INTERNATIONAL JOURNAL OF CONSTRUCTION MANAGEMENT
https://doi.org/10.1080/15623599.2019.1692400

Taylor & Francis
Taylor & Francis Group

An artificial neural network approach for cost estimation of engineering services

Erik Matel[a], Faridaddin Vahdatikhaki[a], Siavash Hosseinyalamdary[c], Thijs Evers[b] and Hans Voordijk[a]

[a]Department of Construction Management and Engineering, University of Twente, Enschede, The Netherlands; [b]Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands; [c]Coordinator of Tender Management, Bilfinger Tebodin, Hengelo, The Netherlands

ABSTRACT
In a globally competitive world, with diminishing profit margins and decreasing market shares, the cost of a project is one of the major criteria in decision making at the early stages of a building design process in the construction industry. To remain competitive in the market, it is crucial for companies to have an accurate estimate of their projects. Nevertheless, given that very little is known about the scope and details of the project, the conventional cost estimation methods tend to be slow and inaccurate. With the rise of computing power, there is now a tendency to use Machine Learning (ML)-based methods, such as Artificial Neural Networks (ANNs), for more accurate cost estimation that can remain reliable in face of insufficient details during the tendering phase. While the use of ANN for cost estimation has been abundantly investigated from the perspective of contractors, there are very limited studies on the development and application of ML-based methods for engineering consultancy firms. Given that the nature of products/services offered by consultancy firms is inherently different from that of contractors (i.e. they are more abstract and less material-based) and also given that the type and level of detail of the available data at the tendering stage is dissimilar, it is important to investigate the applicability of ML-based methods for cost estimation in consultancy firms. To this end, this paper presents an artificial neural network approach for the cost estimation of engineering services. In developing the model, first, the influential factors that affect the costs of engineering services are identified. Thereafter, a model is developed using the data of 132 projects. Subsequently, a heuristic method is developed to systematically improve and fine-tune the performance of the model. Eventually, the findings show that artificial neural networks (ANNs) can obtain a fairly accurate cost estimate, even with small datasets. In fact, the model proposed in this paper performed better than those proposed in other similar works. The model developed in this study showed a 14.5% improvement in the accuracy of the model, considering MAPE.

KEYWORDS
Machine learning; artificial neural networks; cost estimation; engineering services

Introduction

A cost estimate of capital expenditures in the tendering phase of a project greatly influences planning, bidding, design, construction management and cost management (Arage and Dharwadkar 2017). Decisions based on cost estimates commonly lead to resource allocation and other types of major commitments, which may have critical consequences. Cost estimates allow project managers to evaluate the feasibility of projects and control costs effectively. Furthermore, the estimate may influence the client's decision on whether or not to progress with the project (Ahiaga-Dagbui and Smith 2012). In addition, for many clients, completing the project within the predefined budget is a paramount determinant of client satisfaction. Therefore, inaccurate estimates of costs can have significant financial impacts and/or erode clients' trust.

A cost estimate is generally established by a coordinating role of a tender manager supported by a technical expert (e.g. engineers and project managers). Tender managers and technical experts who perform cost estimates are referred to as estimators. Currently, existing estimation methods require detailed information about the project and tend to be very time-consuming and therefore costly. In the tendering phase, estimators have limited information. As a result, they leverage their knowledge/experience and make intuitive judgment calls

CONTACT Faridaddin Vahdatikhaki f.vahdatikhaki@utwente.nl

in order to estimate project costs (Cheng et al. 2010). Estimators have different levels of experience, which leads to tangible differences in the accuracy of cost estimates. Nonetheless, estimation methods in the tender phase of a project need to be quick, realistic and reasonably accurate (Kim et al. 2012). This is very difficult in the face of insufficient information and different levels of experience of estimators.

For a project to realize, first the client's needs should be translated into technical specifications before it can be built by a contractor. The preparation of projects for the actual construction involves many intricacies and complexities that fall well beyond the ability of average clients. This is where consultancy firms come to assist clients. They help to prepare details of projects by leading the feasibility studies, advising on technical details, and guiding the design. Consultancy firms consider a wide range of parameters, such as the type and scope of projects, to determine how much time and effort are going to be put in order to prepare the project, i.e. consultancy services. This is essentially different from the cost estimation applied during the construction, which is mainly concerned with the labour, material, equipment, etc. (Zwaving 2014).

Various estimation methods and techniques are available (Hamaker 1995; Elkjaer 2000; Burke 2009; Chou et al. 2009; GAO 2009; Zwaving 2014; NASA Executive Cost Analysis Steering Group 2015; Lester 2017). Nevertheless, given the fact that the tenders of consultancy firms need to be prepared accurately over a short period of time, the traditional cost estimation methods are not sufficient. This is mainly because the existing methods fail to fully utilize the tacit organizational knowledge which is embedded in past projects. Consequently, while there is usually a rich record of estimates and actual costs for previous projects, this valuable information is often not fully leveraged in preparing new tenders. As a consequence, estimation methods tend to be slow and inaccurate with high variability. This leads to significant financial impacts on the preparation of proposals for engineering projects.

With the advancements in computing power, recent cost estimating approaches tend to use more complex methods and a larger volume of data. Artificial Intelligence (AI) methods, which allow investigating multi- and non-linear relationships between final costs and design variables, have been deployed in recent years (Günaydin and Doğ an 2004). It is shown that through AI applications it is possible to obtain fairly accurate cost estimates even with limited information (Günaydin and Doğ an 2004). Examples of AL methods include, but not limited to, machine-learning (ML), knowledge-based systems (KBS), evolutionary systems (ES), and hybrid systems (HS) (Elfaki et al. 2014). These methods use large volumes of past tender data and identify patterns or relationships within these datasets. This often reduces the sensitivity of the estimates to the experience level or the subjective view of the estimators.

Nevertheless, while there is a myriad of data-driven and ML-based cost estimation methods for contractors, there are very limited studies on the development and application of similar methods for engineering consultancy firms. In other words, although much is known about the parameters that must be considered for accurate cost estimation for contractors, there is very little insight into factors that influence the cost estimation for consultancy firms. The study performed by Hyari et al. (2016) is, to the best of authors' knowledge, one of the few studies on the use of data-driven models for the cot estimation engineering services within the construction industry. In this study, the influential factors are determined by interviewing experts and showing them the available data. As a result, only a very limited number of variables (i.e. 5 variables) are identified and used in the model. Accordingly, the previous research does not provide much insight into the actual influence of a much wider spectrum of variables that impact cost estimation of engineering services. Similarly, not much is known about the relative importance of different variables for accurate cost estimation.

