DOI: 10.38027/ICCAUA2022EN0210

Artificial Neural Network Analysis for Cost Estimation of Building Projects in India

* Ar. **Ankita Gupta** [1], Ar. **Piyali Debnath** [2]
Colliers International, Project Management Consultants, Mumbai, India. [1]
National Institute of Technology, Faculty of Architecture and Planning, Calicut, India [2]
E-mail [1]*: guptaankitain@gmail.com (Corresponding author), E-mail* [2]*: piyalidebnath163@gmail.com*

Abstract
In Construction Management, it is difficult to predict the cost estimate during the preliminary stage of the project because of limited information and unknown factors. Artificial Neural Networks can help in the prediction of estimate because of their simplicity and adaptability to non-linear problems. Due to their self-organizing nature they can be used to solve the problems even with low level programming. This makes them useful in interpreting and generalizing inadequate input information. ANN's are crude derivatives of the biological neural network with single layered or multi-layered neuron in the form of input layer, hidden layer and output layer. The neural network first has to undergo training from historical data in order to make predictions or show results. The size of the data set, number of hidden neurons and the neural network architecture determines the success of the results. Selecting the right data set becomes imperative in this case. For the purpose of cost estimation, the cost drivers were taken as inputs and their estimated costs were taken as the target value. The cost drivers were selected carefully through literature review and survey to provide more accurate results for the estimate. The main drivers identified were: type of building, location, seismic zone, project complexity, ground condition, soil condition, plot area, plinth area, built-up area, number of stories, number of basement, principal structural material, type of foundation, level of design complexity, modular design, market conditions, construction conditions, risk factor, impact of risk, estimated duration of work, specification, quality of work and detailed cost estimate of project. A Pareto analysis performed on the significant drivers showed that the Duration of Work, Complexity of the Building, Plinth Area and Built-up Area, Height and Specifications were the most important cost drivers in a construction project. A problem was formulated based on these drivers with numerical and categorical data. The data set was trained with a neural network using the MATLAB software using feed forward backpropagation. Training was carried out till the greatest correlation and least Mean Squared Error was obtained after multiple iterations. This trained data was used to predict the cost for a new project. The output of the testing was 87% accurate despite the small data set used.
Keywords: Artificial Neural Network, Cost estimation, Construction projects, Building projects.

1. Introduction
In Construction Management, it is a difficult challenge to produce an accurate project cost estimate that takes into account the risks and uncertainties which may result in costs overruns. Cost estimation generally involves predicting labour, material, utilities or other costs over time given a small subset of factual data on "cost drivers."
The different types of costs involved in building construction projects are as follows:

Category I. Direct charges, which are identifiable as being associated with a particular project—standard labour, project labour, material, installation, salaries, and miscellaneous costs;

Category II. Allocated costs, which are charged to the project as the result of applying various formulas to the direct charges—standard overhead, other manufacturing costs, construction overhead, and general sales and administrative expenses.

In project management, cost estimation is not just a simple arithmetic based process. Estimation is an experience based task which includes assumptions, inclusions and exclusions. Cost estimation plays a crucial role in the initial part of a project as it enables decision making for investors, helps in the quantitative assessment of project and selection of alternatives.

This cost estimate is used to set up a budget during the front end of the project. The same budget is used as a baseline to assess the performance of the project during the later stages and to avoid cost overruns. In the worst cases, cost overruns can cause the interruption of construction activities, and even the abandonment of projects, situations that occur when the owner does not have the financial capacity to meet the unexpected costs (Herrera, 2020).

Artificial Neural Network (ANN) can be used here in cost forecasting because it is an analogy based process that works well with limited data- in the same way as the human brain. Our brain has been developed after thousands of years of evolution, and holds several advantages over the von Neumann or modern parallel computers. These include massive tolerance, learning and generalization ability, adaptivity and low energy consumption (Jain, 1996). Neural networks use a process which is analogous to the brain, wherein the existing data undergoes "training". By using an adaptive learning rate for training, a trained neural network is developed for a better test accuracy. (Takase, 2018) The trained network becomes an "expert" in the category of information that it has been given to analyse. This "expert" can then be used to provide projections for new scenarios. Artificial neural networks are purely data driven models which transit from a random state to a final model through iterative training. Artificial neural networks are non-parametric statistical estimators which have proven to be universal approximators, and thus have potential for use in cost estimation modelling.

