%.

Hence, the determinative role of ANNs is highlighted again in Fig. 4. Afterwards, as it is mentioned before, RA is the most powerful method applied in cost estimation studies. As it can be seen, SVM, PSO, RBFNN, and Fuzzy ANN have been used only in building projects.

Figure 5 shows the distribution of these methods and as it is depicted so, ANNs have the first ranking among all methods. This strongly shows the power of Neural Networks as the artificial intelligence tool to deal with estimating problems. Further, Regression Analysis stands on the second step as an outstanding tool in the field of parametric methods.

3.4 Distributions and techniques

Studies on the distribution of the cost estimation techniques suggest the need for categorization. These techniques are based on the studied papers considered as an

Fig. 4 Application area versus method applied

Fig. 5 Applications of different methods in construction cost estimation studies

Fig. 6 Distributions of articles by approach types

analogous, analytical, parametric and intuitive approach. To continue analysing the reviewed papers, based on these criteria, the results have been represented in "Appendix 2". The summary of "Appendix 2" is illustrated in Fig. 6. As the result shown in Fig. 6, most of the adopted techniques belong to the analogous category, and the least one is the analytical one, which is the decision tree method adopted in [43].

In the construction cost estimation, the qualitative model confides in the specialist judgment or heuristic and mathematical rules. The qualitative models can classify into statistical, intuitive, and analytical models. On the other hand, quantitative models can categorize into three main techniques of analogous, parametric, and analogy-based models. Among all the methods applied to the proposal techniques, only 2% of them are qualitative, which belong to intuitive methods such as AHP. Therefore, based on this result, the rests of the studies are done based on quantitative approaches (Fig. 7).

One step further, the sub domains of each approach type are shown in Fig. 8 i.e. intuitive, analytical, analogous, and parametric. As shown in this figure, analytical methods such as decision trees have the least proportion of all methods applied. These categories are also shown in each application area and presented in Fig. 8.

Moreover, the distribution of these approaches in the time spectrum is shown in Fig. 9. As it is presented analogous approaches have the most portion of studies conducted. Total categories of cost estimation methods applied in cost estimation of construction projects can also be seen in "Appendix 3".

3.5 Journals

Table 3 also summarizes the papers reviewed by their journals, and the journal's portion of total.

As shown in the above table, the Journal of Construction Engineering and Management, Building and environment, Construction Management and Economics, Expert System with Applications are the top journals with the most published papers in construction cost estimation scope.

3.6 Year of publication

Figure 10 has also depicted the distribution of cost estimation studies in years. As shown in this figure, a smooth growth has been occurred in years 2009 until 2011, and 2017 until 2019, after a decline in years 2006, and 2007. However, again a diminution has been observed afterwards until 2016. Furthermore, as it presents, the most

Fig. 7 Approach type distribution **Fig. 8** Approach types applied in different areas

studies have been done via ANN as a powerful machine learning technique.

As it has been contemplated more in diverse applied methods, the ANNs' contribution to cost estimation problems observed in Fig. 11. Cost models, expert systems,

AHP,³ CBR, Monte Carlo, fuzzy logic, and decision tree methods are all summarized as other methods in this diagram.

³ Analytical hierarchy process.

Fig. 9 Approaches implemented in time horizon

4 Conclusion

Cost estimation in construction projects has been reviewed in articles published within years from 1985 to 2020. A conclusion is drawn that in almost all the cases, estimating at the very early stage of the project is of a great concern. Most of the proposed estimation techniques tried to meet the expectation by generating models to be applied at even tendering level to help process of decision-making of managers. Fundamentally, effective cost factors shall be explored and scrutinized exactly. Not only, the effective cost factors should be studied, but also the factors affecting the cost model accuracy must be reviewed in deep. One of the cost factors that have been noted repeatedly is the regional factor, which shows the importance of differentiating between projects with diverse geographical origin. Additionally, the ability of the model to expand generally and the applicability to novel cases has the high degree of importance.