This research aims to investigate the potentials of developing an accurate ML-based method that can utilize past data to estimate the cost of engineering services provided by engineering consultancy firms based on the limited information available at the tendering phase. To this end, a thorough analysis of a wide range of input variables (i.e. influential factors on the cost of engineering services) is conducted to create an insight about what and how many of available variables at the tendering phase need to be used to build an accurate ML-based cost estimator. The focus is to build an accurate, yet simple, cost estimators that can work with minimal but significant/impactful input variables.

The remainder of this paper is structured as follows. First, the review of relevant literature is presented. This is, then, followed by the discussion of the proposed method. Next, the results of the case study are presented. Finally, conclusions and future work are elaborated.

Literature review

Automated cost estimation methods

The purpose of automated cost estimation is to identify the correlations between the influential factors and the project cost using predictive models or algorithms. Elfaki et al. (2014) distinguished four different state-of-the-art AI-based approaches, namely machine-learning (ML), knowledge-based systems (KBS), evolutionary systems (ES) and hybrid systems (HS). Kim et al. (2004) analyzed three cost estimating models, namely artificial neural networks (ANNs), multiple regression analysis (MRA) and a case-based reasoning system (CBR), and concluded that ANNs work more accurately then MRA and CBR estimating models. Furthermore, according to Cheng et al. (2010) ANNs represent the most frequently applied approach in estimating the duration and costs of construction projects during the preliminary design stage. ANNs have the ability to self-learn which saves a lot of development time. Also, ANNs can identify non-linear relationships between cost factors and project cost with no additional effort. With an ANN model, it is possible to obtain a fairly accurate prediction, even when sufficient information is not available in early stages of the design process (Günaydin and Doğan 2004).

Artificial neural networks (ANNs)

ANNs are originally inspired by the study of processes in the human brain (Günaydin and Doğ an 2004). ANNs consist of nodes (neurons in ANNs) grouped in interconnecting layers and sets of layers to form a network (Petroutsatou et al. 2012). There are three different types of layers, namely, input, hidden and output layers. The layout or architecture of a network is presented in Figure 1.

Conventionally, neural networks had a very simple structure with only input and output layers, these were called single-layer neural network or shallow neural networks. Neural networks with multiple hidden layers are called multi-layer neural networks or deep neural networks. Most of the contemporary neural networks used in practical applications are deep neural networks (Kim 2017). Every input node has a connection with all the nodes from the next hidden layer. This connection is illustrated by the arrow in Figure 1 and is corresponding to a particular weight. The training of the network consists of two different steps, namely, feedforward propagation and backpropagation. The training of a network begins with feedforward propagation, wherein the inputs and correct outputs from the training data are inputted to the neural network. Eventually, the neural network provides outputs based on the inputs and a random configuration of the weights. Subsequently, the outputs from the neural network are compared to the actual outputs and the error is calculated. The backpropagation is the process in which the weights are updated according to the error contribution in each node, and adjust the weights accordingly to reduce the error. These two steps are repeated for all the available training data. In this way the neural network grows in the accuracy by learning from examples.

Factors affecting the performance of ANNs

The quality and amount of training data are often the single most dominant factor that determines the performance of a model. The amount of data that is needed for a machine learning algorithm depends on the complexity of the problem and on the complexity of the chosen algorithm. A significant amount of practitioners have worked on different ML problems before. Therefore, reasoning by analogy is a way to determine the amount of data that is probably needed. In Table 1, 7 different studies that have comparatively similar scopes to this study are presented. All performances are given in the Mean Absolute Percentage Error

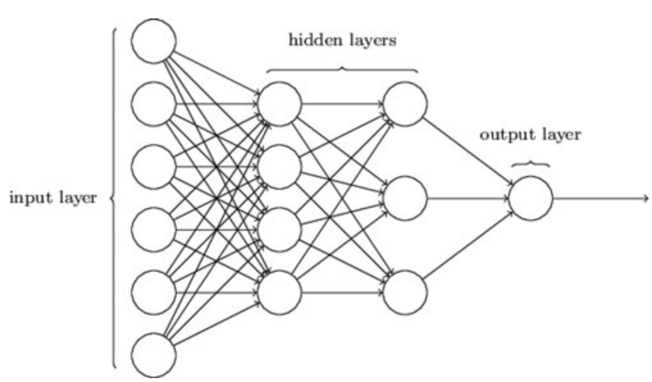

Figure 1. Structure of deep neural network or multilayer perceptron.

Table 1. Other academic work and performance.

Sources:	Performance, %	Data points
Cheng et al. (2010)	10.4	28
Günaydin and Doğan (2004)	7	30
Hyari et al. (2016)	28.2	224
Emsley et al. (2002)	16.6	288
Arafa and Alqedra (2011)	4	71
Mahamid (2013)	17	52
Arage and Dharwadkar (2017)	6.2	813

(MAPE). Hyari et al. (2016) used 224 datasets and achieved a performance of 28.2% which is lower than the 10.4% achieved by Cheng et al. (2010) with only 28 data points. Given the fact that they try to solve relatively similar cost estimation problems, this indicates that the amount of data does not have a direct relationship with the performance. The amount of data that is needed for an accurate model is difficult to determine prior to training the model. The performance analysis of the network can help decide if more data are required (Hagan et al. 2014).

In the supervised learning process of an ANN, the learning process is based on datasets that provide both input and output values. While the output value is normally determined by the purpose of the model (e.g. cost or time), the selection of inputs (or features) to be considered in the model is at the discretion of the modeller. It is possible that some features in the training data have irrelevant elements. For example, when the dimension of the potential input vector is very large, it can be beneficial to eliminate redundant or irrelevant features. This can reduce the required computation and can assist in preventing overfitting during training (Hagan et al. 2014).

Akintoye (2000) conducted a feature analysis and principal component analysis based on 24 different factors that influence project cost estimation in the construction industry. Example features are site constrains, availability and supplies of labour and materials, likely production time, and off/on-site operations sequencing. While the cost factors that are proposed by Akintoye (2000) are not all relevant for the cost estimation of engineering services, other studies provide information about the relevant factors for engineering services. Zwaving (2014) adopted a probabilistic estimating approach for cost estimation of engineering services within the energy and chemical industry. They proposed a set of features that are relevant for cost estimation of engineering services. Features that are distinguished in this research are for example the quality of information, scale of work, amount of work-sharing, and project team experience. Furthermore, Hyari et al. (2016) developed a conceptual cost estimation model for engineering services in a public construction project. They used 5 features, namely, project type, engineering services category, project location, total construction costs, and project scope.

Context of cost estimation for engineering services

Before an engineering project is started, a Request for Quotation (RFQ) is received. This request basically means that the client inquires a request for the cost of engineering services. This request is then appointed to a specific tender manager who is responsible to prepare a proposal for the client. It is crucial to understand the client's business case. Therefore, a kick-off meeting with an appointed team is organized. In this kick-off session, the scope and planning of the project are discussed. Subsequently, the proposal and calculations are prepared in detail. Each team member is asked to make a scoping document (WBS) with the required activities and deliverables that are needed to be carried out or delivered. These are usually made by experienced engineers in consultation with their specific department. Furthermore, the corresponding man-hour estimates of the expected activities and deliverables are requested. The estimated man-hours are then collected by the responsible tender manager and a final price is calculated. The estimation of the final price is basically calculated by multiplying the estimated required man-hours by the corresponding wage rates.