1.1 Aim

This research aims to examine the performance of cost estimation modelling through neural networks to predict the initial cost estimate for building construction projects in India.

1.2 Need

Conceptual cost estimates play a crucial role in initial project decisions, even when scope is not finalized and very limited design information is available during early project stages. Some categories of problems and challenges faced in the project environment may depend on so many subtle factors that a computer algorithm cannot be created to calculate the initial cost. There exists a need to predict reliable results from complicated or imprecise data for obtaining the initial cost estimate of a project. Artificial Neural Networks (ANN) are a family of statistical learning models inspired by the way biological nervous systems that process information. Unlike some conventional statistical systems, ANN adapt continuously to new data and possess an ability to learn from arbitrary and noisy data permits (Abraham, 2005). Hence they can lend themselves towards cost estimation in the early stage of the project. ANN method is appropriate for cost prediction due to the following reasons: (a) it is able to determine the interdependencies between data when considering significant variables in construction, (b) it can deal with non-linear relationships (c) it can handle incomplete data sets (Arafa, 2010, Amen,2022).

1.3 Objectives

The objectives of the paper are as follows:

- Analysis of the cost drivers based on historical data and identification of significant cost drivers.
- Formulating a problem based on the identified data set for cost estimation.
- Finding the cost estimate for a test project based on input from simulated and actual data sets using artificial neural network analysis tool for projects in residential, office, commercial and institutional category.
- Analysis of artificial neural network analysis tools in cost estimation in terms of accuracy, variability, model creation and model examination.

2. Methodology

Step 1- Identification and collection of data set for cost estimation. The data set is a key factor in cost estimation of the project. If the data set is not reliable, the results from the networks calculations will not be reliable. This dataset is collected through literature study and actual survey.

Step 2- Identifying the variables in the data set and classifying them into the following types.

Categorical data. Dependent Category: Dependent or output variable whose possible values are taken from a set of possible categories; for example, yes or no, or red, green, or blue.

- Dependent Numeric: Dependent or output variable whose possible values are numeric.
- Independent Category: Independent variable whose possible values are taken from a set of possible categories; for example, yes or no, or red, green, or blue.

Numeric data. Independent variable whose possible values are numeric.

Step 3- Setting the input variables for cost estimation of the project. These input variable will differ based on typology, complexity, location, materials, labor and other specifications of the project.

Step 4- Estimating the cost using neural network analysis technique for the sample set. Regression analysis is a statistical technique used to measure the extent to which a change in one quantity (variable) is accompanied by a change in some other quantity (variable).

Step 5- Comparing the results of ANN with the actual results based on accuracy, variability, model creation and model examination.

Step 6- Determining the relevance of the quantitative tools for the project environments.

3. Artificial Neural Network – An overview

- An Artificial Neural Network (ANN) provides a method to characterize synthetic neurons to solve complex problems in a manner which is similar to the human brain. (Huang, 2009).
- An artificial neural network is composed of many artificial neurons that are linked together according to a specific network architecture.
- The objective of the neural network is to transform the inputs into meaningful outputs.
- The ANNs have a high degree of self-organizing ability. In its crude imitation of the biological neural network, it enables low level programming to solve complex problems- especially those which are non-linear, non-analytical, non-stationary and stochastic.
- The types of neural networks are classified based on learning systems which include a) Supervised, b) Unsupervised and c) Deep learning. A feedforward neural network is a simple type of artificial neural network with supervised learning where the units do not form a cycle between the connections of the units. It is the first and simplest type of ANN. (Mishra, 2018)

4. Identifying the Cost Drivers for A Construction Project

In order to write a neural network program for cost estimation in a project (with reasonable accuracy), the factors which have an impact on the overall cost have been listed.

The cost of building materials can themselves account for 60% of the cost of residential buildings (Olotuah, 2002). However, it is important to understand that there are several factors which may affect the cost of construction. They can range from accountable factors like quantity of materials to unaccountable factors like bribes resulting from corruption. It is necessary to narrow down the factors to streamline the neural network and get a logical output.