As shown by results, among the various methods (ANN, Fuzzy NN, SVM, PSO, RBF, RA, CBR, PSO, Decision Tree, AHP, Monte Carlo, fuzzy logic) used by researchers, the most popular machine learning techniques that used in the reviewed papers are ANN and RA respectively. In contrast to other methods, the ANN and RA are the most popular and successful methods implemented in these

studies respectively. However, the hybrid models of ANN with fuzzy logic, CBR, GA and so forth have surpassed the mere ANN applied. The point that shall be considered in ANN application is its sensitivity to input data. Since this machine learning technique is data driven, it will perform more accurately, if a large amount of data and homogenous dataset exists to extract relations between available data. On the other hand, the number of input neurons (known as cost factors), has a direct effect on system malfunction. Accordingly, when the number of input cost factors increases, the complexity of the system will increase and in case of construction cost estimation, it showed the accuracy of the estimation will decrease. This study finds out in the hidden layer, the number of neurons and the corresponding weights have a direct effect on the generalization ability of the model. Indeed, the number of factors is important rather than learning parameters, and it directly affects the estimation model accuracy. In addition, ANN is known as a powerful model in tackling with nonlinear problems. Tuning the ANN parameters, such as the number of hidden factors and weights have also been the concern of many studies, which have been overcome by combining it with GA algorithm. Nevertheless, the expert knowledge to select cost factors in the estimation models has a valuable influence. Furthermore, the building and highway projects assign the most attention

Table 3 Journals and related articles

Journal	Papers	Total (%)
AACE International Transactions	1	1.54
Advanced Materials Research	2	3.08
Advances in Civil Engineering	1	1.54
American Journal of Applied Sciences	1	1.54
Automation in Construction	3	4.62
Building and Environment	6	9.23
Canadian Journal of Civil Engineering	2	3.08
Civil Engineering Journal	1	1.54
Complexity	1	1.54
Construction Management and Economics	5	7.69
Energy for Sustainable Development	1	1.54
Engineering, Construction and Architectural Management	1	1.54
Expert Systems with Applications	4	6.15
Facilities	1	1.54
Intelligent Information Technology Application	1	1.54
International Journal of Civil and Structural Engineering	1	1.54
International Journal of Emerging Technology and Advanced Engineering	1	1.54
International Journal of Project Management	3	4.62
International Journal of Science and Research (IJSR)	1	1.54
Journal of Artificial Intelligence	1	1.54
Journal of Civil Engineering and Management	2	3.08
Journal of Cleaner Production	1	1.54
Journal of Computing in Civil Engineering	2	3.08
Journal of Construction Engineering and Management	8	12.31
Journal of Construction Engineering and Project Management	1	1.54
Journal of Discrete Mathematical Sciences and Cryptography	1	1.54
Journal of Engineering, Design and Technology	2	3.08
Journal of Management in Engineering	1	1.54
Journal of Soft Computing in Civil Engineering		1.54
Journal of Taibah University for Science	1	1.54
Journal of the Eastern Asia Society for Transportation Studies	1	1.54
KSCE Journal of Civil Engineering	1	1.54
Neural Computing and Applications	2	3.08
Operational Research	1	1.54
Sustainable Construction and Building Materials	1	1.54
Transportation Research Part B: Methodological	1	1.54

of the researchers to themselves in cost estimation studies. Among these studies, the methods have been categorized based on their approach, including intuitive, parametric, analogous, and analytical, which the most studies belong to the analogous group.

This study provides several guidelines for applying machine learning models in construction projects as follows: (1) understand the fundamental and validation of machine learning models and cooperate with existing applications and models; (2) select the best models, which ability is well matched with the research impacts and goals; (3) construct the dataset priority for proposal machine learning models and check the sufficiency and

efficiency of the dataset; (4) parallel use of machine learning models with current or ordinary models at the early stage of a project; and (5) find the project priority of factors and required datasets in the research association.

The limitations of this research paper can be summarised as: (a) data is collected from Google Scholars and Science Direct scientific database, therefore, the articles did not cite in these two databases did not consider in the study as well; (b) the study had the limitation of exploring the English language papers in the cost estimation for construction projects domain only and not considered the other languages.