In order to verify and determine cost influencing factors that are used in the estimation process, interviews were held with 13 employees that have experience with preparing bidding offers for engineering services. The interviewees consisted of three project managers, five lead engineers (different departments), two heads of departments, and three tender managers. To start the interview, 14 influential factors were identified from the literature, as shown in Table 2. The interviewees were asked to rank the factors from the most important (rank 1) to the least important (rank 14). The average of the scores was taken to identify the average relative importance of the 14 different variables by expert opinion, as shown in Table 2. Next to the question about the ranking of influential factors, the interviewees were asked to specify other important factors that might have been left out from the original list. Based on the frequency of suggestions about the additional important factors, two other factors are added, namely *contract type*, and *intensity*. However, given the fact that the

Table 2. Final input variables.

Influencing factor	Description	Unit	Rank (by experts)
Scale of work	The costs of the total construction	Category Value in €	1
Project phases	The level of detail of the design	1 = Masterplan 2 = Conceptual design 3 = Basic design 4 = Detailed design 5 = Basic + detailed	2
Project duration	Number of weeks the project will take	Positive real number	3
Scope of work	The activities that are included in the contract	1 = Engineering (E) 2 = Engineering, Procurement, Construction (EPC) 3 = Engineering, Procurement, Construction Management (EPCm)	4
Type of work	The extent in which the project is a brownfield (modification) or greenfield (new construction) project	1 = 100% GF - 0% BF 2 = 75% GF - 25% BF 3 = 50% GF - 50% BF 4 = 25% GF - 75% BF 5 = 0% GF - 100% BF	5
Level of experience on clients side	The level of experience on the client side	1 = Very low level of experience 2 = Low level of experience 3 = Moderate level of experience 4 = High level of experience 5 = Very high level of experience	6
Scope definition	The extent in which the scope is defined	1 = Very poor scope definition 2 = Poor scope definition 3 = Moderate scope definition 4 = Good scope definition 5 = Very good scope definition	7
Size project team	Number of team members	Positive real number	8
Multidisciplinarity	Number of disciplines involved	Positive real number	9
Type of client and requirements	How demanding the client is towards standards and documentation	1 = Very low demands 2 = Low demands 3 = Standard demands 4 = High demands 5 = Very high demands	10
Main market type	The main market in which the project takes place	1 = Oil & Gas 2 = Chemicals 3 = Energy & Environment 4 = Health and Nutrition 5 = Infrastructure 6 = Industrial 7 = Property 8 = Public sector 9 = Pharma	11
Attitude towards design changes	The attitude of the client toward design changes	1 = Very high level of cooperation 2 = High level of cooperation 3 = Average level of cooperation 4 = Low level of cooperation 5 = Very low level of cooperation	12
Project manager experience	The amount of hours of experience the selected project manager has	1 = Project manager D (<2000 hours) 2 = Project manager C (<10.000 hours) 3 = Project manager B (<25.000 hours) 4 = Project manager A (<100.000 hours) 5 = Project director (>100.000 hours)	13
Pre-contract design	The extent to which the pre-contract design is complete	1 = To a small extent 2 = To some extent 3 = To a moderate extent 4 = To a great extent 5 = To a very great extent	14
Contract type	The type of contract in which the project is carried out	1 = Fixed Price 2 = Reimbursable	N/A
Intensity	The average hours a team member spend per week	1 = 8 hours/team member/week 2 = 16 hours/team member/week 3 = 32 hours/team member/week 4 = >32 hours/team member/week	N/A

interviews were held in separate sessions and also given that the decision about the addition of the two new factors was made only after all interviews were concluded, it was not possible to ask interviewees to rank the additional two factors with respect to the other 14. That is why the last two factors in Table 2 are not ranked. Nonetheless, for the remainder of the analysis and for the model development these two factors were taken into account, making the total number of influential factors considered in the model 16.

Model development

In this section, the model and the development methodology are described. The methodology is adapted from Hagan et al. (2014) and consists of data collection, network training, and validation as shown in Figure 2. The data collection concerns the development of the dataset. In this phase, the input variables of the model are determined. Furthermore, the data that is needed to train the model is collected. Subsequently, the network training covers the development of the actual model. The training phase consists of creating an ANN model and improving its performance by carrying out a heuristic optimization strategy, which is developed for this study and will be explained in detail in section 'Training'. Briefly, this methodology consists of three iterative phases. The first iterative phase is about determining the best training algorithm and best network architecture using the complete dataset. The second iterative phase determines whether the model can have better performance by using fewer input variables. The third iterative phase in the optimization strategy to identify the most relevant scope of the input variables, i.e. different proposal value ranges. Lastly, the validation phase is about the internal validation of the model.

Data collection

This step is concerned with the selection, gathering and pre-processing of data. Given that ANNs are not efficient at handling extrapolation, the training data must be as comprehensive as possible to cover the entire range of the model application (Hagan et al. 2014). As shown in Table 3, 132 projects, ranging in value, were selected for this study. The full dataset was divided into 11 categories based on the value. Due to the confidentiality agreement with the company, the value ranges must remain undisclosed. The number of projects within each cost range is presented in Table 3. Because the accuracy of the model depends on the size of the input data, three different scenarios were considered to identify ranges at which an accurate estimator can be trained, given the uneven distribution of project values. These scenarios are shown in Table 3. Given the insufficiency of data and also on the advice of company experts who pointed out the rarity of projects within certain ranges, the last three ranges were excluded from the analysis. In Scenario 1, all but the last three ranges are considered. In Scenarios 2 and 3, based on the expert opinion, four consecutive ranges, which accommodate most of the more recent projects, were chosen. The application of these scenarios will be explained in section 'Phase 3: Determining the model scope/range'.

As for the input variables, the 16 input variables that were identified in the interviews/literature (see Table 2) are used as the basis for the model development. Input variables were of both qualitative and quantitative natures. Quantitative values were represented using positive real numbers and can be directly used in the model. The qualitative data were quantified using a mapping scheme. For instance, the type of client and his/her requirements can be categorized into very high demands, high demands, standard demands, low demands, and very low demands. This can be mapped to a quantitative scale of 1-5, as shown in Table 2. The output variable for the dataset is the final proposal cost estimated by the experts.