The following steps are used to sort through cost drivers and decide the input parameters:

4.1 Identification of Significant Factors Based on Literature Study

Table-1 demonstrates the list of factors elicited through literature study. (Stoy, 2007), (Ferry, 1999), (Toh, 2012), (Chan. S, 2005), (Ji S.H, 2019). Based on their frequency of occurrence, they were classified as significant and non-significant factors.

Table 1: Identified factors from literature study.

Sl. No.	Parameters	Significance
1	Historical Database	Non-significant
2	Accuracy and experience of estimator	Significant
3	Knowledge of Client	Significant
4	Weather	Non-significant
5	Design Team Resources	Non-significant
6	Type of Estimation Method Used	Significant
7	Site Investigation/ Geotechnical Data	Significant
8	Structural Materials used	Significant
9	Cost of Construction Materials	Significant
10	Availability of Labour	Significant
11	Labour Wage Rate	Non-significant
12	Contractor’s / Sub Contractor’s Experience	Non-significant
13	Project Manager’s Experience	Non-significant
14	Size of project	Significant
15	Location of Project	Significant
16	Typology of Project	Significant
17	Regulatory Requirements	Non-significant
18	Political Influence	Non-significant
19	Environmental Conditions	Non-significant

20	Social and Religious Impact	Non-significant
21	Construction Site Conditions	Non-significant
22	Availability of services	Non-significant
23	Insurance	Non-significant
24	Number of Stories/ Height	Significant
25	Gross Floor Area, Plot Area, Built-up Area	Significant
26	Shape of the Building	Non-significant
27	Probability of Force Majeure (Earthquake, Flood or Cyclone, Fire)	Significant
28	Ground Conditions	Significant
29	Design Complexity	Significant
30	Market Conditions	Significant
31	Type of Foundation	Significant
32	Availability of Labour and Equipment Required	Significant
33	Labour Productivity	Non-significant
34	Quality of Work Required	Significant
35	Specification of Finishes	Significant
36	Total Duration of Project	Significant
37	Project Program	Non-significant
38	Changes in design/ Scope Change	Non-significant
39	Changes in Material	Non-significant
40	Change in Schedule	Non-significant
41	Risk Factor	Significant
42	Impact of Risk	Significant
43	Volume of HVAC work	Non-significant
44	Overheads	Significant
45	Mark-up and Profit	Non-significant
46	Amount of Rework	Non-significant
47	Inflation Rates	Non-significant
48	Delay in construction	Significant

4.2 Finding the Rank of The Cost Drivers Based on Relative Importance Index from Survey

A survey was carried out amongst construction industry professionals, project managers and architects to identify the critical cost drivers amongst the significant factors which were identified. Table-2 and Table-3 exhibits the weightages and ranking of the previously identified cost drivers consecutively, which are derived by analysing the surveyed data by using RII (Relative importance index) by using Eq. 1.

$$\mathrm{RII} = \frac{\sum W}{A * N} \qquad (1)$$

Where, W—weighting given to each statement by the respondents and ranges from 1 to 5; A—Higher response integer (5); and N—total number of respondents.

Ranking is done based on relative importance index.

Table 2: Identified factors from literature study.