Fig. 10 Distributions of applied cost estimation methods in years

Fig. 11 Proportions of each method studied in time horizon

Based on this study, deep-learning techniques did not get attention of researchers in the field of cost estimation for construction projects; therefore, this systematic review suggests these techniques and models for future propose work and study.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix 1: Reviewed articles by methods used

Row	References	Year	First Author	Area	Method1	Method2	Method3	Method4
2	[78]	2019	Ksenija Tijanic	Road way projects	ANN			
3	[1]	2019	Sae-Hyun Ji	Building projects	Cost Model	CBR		
4	[66]	2019	G. Mahalakshmi	Highway projects	ANN			
5	[29]	2019	V.B. Chandan-shive1	Building projects	ANN			
6	[28]	2019	Qinghua Jiang	Building projects	ANN	RBF		
7	[92]	2017	Sanaz Tayefeh Hashemi	Power plant projects	ANN			
8	[30]	2017	Xinzheng Wang*	Building projects	Fuzzy Logic			
9	[79]	2017	Omar Swei	Road way projects	RA			
10	[73]	2017	Othman Subhi Alshamrani	Public projects	RA			
11	[31]	2017	Zeyan Du	Building projects	ANN			
12	[33]	2017	Abbas Mohammed Burhan Alshemosi	Building projects	RA			
13	[32]	2017	Mawardi Amin	Building projects	RA			
14	[80]	2017	Igor Peško	Road way projects	ANN	SVM		
15	[84]	2016	Marzouk	Water-related construction projects	ANN			
16	[15]	2015	Gunduz	Hydropower projects	ANN	RA		
17	[74]	2015	Hyari	Public projects	ANN			
18	[34]	2014	Bala	Building projects	ANN			
19	[35]	2014	El-Sawalhi	Building projects	ANN			
20	[36]	2014	Roxas	Building projects	ANN			
21	[39]	2013	Jafarzadeh	Building projects	ANN			
22	[37]	2013	Feng	Building projects	ANN			
23	[38]	2013	Kim	Building projects	CBR			
24	[67]	2013	Kim	Highway projects	CBR	AHP		
25	[43]	2012	Son	Building projects	SVM	ANN	DT	RA
26	[41]	2012	El Sawalhi	Building projects	Fuzzy Logic			
27	[42]	2012	Jin	Building projects	CBR			
28	[85]	2012	Ahiaga-Dagbui	Water-related construction projects	ANN			
29	[40]	2012	Wang	Building Projects	SVM	ANN		
30	[44]	2011	Ji	Building projects	CBR			
31	[46]	2011	He	Building projects	Fuzzy NN			
32	[45]	2011	Tatari	Building projects	ANN	RA		
33	[81]	2011	Chou	Road way projects	Monte Carlo			
34	[87]	2011	Petroutsatou	Road tunnel projects	ANN	RA		

Row	References	Year	First Author	Area	Method1	Method2	Method3	Method4
35	[47]	2011	Arafa	Building projects	ANN			
36	[89]	2011	Gunduz	Railway projects	RA	ANN		
37	[90]	2011	Kim	Railway projects	CBR			
38	[91]	2010	Singal	Hydropower projects	RA			
39	[48]	2010	Cheng	Building projects	Fuzzy NN			
40	[82]	2010	Kim	Road way projects	CBR			
42	[86]	2009	Alex	Water-related construction projects	ANN			
43	[69]	2009	Pewdum	Highway projects	ANN			
44	[68]	2009	Chou	Highway projects	Expert System			
45	[49]	2009	Cheng	Building Projects	fuzzy NN			
46	[83]	2008	Bouabaz	Road way projects	ANN			
47	[50]	2008	Lai	Building projects	AHP			
48	[51]	2008	Shi	Building projects	ANN			
49	[52]	2007	An	Building projects	CBR	AHP		
50	[75]	2006	Yu	Public projects	Cost Model			
51	[71]	2005	Sodikov	Highway projects	ANN	RA		
52	[55]	2005	Li	Building projects	RA			
53	[53]	2005	Kim	Building Projects	ANN			
54	[70]	2005	Wilmot	Highway projects	ANN	RA		
55	[54]	2005	Chan	Building projects	RA			
56	[57]	2004	Gunaydin	Building projects	ANN	RA		
57	[56]	2004	Kim	Building projects	ANN	CBR	RA	
58	[59]	2004	Sonmez	Building projects	ANN	RA		
59	[58]	2004	Kim	Building projects	ANN			
60	[76]	2003	Skitmore	Public projects	RA			
61	[60]	2003	Trost	Building projects	RA			
62	[61]	2002	Emsley	Building projects	ANN	RA		
63	[62]	2002	Setyawati	Building projects	ANN			
64	[64]	1998	Boussabaine	Building projects	ANN			
65	[63]	1998	Elhag	Building projects	ANN			
66	[23]	1998	Hegazy	Highway projects	ANN			
67	[72]	1998	Adeli	Highway projects	ANN			
68	[65]	1996	Khosrowshahi	Building projects	RA			
69	[77]	1985	Bowen	Public projects	Expert system			