When the required input and output criteria were determined, the data were gathered from various sources. Most of the required data were available in the database of the company. The data that were not available in the database were gathered using an online survey. In this survey, the responsible tender managers or project managers were asked to provide the missing data. Upon the completion of data gathering, a database with only the relevant data was built. Then, the data was cleaned to ensure homogeneity in the data. During this step, incomplete data sets were eliminated.

Training

Once the dataset became ready, a methodology was applied to build, train, and fine-tune the cost estimation model. This methodology is shown in Figure 3. This methodology consisted of three iterative sequential phases to determine (1) the best training function, (2) the most relevant input variables, and (3) the most fitting scope for the estimator.

Phase 1: determination of training function

In this phase, the best training function that can model the data was determined. In this research, the three most common training functions were considered, namely, the Levenberg-Marquardt (LM), Bayesian

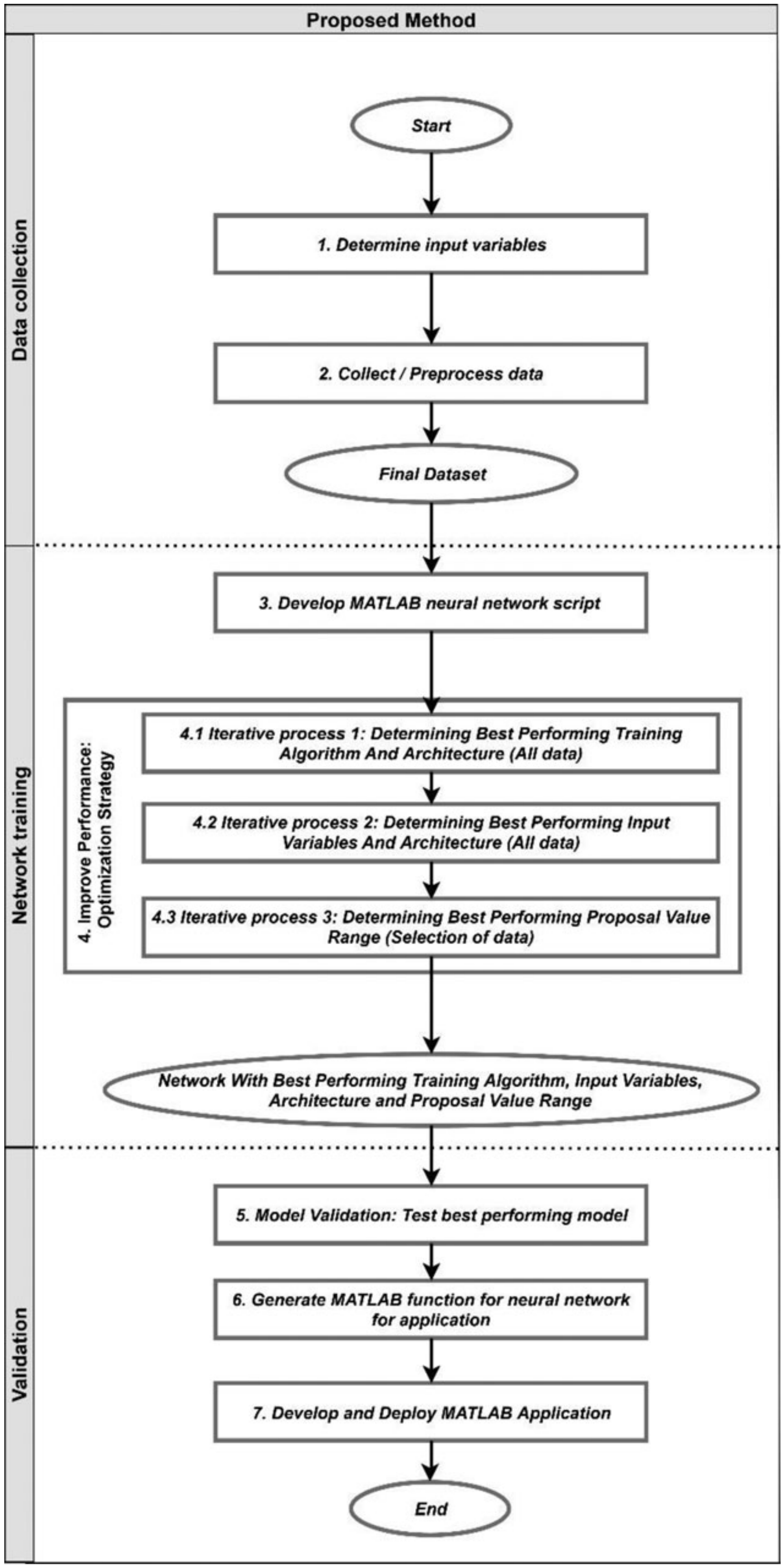

Figure 2. Proposed method.

Regularization (BR), and Resilient Backpropagation (RB) functions. The Correlation Coefficient (R) and Mean Absolute Percentage Error (MAPE) were used for the assessment of the models' performances. The model was considered a good fit when (1) R was close to 1, which indicates the goodness of the fit and (2) the difference between R values of the training and test sets was small, which indicates the model is stable and generalizable. In addition, the MAPE value was used to estimate the error of the test predictions. When the model was determined as a good fit, the model with the least MAPE value was selected.

First, a training function was selected and the network architecture was optimized using the growing method. In this method, the training is initiated by a single hidden neuron and then neurons are added one at a time until a threshold is reached on the over-fitting of the model.

Given the fact that the training of a multilayer neural network involves two stochastic elements, every training run may result in slightly different models. These two elements are (1) the initial weights and biases and (2) the random selection of training, testing and validation sets (Hagan et al. 2014). To get a robust estimate of the stochastic model, this variance must be taken into account. This was done by applying the Bootstrapping method. In this method, the distribution of an estimator or test statistic is captured by resampling the data several times (Allende et al. 2004). In this research, each configuration of the model was trained 100 times and the performance was calculated. For each configuration, the best model was identified and selected.

When the network architecture was optimized and the best model was identified, the next training algorithm was selected until all the training algorithms were tested. At the end of this phase, the best training algorithm and the best network architecture that explains the total dataset were found.

Table 3. Data selection: project value range.