Type of building	Location of Building	Sector	Complexity	Ground Conditions	Soil Condition	Plot Area, Plinth Area, Built-up Area	Number of Stories (or Height)	Number of Basements	Principal Structural Members (Masonry, R.C.C or Steel)	Type of Foundation (Isolated, Raft or Pile)	Design Complexity (Modular or Non-Modular Design)	Market Conditions	Seismic Zone	Risk factor (High, Medium, Low)	Construction Conditions (Climate, Site access, storage)	Duration of Work	Specifications (for finishes)
4	5	3	5	3	5	4	5	5	5	5	4	4	4	4	3	5	2
5	3	3	3	2	2	5	5	5	5	4	4	3	3	3	3	5	5
4	5	2	4	5	5	4	3	2	4	4	2	3	4	3	3	4	3
3	3	2	3	3	4	4	4	2	2	2	3	3	3	3	2	2	2
4	4	2	5	5	2	3	4	1	3	4	5	5	2	5	5	5	5
5	4	3	4	4	4	4	3	2	4	4	2	5	3	2	4	5	5
5	5	4	5	4	5	5	5	5	5	5	5	4	4	5	3	5	3
1	5	5	3	1	1	2	2	1	1	1	1	3	1	4	1	4	1
5	5	3	3	3	3	3	4	2	4	2	2	5	2	3	2	4	5
5	3	4	4	5	5	4	3	4	3	5	4	5	4	5	5	5	5
4	5	2	4	5	5	2	3	2	3	3	3	4	3	3	2	2	2
2	4	2	5	5	4	3	3	5	4	4	4	3	2	3	2	4	3
5	5	3	5	5	5	5	5	5	5	5	5	5	3	4	3	5	5
3	5	4	4	5	4	5	5	4	5	5	4	4	5	4	3	4	3
4	4	3	4	3	3	5	5	5	4	3	4	4	4	3	3	4	4
4	3	4	5	4	4	4	4	4	3	3	3	4	4	3	3	4	4
1	1	2	3	3	3	3	2	1	1	1	3	3	2	2	3	3	3
3	3	3	4	1	3	5	1	1	4	1	3	4	3	1	4	3	3
5	4	4	5	5	4	4	2	1	4	4	5	5	2	3	4	5	5
3	5	1	3	3	4	2	2	5	5	5	3	1	3	3	4	5	5
5	4	2	3	4	4	5	5	4	5	4	3	3	3	3	4	4	5
5	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	3	5	5	5	5	5	5	5	5	3	5	5	5	5	5	5
4	5	2	4	3	4	1	4	2	3	3	3	4	3	4	3	4	4
5	5	5	5	3	2	5	5	5	5	5	2	5	3	3	4	5	5
5	5	3	5	3	3	5	5	5	4	3	3	4	5	3	5	5	5
3	2	5	5	4	3	3	5	2	4	2	3	5	4	3	4	3	4
5	3	2	3	3	2	5	5	4	4	4	2	3	2	2	2	3	4
5	2	2	4	2	3	5	5	4	4	4	3	2	2	3	2	3	4
4	3	3	3	3	2	5	5	4	4	4	3	3	2	2	2	4	5
0.8066667	0.786667	0.6066667	0.82	0.726667	0.72	0.8	0.793333	0.68	0.78	0.7266667	0.66	0.773333	0.633333	0.66	0.653333333	0.82666667	0.793333

The 48 cost drivers which were initially identified were narrowed down to 18 critical factors based on the weightages assigned to them by the surveyors. This was carried out to create a compact list of parameters as input for the artificial neural network, as a higher number of parameters could lead to inaccuracy of test result.

Table 2: Ranking of the significant cost drivers.

Parameter	RII	Rank
Sector	0.606666667	18
Seismic Zone	0.633333333	17
Construction Conditions	0.653333333	16
Level of Design Complexity (Modular or Non-Modular Design)	0.66	15
Risk Factor (High, Medium, Low)	0.66	14
Number of Basements	0.68	13
Soil Conditions	0.72	12
Ground Conditions	0.726666667	11
Type of Foundation (Isolated, Raft or Pile)	0.726666667	10
Market Conditions	0.773333333	9
Principal Structural Members (Masonry, R.C.C or Steel)	0.78	8
Location of buildings	0.786666667	7
Specifications (for finishes)	0.793333333	6
Number of Stories (or Height)	0.793333333	5
Plot Area, Plinth Area, Built-up Area	0.8	4
Type of Building	0.806666667	3
Complexity	0.82	2
Duration of Work	0.826666667	1