Appendix 2: Reviewed articles by approach type implemented

Row	References	Year	First Author	Area	Approach Type 1	Approach Type 2	Approach Type 3	Approach Type 4
1	[27]	2020	Tarq Zaed Khalaf	Building projects	Analogous			
2	[78]	2019	Ksenija Tijanic	Road way projects	Analogous			
3	[1]	2019	Sae-Hyun Ji	Building projects	Parametric	Analogous		
4	[66]	2019	G. Mahalakshmi	Highway projects	Analogous			

Row	References	Year	First Author	Area	Approach Type 1	Approach Type 2	Approach Type 3	Approach Type 4
5	[29]	2019	V.B. Chandan-shive1	Building projects	Analogous			
6	[28]	2019	Qinghua Jiang	Building projects	Analogous	Analogous		
7	[92]	2017	Sanaz Tayefeh Hashemi	Power plant projects	Analogous			
8	[30]	2017	Xinzheng Wang*	Building projects	Analogous			
9	[79]	2017	Omar Swei	Road way projects	Parametric			
10	[73]	2017	Othman Subhi Alshamrani	Public projects	Parametric			
11	[31]	2017	Zeyan Du	Building projects	Analogous			
12	[33]	2017	Abbas Mohammed Burhan Alshemosi	Building projects	Parametric			
13	[32]	2017	Mawardi Amin	Building projects	Parametric			
14	[80]	2017	Igor Peško	Road way projects	Analogous	Analogous		
15	[84]	2016	Marzouk	Water-related construction projects	Analogous			
16	[15]	2015	Gunduz	Hydropower projects	Analogous	Parametric		
17	[74]	2015	Hyari	Public projects	Analogous			
18	[34]	2014	Bala	Building projects	Analogous			
19	[35]	2014	El-Sawalhi	Building projects	Analogous			
20	[36]	2014	Roxas	Building projects	Analogous			
21	[39]	2013	Jafarzadeh	Building projects	Analogous			
22	[37]	2013	Feng	Building projects	Analogous			
23	[38]	2013	Kim	Building projects	Analogous			
24	[67]	2013	Kim	Highway projects	Analogous	Intuitive		
25	[43]	2012	Son	Building projects	Analogous	Analogous	Analytical	Parametric
26	[41]	2012	El Sawalhi	Building projects	Analogous			
27	[42]	2012	Jin	Building projects	Analogous			
28	[85]	2012	Ahiaga-Dagbui	Water-related construction projects	Analogous			
29	[40]	2012	Wang	Building projects	Analogous	Analogous		
30	[44]	2011	Ji	Building projects	Analogous			