Project value range (V in K€)	Number of projects	Inclusion in different scenarios		
		Scenario 1	Scenario 2	Scenario 3
Range category 1	2	✓		
Range category 2	12	✓		
Range category 3	23	✓		
Range category 4	21	✓		✓
Range category 5	19	✓	✓	✓
Range category 6	15	✓	✓	✓
Range category 7	15	✓	✓	✓
Range category 8	11	✓	✓	
Range category 9	8			
Range category 10	5			
Range category 11	1			

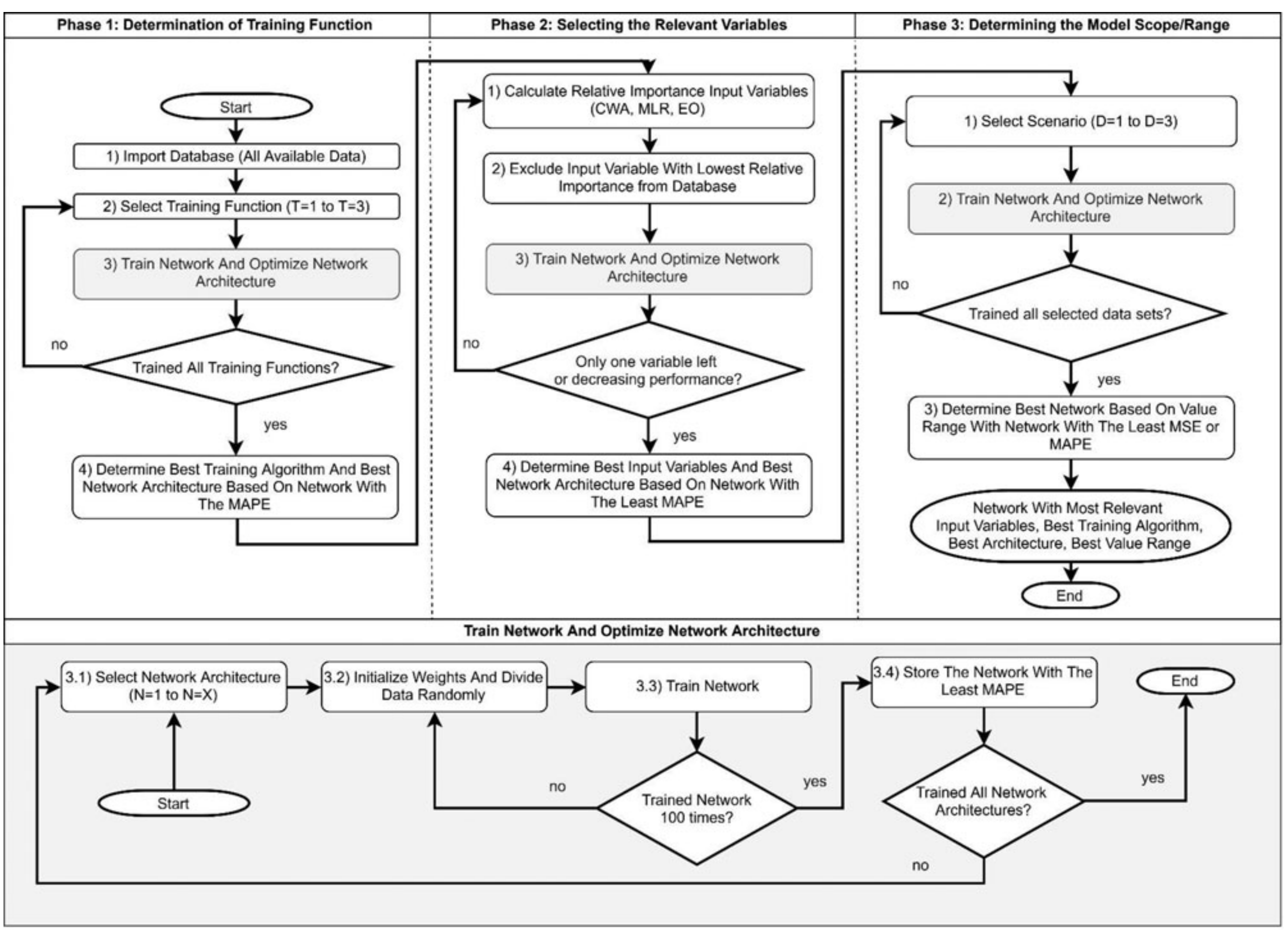

Figure 3. Optimization strategy.

The best results for the best network architectures for the three different training algorithms are shown in Table 4. It can be discerned that with 16 input variables and the complete dataset, the Bayesian Regularization (BR) training algorithm with one hidden layer and 4 hidden neurons performed the best. For this architecture, R-value is 0.98 and the difference between training and testing dataset is 0.03. In addition, this architecture has the lowest MAPE score. Which means that it has the lowest mean average percentage error for both the training and the testing set.

Phase 2: feature selection

In order to find the simplest model that explains the data, it could be helpful to eliminate redundant or irrelevant input variables. This process is also known as feature selection. By calculating the relative importance of input variables of the network with the highest performance, redundant input variables can be removed and thus make the model more generic. A method called Connection Weights Algorithm (CWA) (Olden and Jackson 2002) was used to calculate the relative importance of input variables of the model. This approach is based on the estimation of the network's final weights obtained by training the network (Ibrahim 2013; Janssen 2018). The relative importance of the input variables is shown in Table 5. Once the relative importance of the input variables was identified, the next step was to eliminate the variables that have low impact from the best model identified in the previous phase. The elimination started by removing the variable with the lowest relative importance. The elimination of the variables continued until the change in the performance of the model was decreasing with regard to the previous configuration. Due to the fact that the number of input neurons decreased, the number of neurons in the hidden layers also needed readjustment. Therefore, the training entered the network architecture optimization module again. Eventually, the simplest model that explains the data was determined.

In this research, to ensure that the most accurate model is achieved, in addition to CWA, two other methods were used to filter the input variables, namely, (1) multiple linear regression analysis and (2) expert opinion (shown in Table 5). MLR analysis is a suitable method to identify which variables have a significant influence on the proposal price (van der Steen 2018). It can help determine whether there is a linear association between the independent variables and proposal price. The relative importance of the

Table 4. Results of the first phase.

Network architecture	R train	R test	R all	MAPE train (%)	MAPE test (%)	MAPE all (%)
LM-16-6-1	0.9997	0.9168	0.9684	57.05	100.18	77.98
BR-16-4-1	0.9998	0.9645	0.9796	37.25	50.36	39.24
RP-16-6-1	0.9966	0.7509	0.8666	59.48	88.68	89.51

The network architecture is described:
'Training algorithm' – 'No. of input Variables' – 'No. of hidden neurons' – 'No. of output Variables'.
LM: Levenberg-Marquardt.
BR: Bayesian regularization.
RP: Resilient backpropagation.

Table 5. Comparison of variable importance using different methods.

Input variables	Variable relative importance based on CWA (%)	Variable significance based on MLR	Variable ranking by experts
Scale of work	7	12	1
Project phases	6	11	2
Project duration	3	1	3
Scope of work	14	10	4
Type of work	12	16	5
Level of experience on clients side	10	9	6 (was complexity of design when ranked by experts)
Scope definition	16	15	7
Number of project team members	2	4	8
Collaborating disciplines	4	3	9
Type of client and requirements	13	14	10
Main market type	8	7	11
Client's attitude towards design changes	11	13	12
Project manager experience	15	8	13
Pre-contract design	9	6	14
Contract type	5	5	(added later)
Intensity	1	2	(added later)

Table 6. Results of the second phase.

Variable selection method	Network architecture	R train	R test	R all	MAPE train (%)	MAPE test (%)	MAPE all (%)
Connection weights algorithm	BR-9-5-1	0.9992	0.9640	0.9813	48.27	51.32	48.73
	BR-8-6-1	0.9995	0.9556	0.9979	55.07	42.26	53.13
	BR-7-6-1	0.9994	0.9648	0.9952	46.73	37.41	45.32
	BR-6-8-1	0.9997	0.9460	0.9985	35.19	32.83	34.83
	BR-5-7-1	0.9996	0.9952	0.9991	33.15	27.41	32.28
Multiple linear regression	BR-7-6-1	0.9998	0.8419	0.9921	46.68	52.70	47.87
	BR-6-6-1	0.9996	0.9784	0.9939	38.83	42.56	39.56
	BR-5-7-1	0.9999	0.9065	0.9806	23.56	42.47	27.28
Expert opinion	BR-7-2-1	0.9583	0.9449	0.9519	167.69	121.04	158.50
	BR-6-7-1	0.9824	0.9405	0.9668	210.61	105.18	189.84
	BR-5-6-1	0.7530	0.8609	0.7664	274.98	93.25	239.19

Table 7. Best results third iterative process.

Scenario	Network architecture	R train	R test	R all	MAPE train (%)	MAPE test (%)	MAPE all (%)
1	BR-5-7-1	0.9989	0.9705	0.9819	24.04	23.612	23.971
2	BR-7-4-1	0.9957	0.9944	0.9954	13.648	13.648	13.648
3	BR-5-6-1	0.9915	0.9581	0.9832	16.246	21.870	17.130

independent variables is determined by the unit drop in R^2 when a variable is deleted from the sample. R^2 is the coefficient of determination and shows the percentage of variation in a dependent variable which is explained by all the independent variables together. The larger the drop in R^2 when a variable is removed from the sample, the more important it is assumed to be. Also, expert estimators can be interviewed to rank the importance of the input variables. These two methods were applied and the results were compared to the model formed by CWA, as shown in Table 5.

Using CWA, it was observed that the performance decreased significantly after the elimination of the 12th variable (Contract type). As shown in Table 6, the best performances occurred when the model was trained using between 5 to 9 top variables from Table 5. Within this range, the differences between performances were rather minor. The phase has proven effective in improving the accuracy of the model. To put this into perspective, the lowest MAPE with all 16 variables was 50.36%, i.e. the best performance in Phase 1. Nevertheless, when only the top 5 variables were used, the MAPE plummet to 27.41%. This indicates a 45.6% improvement in accuracy. It could be concluded that the most dominant variables for the estimation of engineering services based on CWA method are intensity, number of project team members, project duration, collaborating disciplines, contract type, project phases, scale of work, main market type, and pre-contract design.

When the MLR method was used, the best performances were achieved when the top 5 to 7 variables are used. As shown in Table 6, the best results based on MLR ranking was achieved when 5 top variables (project duration, intensity, collaborating disciplines, number of project team members and contract type) are considered. In this case, the MAPE of the model is 42.47%. While this indicates a slight improvement of accuracy compared to the best model from Phase 1, it is evident that CWA performs better in capturing the relative importance of variables for this particular case. This can be justified by the fact that MLR assumes linear relationships between independent variables and a dependent variable. In reality, some of the variables could have a non-linear relationship with the dependent variable, which cannot be captured by MLR. CWA-based variable selection, on the other hand, is capable of handling non-linearity in the data.

Finally, as shown in Table 6, the best results of the training based on the ranking of variables by expert opinion was achieved when the top 5 variables (scale of work, project phases, project duration, the scope of work and type of work) are used for the training. In this model, MAPE was 93.25% when 5 hidden neurons are used. Based on the comparison of MAPE values of different methods, expert opinion performed poorly in terms of capturing the important variables. This can be partially attributed to the fact that the final ranking was based on the average scores by all the interviewees. It is conceivable that some of the interviewees were ranking the factors from their own standpoint and lacked a global view on how different factors impact the entire service provided by the firm.

Phase 3: determining the model scope/range

In the last phase of the training, the best scope/range for the input data was determined. As stated earlier