Row	References	Year	First Author	Area	Approach Type 1	Approach Type 2	Approach Type 3	Approach Type 4
31	[46]	2011	He	Building projects	Analogous			
32	[45]	2011	Tatari	Building projects	Analogous	Parametric		
33	[81]	2011	Chou	Road way projects	Parametric			
34	[87]	2011	Petroutsatou	Road tunnel projects	Analogous	Parametric		
35	[47]	2011	Arafa	Building projects	Analogous			
36	[89]	2011	Gunduz	Railway projects	Parametric	Analogous		
37	[90]	2011	Kim	Railway projects	Analogous			
38	[91]	2010	Singal	Hydropower projects	Parametric			
39	[48]	2010	Cheng	Building projects	Analogous			
40	[82]	2010	Kim	Road way projects	Analogous			
41	[88]	2010	Petroutsatou	Road tunnel projects	Parametric			
43	[69]	2009	Pewdum	Highway projects	Analogous			
44	[68]	2009	Chou	Highway projects	Analogous			
45	[49]	2009	Cheng	Building projects	Analogous			
46	[83]	2008	Bouabaz	Road way projects	Analogous			
47	[50]	2008	Lai	Building projects	Intuitive			
48	[51]	2008	Shi	Building projects	Analogous			
49	[52]	2007	An	Building projects	Analogous	Intuitive		
50	[75]	2006	Yu	Public projects	Parametric			
51	[71]	2005	Sodikov	Highway projects	Analogous	Parametric		
52	[55]	2005	Li	Building projects	Parametric			
53	[53]	2005	Kim	Building projects	Analogous			
54	[70]	2005	Wilmot	Highway projects	Analogous	Parametric		
55	[54]	2005	Chan	Building projects	Parametric			
56	[57]	2004	Gunaydin	Building projects	Analogous	Parametric		
57	[56]	2004	Kim	Building projects	Analogous	Analogous	Parametric	
58	[59]	2004	Sonmez	Building projects	Analogous	Parametric		
59	[58]	2004	Kim	Building projects	Analogous			

Row	References	Year	First Author	Area	Approach Type 1	Approach Type 2	Approach Type 3	Approach Type 4
60	[76]	2003	Skitmore	Public projects	Parametric			
61	[60]	2003	Trost	Building projects	Parametric			
62	[61]	2002	Emsley	Building projects	Analogous	Parametric		
63	[62]	2002	Setyawati	Building projects	Analogous			
64	[64]	1998	Boussabaine	Building projects	Analogous			
65	[63]	1998	Elhag	Building projects	Analogous			
66	[23]	1998	Hegazy	Highway projects	Analogous			
67	[72]	1998	Adeli	Highway projects	Analogous			
68	[65]	1996	Khosrowshahi	Building projects	Parametric			
69	[77]	1985	Bowen	Public projects	Analogous			

Appendix 3: Categorization of cost estimation methods applied in construction projects

Quantitative	Parametric				
	RA		Cost Model	Monte Carlo	
<div> </div>
	Analogous				
	ANN	Fuzzy NN	CBR	Expert System	SVM