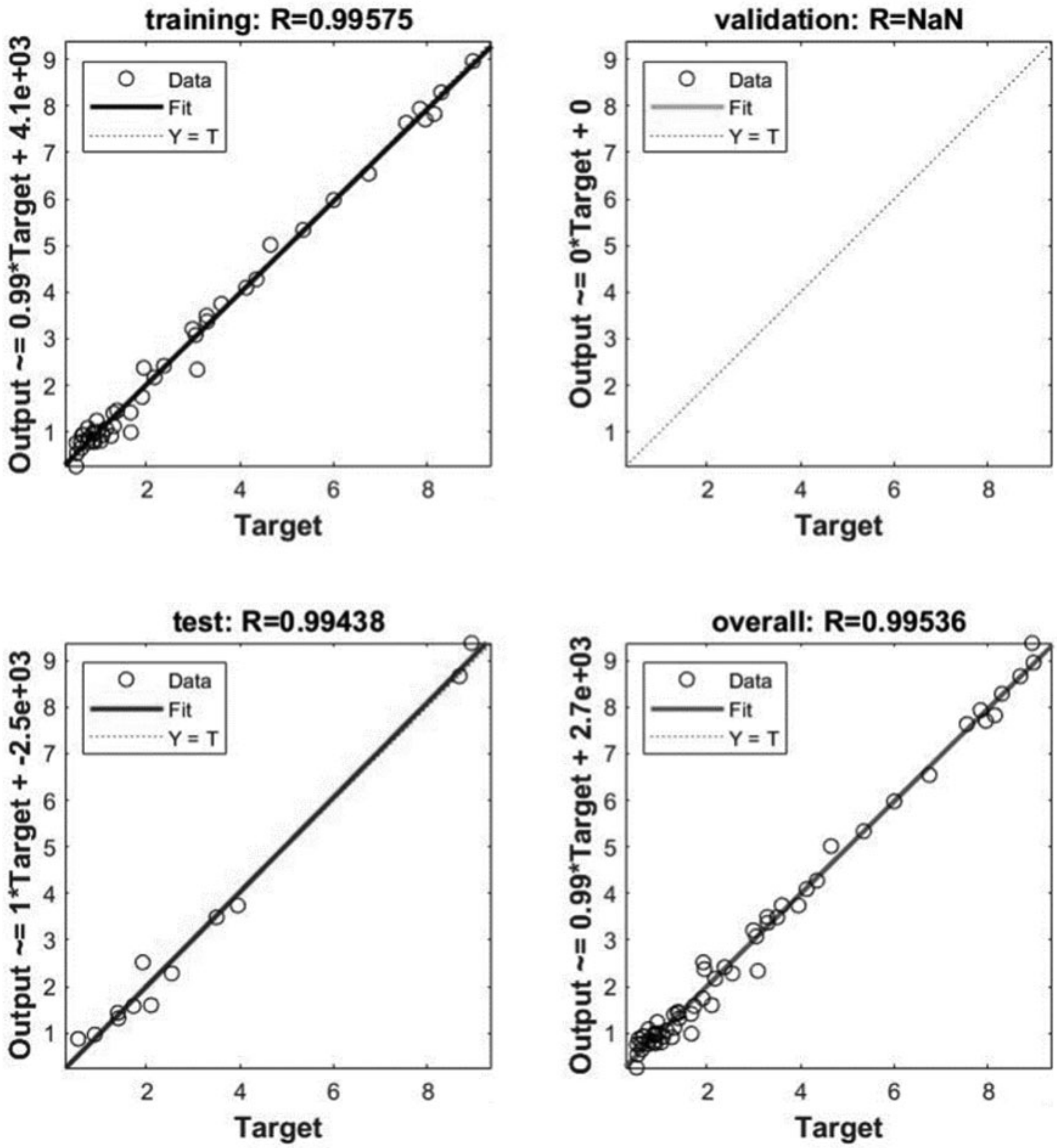

Figure 4. Regression plot BR-7-4-1.

in section 'Data collection', because ANN is not efficient in (1) handling extrapolation outside the range of the training set and (2) fitting an accurate model with insufficient data, it is important to determine at which scope/range of the input data the model performs the best. For this purpose, the 3 scenarios that are presented in Table 3 were used to investigate the impact of scope reduction/adjustment on the accuracy of the model. As mentioned in section 'Data collection', these scenarios were formed based on the amount of data that is available for the different project value categories. It should be noted that the elimination of certain input data from the dataset will improve the overall accuracy of the model at the expense of sacrificing the generic-ness of the model. First of all, it is decided to proceed the training with the 5 different network architectures (top 9 to 5 variables determined by CWA method). However, when a data selection was made, the complexity of the underlying function of the data could be different compared to the full database. Therefore, the growing technique was used again and the number of hidden neurons was changed for every network in each training set.

In this phase, a model is developed for each scenario. Given the fact that the change of the input dataset may affect the architecture of the model, a new model was trained and optimized for each scenario. Ultimately, the performance of the models was compared and analyzed.

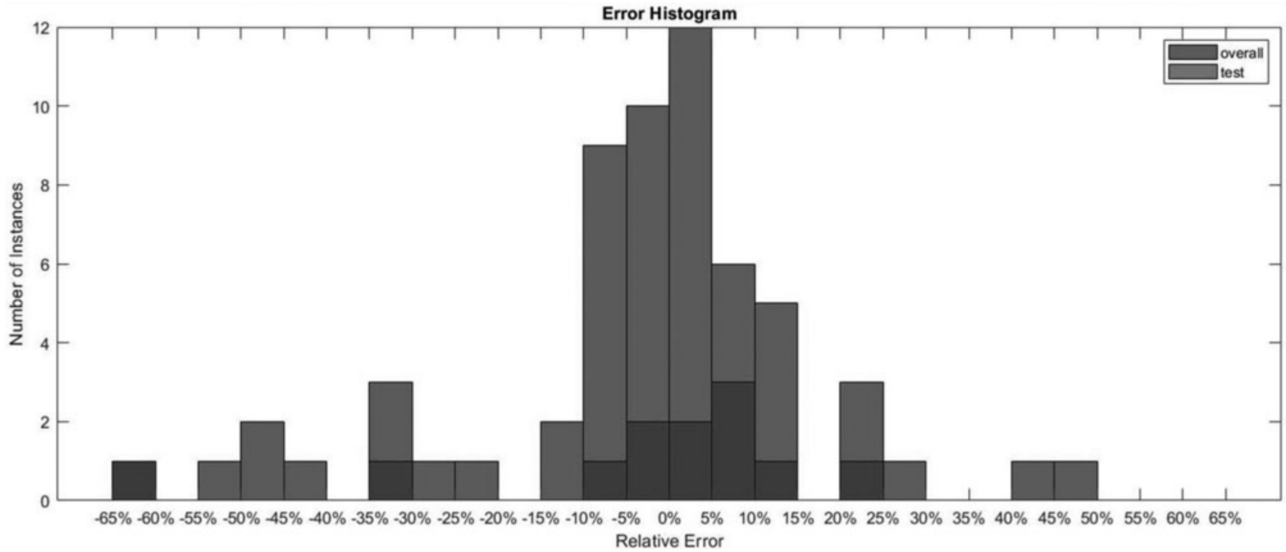

Figure 5. Error histogram, with bin sizes of 5%, for BR-7-4-1 with 60 data points.

Table 8. Test results best model.

Test case	1	2	3	4	5	6	7	8	9	10	11	12
Relative error in %	0.3	−62.1	−8.5	−31.2	8.2	5.8	−3.4%	5.4	0.3	10.2	23.7	−4.8

The results of this phase are shown in Table 7. While the performance of the model has improved for all scenarios, compared to the best models from phase 2, scenario 2 has shown the greatest improvement. In this scenario, projects with a value range category 2 until 5 are chosen. The best model for this input dataset had 7 input variables and 4 hidden neurons. In this case, MAPE of 13.65% was achieved, which indicates 50.2% and 72.9% improvement over the best model of Phases 1 and 2, respectively. Very important is that the R-value for both training and test sets are very similar and only differed by 0.0013, which indicates high generalizability. Furthermore, the R values for both sets were very close to 1, which indicates high goodness of the fit. Finally, the performances of the training MAPE, test MAPE, and overall MAPE were very similar.

Validation

As the next step in the model development, the best model architecture identified in the training step (i.e. BR-7-4-1) was validated by investigating the performance of the model outside the training sample.

In addition, the Bootstrapping method was again used to reduce the impact of randomly selected elements in the training. At the end of each iteration, the performance of the model was analyzed and the MAPE of the model was calculated. Subsequently, the mean MAPE and the standard deviation were calculated for all the models combined. By doing so, a more robust estimate of the variance of the model can be acquired.

The regression plot of the validation step is shown in Figure 4. The regression plot shows that both the training and testing results were very promising, with the respective R-value of 0.99575 and 0.99438. The distribution of the relative error of the individual estimates is shown in Figure 5. In addition, the relative errors in percentage for the individual test results are provided in Table 8. For this model, the MAPE of the total set was 13.65%. with a maximum error of 62% and a minimum error of 0.32%. For the test set, about 66% of the predictions had a relative error of lower than 10%. In addition, 37% of the predictions, both on training and test data, had a relative error of less than 5%.

Discussions and conclusions

The goal of this research was to investigate the possibility of developing an accurate ML-based cost estimation method for tendering of engineering services. This was done by systematically developing and optimizing a neural network model to estimate the preliminary costs of engineering services. The research applied a systematic methodology that provides a guideline for developing and optimizing an artificial neural network for cost estimation. The development of the neural network included measures to remove the nuisance from the data. The systematic methodology applied for the optimization of the network proved to be very efficient in improving the performance of the model.

The results showed that ANN can be used to obtain a fairly accurate cost estimate, even with minimum data that is available during the tendering phase. The accuracy that is obtained with the ANN model is well within the range achieved by models developed in comparatives studies in the literature, see Table 1. However, these results were achieved for cost estimation from the perspective of contractors and not for engineering services. When looking at the ML-based cost estimation model for engineering services, i.e. the model developed by Hyari et al. (2016), the model developed in this study showed a 14.5% improvement in the accuracy of model, considering MAPE.

One of the main contributions of this study is the analysis of the relevant and important variables for the cost estimation of engineering services. It is shown that by decreasing the number of variables and excluding the less important variables, the performance of the model can improve. Furthermore, it is shown that the cost of the engineering services can be accurately estimated using the following 7 input variables: (1) intensity, (2) number of project team members, (3) project duration, (4) collaborating disciplines, (5) contract type, (6) project phases, (7) scale of work. It is discovered that the variables that are found more prominent for the cost estimation based on CWA and MLR methods are different from those identified by the expert opinion.

However, there are a number of limitations with the present research. First, 132 individual data points were collected and the best neural network was developed using 60 data points. Due to the split-sample technique that is used, the test results were based only on 12 cases. This is considered a small sample and the model needs to be validated with a larger set of data. Second, the model also needs to be externally validated by applying it to new projects and comparing the results with the actual estimate by experts. Finally, more research needs to be done on the adoption of ML-based cost estimation in practice. Given the black-box nature of ANNs, building trust in the model within an organization seems challenging. Neural networks are accurate predictors, however it is a challenge to offer a justification for the structure and behaviour of the model.

Acknowledgements

The authors of this paper would like to thank Bilfinger Tebodin as their case acted as a context for the research that is conducted. They allowed their employees to participate in this research and gave us access to their data. Therefore, special thanks to all the participants from the company.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Ahiaga-Dagbui D, Smith SD. 2012. Neural networks for modelling the final target cost of water projects. In: Smith SD, editor. Proceedings 28th Annual ARCOM Conference, Sep 3–5; Edinburgh, UK. Association of Researchers in Construction Management. p. 307–316.

Akintoye A. 2000. Analysis of factors influencing project cost estimating practice. Constr Manage Econ. 18(1): 77–89.

Allende H, Ñanculef R, Salas R. 2004. Robust bootstrapping neural networks. In Mexican International Conference on Artificial Intelligence. Berlin, Heidelberg: Springer. p. 813–822.

Arafa M, Alqedra M. 2011. Early stage cost estimation of buildings construction projects using artificial neural networks. J Artif Intell. 4(1). http://hdl.handle.net/20.500.12358/24892.

Arage SS, Dharwadkar NV. 2017. Cost estimation of civil construction projects using machine learning paradigm. 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 594–599.

Burke R. 2009. Project management - planning and control techniques. 4th ed. India: Wiley. p. 251–266.

Cheng MY, Tsai HC, Sudjono E. 2010. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. Expert Syst Appl. 37(6):4224–4231.

Chou JS, Yang IT, Chong WK. 2009. Probabilistic simulation for developing likelihood distribution of engineering project cost. Autom Constr. 18(5):570–577.

Elfaki AO, Alatawi S, Abushandi E. 2014. Using intelligent techniques in construction project cost estimation: 10-year survey. Adv Civil Eng. 2014(December):1.

Elkjaer M. 2000. Stochastic budget simulation. Int J Project Manage. 18(2):139–147.

Emsley MW, Lowe DJ, Duff AROY, Hickson A. 2002. Data modelling and the application of a neural network approach to the prediction of total construction costs. Constr Manage Econ. 20(6):465–472. DOI: 10.1080/01446190210151050.

GAO. 2009. GAO cost estimating and assessment guide. Washington, D.C.: US Government Accountability Office, 440.

Günaydin HM, Doğ an SZ. 2004. A neural network approach for early cost estimation of structural systems of buildings. Int J Project Manage. 22(7):595–602.

Hagan MT, Demuth HB, Beale MH, Orlando DJ. 2014. Neural network design. 2nd ed. Natick, Massachusetts, United States: MathWork.

Hamaker J. 1995. No parametric estimating. In: Stewart RD, Wyskida RM, Johannes JD, editors. Cost estimator's reference manual. New York (NY): John Wiley & Sons. p. 233–265.

Hyari KH, Al-Daraiseh A, El-Mashaleh M. 2016. Conceptual cost estimation model for engineering services in public construction projects. J Manage Eng. 32(1):04015021.

Ibrahim OM. 2013. A comparison of methods for assessing the relative importance of input variables in artificial neural networks. J Appl Sci Res. 9(11):5692–5700.

Janssen J. 2018. Developing a model for estimating road capacity values of weaving sections. Master's Thesis, University of Twente, Enschede, the Netherlands. p. 1–90.

Kim GH, An SH, Kang KI. 2004. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. Build Environ. 39(10):1235–1242.

Kim HJ, Seo YC, Hyun CT. 2012. A hybrid conceptual cost estimating model for large building projects. Autom Constr. 25:72–81.

Kim P. 2017. MATLAB deep learning. Berkeley, CA: Apress. DOI: 10.1007/978-1-4842-2845-6.

Lester EIA. 2017. Estimating. In: Project management, planning and control. The Netherlands: Elsevier. p. 61–65. DOI: 10.1016/B978-0-08-102020-3.00013-9.

Mahamid I. 2013. Conceptual cost estimate of road construction projects in Saudi Arabia. Jordan J Civ Eng. 7(3):285–294. DOI: 10.12816/0024836.

Nanculef R, Salas R. 2004. Robust bootstrapping neural networks, (August 2018).

NASA Executive Cost Analysis Steering Group. 2015. NASA cost estimating handbook. NASA 63(4).

Olden JD, Jackson DA. 2002. Illuminating the "“black box”: a randomization approach for understanding variable contributions in artificial neural networks. Ecol Model. 154:135–150.

Petroutsatou K, Georgopoulos E, Lambropoulos S, Pantouvakis JP. 2012. Early cost estimating of road tunnel construction using neural networks. J Constr Eng Manage. 138(June):679–687.

van der Steen E. 2018. Assessing the influence of tender and project characteristics on project performance. Master's Thesis, University of Twente, Enschede, the Netherlands.

Zwaving J. 2014. Probabilistic estimating of engineering costs. Delft (The Netherlands): Delft University of Technology.