	Analytical				
	Decision Tree

Qualitative	Intuitive				
	AHP
<div> <p>© bpmsg.com</p></div>				

References

- Ji S-H et al (2019) Cost estimation model using modified parameters for construction projects. *Adv Civ Eng* 2019:1–10
- Aamodt A, Plaza E (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Commun* 7(1):39–59
- Hegazy T (2002) Computer-based construction project management. Prentice Hall, Upper Saddle River
- Alpaydin E (2014) Introduction to machine learning. MIT Press, Cambridge
- Graupe D (2013) Principles of artificial neural networks, vol 7. World Scientific, Singapore
- Anderson D, McNeill G (1992) Artificial neural networks technology. *Kaman Sci Corp* 258:1–83
- Hsu KL, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall-runoff process. *Water Resour Res* 31(10):2517–2530
- Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 49(11):1225–1231
- Bjornson C, Barney DK (1999) Identifying significant model inputs with neural networks: tax court determination of reasonable compensation. *Expert Syst Appl* 17(1):13–19
- Bai Z et al (2014) Predictive model of energy cost in steelmaking process based on BP neural network. In: 2nd international conference on software engineering, knowledge engineering and information engineering (SEKEIE). Singapore
- Tkáč M, Verner R (2016) Artificial neural networks in business: two decades of research. *Appl Soft Comput* 38:788–804
- Mitchell M (1998) An introduction to genetic algorithms. MIT Press, Cambridge
- Back T, Fogel DB, Michalewicz Z (1997) Handbook of evolutionary computation. IOP Publishing Ltd., Bristol
- Yang J-B, Yau N-J (1996) Applications of case-based reasoning in construction engineering and management. In: Computing in civil engineering, vol 15. ASCE, pp 663–669
- Gunduz M, Sahin HB (2015) An early cost estimation model for hydroelectric power plant projects using neural networks and multiple regression analysis. *J Civ Eng Manag* 21(4):470–477
- Djukova EV, Peskov N (2007) A classification algorithm based on the complete decision tree. *Pattern Recognit Image Anal* 17(3):363–367
- Loh WY (2008) Classification and regression tree methods. In: Encyclopedia of statistics in quality and reliability. Wiley, pp 1–13
- Perner P (2015) Decision tree induction methods and their application to big data. In: Xhafa F, Barolli L, Barolli A, Papajorgji P (eds) Modeling and processing for next-generation big-data technologies. Springer, Cham, pp 57–88
- Dym CL, Levitt RE (1991) Knowledge-based systems in engineering. McGraw-Hill Book Company, New York
- Layer A et al (2002) Recent and future trends in cost estimation. *Int J Comput Integr Manuf* 15(6):499–510
- Rad PF (2002) Project estimating and cost management. Management Concepts Incorporated, Virginia
- Qian L, Ben-Arieh D (2008) Parametric cost estimation based on activity-based costing: a case study for design and development of rotational parts. *Int J Prod Econ* 113(2):805–818
- Hegazy T, Ayed A (1998) Neural network model for parametric cost estimation of highway projects. *J Constr Eng Manag* 124(3):210–218
- Duverlie P, Castelain J (1999) Cost estimation during design step: parametric method versus case based reasoning method. *Int J Adv Manuf Technol* 15(12):895–906
- Caputo AC, Pelagagge PM (2008) Parametric and neural methods for cost estimation of process vessels. *Int J Prod Econ* 112(2):934–954
- Elfaki AO, Alatawi S, Abushandi E (2014) Using intelligent techniques in construction project cost estimation: 10-year survey. *Adv Civ Eng* 2014:1–11
- Khalaf TZ et al (2020) Particle swarm optimization based approach for estimation of costs and duration of construction projects. *Civ Eng J* 6(2):384–401
- Jiang Q (2019) Estimation of construction project building cost by back-propagation neural network. *J Eng Des Technol* 18(3):601–609
- Chandanshive V, Kambekar AR (2019) Estimation of building construction cost using artificial neural networks. *J Soft Comput Civ Eng* 3(1):91–107
- Wang X (2017) Application of fuzzy math in cost estimation of construction project. *J Discrete Math Sci Cryptogr* 20(4):805–816
- Du Z, Li B (2017) Construction project cost estimation based on improved BP Neural Network. In: 2017 international conference on smart grid and electrical automation (ICSGEA). IEEE
- Amin M (2017) Development of cost estimation model for residential building. *Int J Civ Struct Eng Res* 5(1):1–4
- Alshemosi AMB, Alsaad HSH (2017) Cost estimation process for construction residential projects by using multifactor linear regression technique. *Int J Sci Res* 6(6):151–156
- Bala K, Ahmad Bustani S, Shehu Waziri B (2014) A computer-based cost prediction model for institutional building projects in Nigeria: an artificial neural network approach. *J Eng Des Technol* 12(4):519–530
- El-Sawalhi NI, Shehatto O (2014) A neural network model for building construction projects cost estimating. *KICEM J Constr Eng Proj Manag* 4(4):9–16
- Roxas CLC, Ongpeng JMC (2014) An artificial neural network approach to structural cost estimation of building projects in the Philippines. In: DLSU Research Congress, vol 2, no 2, pp 1–8
- Feng GL, Li L (2013) Application of genetic algorithm and neural network in construction cost estimate. In: Advanced materials research. Trans Tech Publ 756:3194–3198
- Kim S, Shim JH (2013) Combining case-based reasoning with genetic algorithm optimization for preliminary cost estimation in construction industry. *Can J Civ Eng* 41(1):65–73
- Jafarzadeh R et al (2013) Application of artificial neural network methodology for predicting seismic retrofit construction costs. *J Constr Eng Manag* 140(2):04013044
- Wang Y-R, Yu C-Y, Chan H-H (2012) Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models. *Int J Project Manag* 30(4):470–478
- El Sawalhi NI (2012) Modelling the parametric construction project cost estimate using fuzzy logic. *Int J Emerg Technol Adv Eng Website* 2(4) ISSN 2250-2459
- Jin R et al (2012) MRA-based revised CBR model for cost prediction in the early stage of construction projects. *Expert Syst Appl* 39(5):5214–5222
- Son H, Kim C, 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. *Autom Constr* 27:60–66
- Ji S-H, Park M, Lee H-S (2011) Case adaptation method of case-based reasoning for construction cost estimation in Korea. *J Constr Eng Manag* 138(1):43–52
- Tatari O, Kucukvar M (2011) Cost premium prediction of certified green buildings: a neural network approach. *Build Environ* 46(5):1081–1086
- He X et al (2011) Cost estimation of construction project using fuzzy neural network model embedded with modified particle

- optimizer. In: Advanced materials research. Trans Tech Publ 243:6296–6301
47. Arafat M, Alqedra M (2011) Early stage cost estimation of buildings construction projects using artificial neural networks. *J Artif Intell* 4(1):63–75
 48. Cheng M-Y, Tsai H-C, Sudjono E (2010) Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Syst Appl* 37(6):4224–4231
 49. Cheng M-Y, Tsai H-C, Hsieh W-S (2009) Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. *Autom Constr* 18(2):164–172
 50. Lai Y-T, Wang W-C, Wang H-H (2008) AHP-and simulation-based budget determination procedure for public building construction projects. *Autom Constr* 17(5):623–632
 51. Shi H, Li W (2000) The integrated methodology of rough set theory and artificial neural-network for construction project cost prediction. In: Second international symposium on intelligent information technology application, 2008. IITA'08. IEEE
 52. An S-H, Kim G-H, Kang K-I (2007) A case-based reasoning cost estimating model using experience by analytic hierarchy process. *Build Environ* 42(7):2573–2579
 53. Kim G, Seo D, Kang KI (2005) Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates. *J Comput Civ Eng* 19(2):208–211
 54. Chan SL, Park M (2005) Project cost estimation using principal component regression. *Constr Manag Econ* 23(3):295–304
 55. Li H, Shen Q, Love PE (2005) Cost modelling of office buildings in Hong Kong: an exploratory study. *Facilities* 23(9/10):438–452
 56. Kim G-H, An S-H, Kang K-I (2004) Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Build Environ* 39(10):1235–1242
 57. Günaydin HM, Doğan SZ (2004) A neural network approach for early cost estimation of structural systems of buildings. *Int J Proj Manag* 22(7):595–602
 58. Kim G-H et al (2004) Neural network model incorporating a genetic algorithm in estimating construction costs. *Build Environ* 39(11):1333–1340
 59. Sonmez R (2004) Conceptual cost estimation of building projects with regression analysis and neural networks. *Can J Civ Eng* 31(4):677–683
 60. Trost SM, Oberlender GD (2003) Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *J Constr Eng Manag* 129(2):198–204
 61. Emsley MW et al (2002) Data modelling and the application of a neural network approach to the prediction of total construction costs. *Constr Manag Econ* 20(6):465–472
 62. Setyawati BR, Sahirman S, Creese RC (2002) Neural networks for cost estimation. *AACE International Transactions*, Durham, p ES131
 63. Elhag T, Boussabaine A (1998) An artificial neural system for cost estimation of construction projects. In: Proceedings of the 14th ARCOM annual conference
 64. Boussabaine A, Kaka A (1998) A neural networks approach for cost flow forecasting. *Constr Manag Econ* 16(4):471–479
 65. Khosrowshahi F, Kaka AP (1996) Estimation of project total cost and duration for housing projects in the UK. *Build Environ* 31(4):375–383
 66. Mahalakshmi G, Rajasekaran C (2019) Early cost estimation of highway projects in India using artificial neural network. In: Das B, Neithalath N (eds) Sustainable construction and building materials. Springer, Singapore, pp 659–672
 67. Kim S (2013) Hybrid forecasting system based on case-based reasoning and analytic hierarchy process for cost estimation. *J Civ Eng Manag* 19(1):86–96
 68. Chou J-S (2009) Generalized linear model-based expert system for estimating the cost of transportation projects. *Expert Syst Appl* 36(3):4253–4267
 69. Pewdum W, Rujiranyong T, Sooksatra V (2009) Forecasting final budget and duration of highway construction projects. *Eng Constr Arch Manag* 16(6):544–557
 70. Wilmot CG, Mei B (2005) Neural network modeling of highway construction costs. *J Constr Eng Manag* 131(7):765–771
 71. Sodikov J (2005) Cost estimation of highway projects in developing countries: artificial neural network approach. *J East Asia Soc Transp Stud* 6:1036–1047
 72. Adeli H, Wu M (1998) Regularization neural network for construction cost estimation. *J Constr Eng Manag* 124(1):18–24
 73. Alshamrani OS (2017) Construction cost prediction model for conventional and sustainable college buildings in North America. *J Taibah Univ Sci* 11(2):315–323
 74. Hyari KH, Al-Daraiseh A, El-Mashaleh M (2015) Conceptual cost estimation model for engineering services in public construction projects. *J Manag Eng* 32(1):04015021
 75. Yu WD (2006) PIREM: a new model for conceptual cost estimation. *Constr Manag Econ* 24(3):259–270
 76. Skitmore RM, Ng ST (2003) Forecast models for actual construction time and cost. *Build Environ* 38(8):1075–1083
 77. Bowen P, Edwards P (1985) Cost modelling and price forecasting: practice and theory in perspective. *Constr Manag Econ* 3(3):199–215
 78. Tijanić K, Car-Pušić D, Šperac M (2020) Cost estimation in road construction using artificial neural network. *Neural Comput Appl* 32:9343–9355
 79. Swei O, Gregory J, Kirchain R (2017) Construction cost estimation: a parametric approach for better estimates of expected cost and variation. *Transp Res Part B Methodol* 101:295–305
 80. Peško I et al (2017) Estimation of costs and durations of construction of urban roads using ANN and SVM. *Complexity* 2017:1–13
 81. Chou J-S (2011) Cost simulation in an item-based project involving construction engineering and management. *Int J Proj Manag* 29(6):706–717
 82. Kim KJ, Kim K (2010) Preliminary cost estimation model using case-based reasoning and genetic algorithms. *J Comput Civ Eng* 24(6):499–505
 83. Bouabaz M, Hamami M (2008) A cost estimation model for repair bridges based on artificial neural network. *Am J Appl Sci* 5(4):334–339
 84. Marzouk M, Elkadi M (2016) Estimating water treatment plants costs using factor analysis and artificial neural networks. *J Clean Prod* 112:4540–4549
 85. Ahiaga-Dagbui DD, Smith SD (2012) Neural networks for modelling the final target cost of water projects. In: Smith SD (ed) *Procs 28th Annual ARCOM Conference*, vol 2, no 1, pp 307–316
 86. Alex DP et al (2009) Artificial neural network model for cost estimation: City of Edmonton's water and sewer installation services. *J Constr Eng Manag* 136(7):745–756
 87. Petroutsatou K et al (2011) Early cost estimating of road tunnel construction using neural networks. *J Constr Eng Manag* 138(6):679–687
 88. Petroutsatou K, Lambropoulos S (2010) Road tunnels construction cost estimation: a structural equation model development and comparison. *Oper Res Int J* 10(2):163–173
 89. Gunduz M, Ugur LO, Ozturk E (2011) Parametric cost estimation system for light rail transit and metro trackworks. *Expert Syst Appl* 38(3):2873–2877
 90. Kim BS (2011) The approximate cost estimating model for railway bridge project in the planning phase using CBR method. *KSCSE J Civ Eng* 15(7):1149–1159

91. Singal S, Saini R, Raghuvanshi C (2010) Analysis for cost estimation of low head run-of-river small hydropower schemes. *Energy Sustain Dev* 14(2):117–126
92. Hashemi ST, Ebadati EOM, Kaur H (2017) A hybrid conceptual cost estimating model using ANN and GA for power plant projects. *Neural Comput Appl* 31:2143–2154

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.