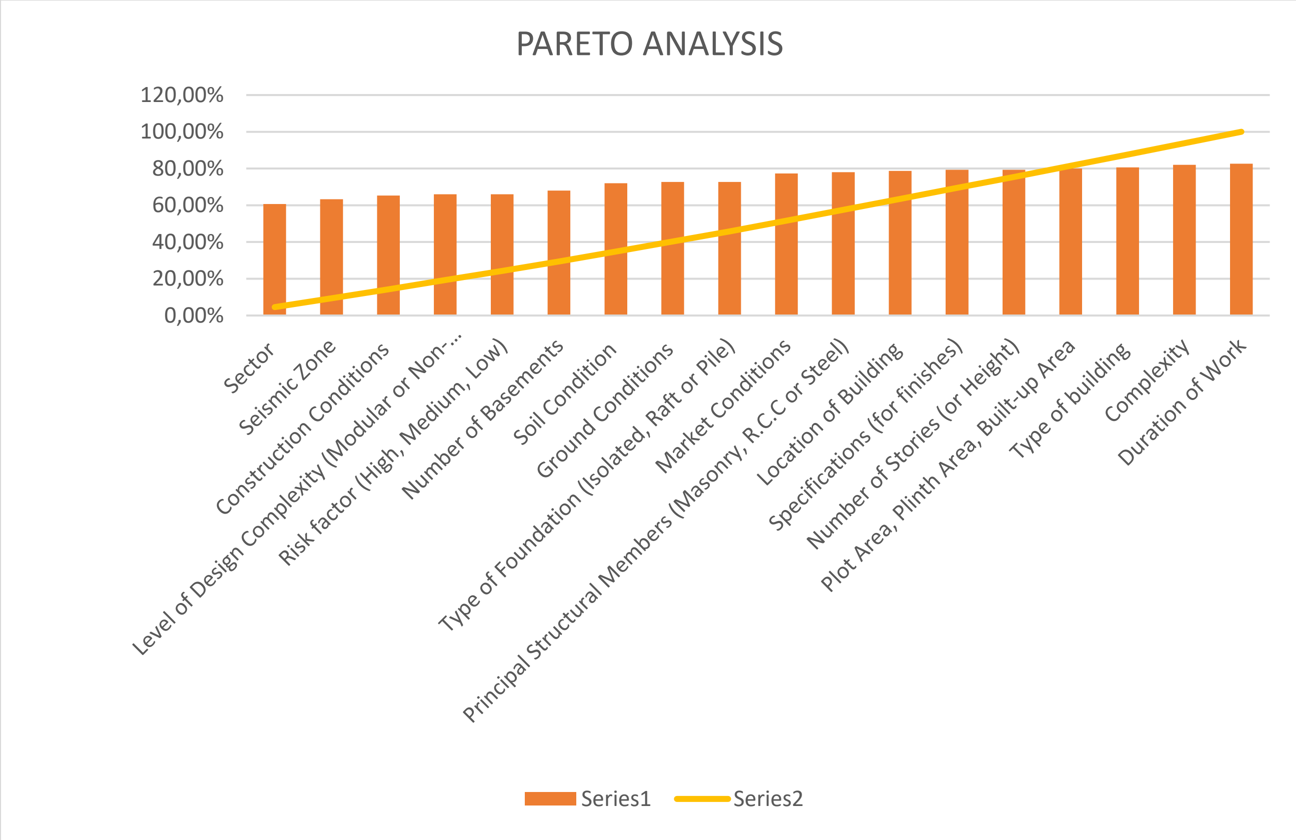

Figure 1. Pareto analysis of the significant cost drivers.

Fig. 1 depicts a Pareto analysis deployed to elicit the major issues to focus on to sort the cost estimate effectively. For example, targeting first 5 cost drivers would automatically take care of ~80% of the cost of construction.
From the rank chart based on survey, it is found that the following drivers have maximum impact on the cost of construction:

- Duration of Work
- Complexity of the Building (In terms of plan, elevation, interior spaces etc.)
- Plinth Area and Built-up Area
- Height
- Specifications (in terms of finishes, furniture etc.)

5. Problem Formulation for Initial Cost Estimation Using Artificial Neural Networks

A problem was formulated for cost estimation by dividing the critical cost drivers, identified after survey and ranking into numerical and categorical data with description (Refer to Table 4).

Table 4: Problem formulation for initial cost estimation of project.

Project characteristics	Type of information	Unit	Description
Type of building	Categorical	Number	1-Residential 2- Educational 3-Institutional 4- Assembly 5-Mercantile 6-Business and Industrial 7-Storage 8-Healthcare
Location	Categorical	Number	1-Central Business District 2-Metropolitan

			3-Regional
Seismic Zone	Numerical		
Sector	Categorical	Number	1-Government 2- Public (Statutory and Departmental) 3-Private Sector 4-Public-Private
Complexity (Dependent on constraints of material procurement, technological availability, legal and organizational issues)	Categorical	Number	1-Low 2-Medium 3-High
Ground Condition	Categorical	Number	1-Plane Land 2-Gentle Slope 3-Steep Slope
Soil Condition	Categorical	Number	1-Soft Soil (Sand and Gravel) 2-Silt and Clay 3-Rock 4-Black Cotton Soil
Plot Area	Quantitative	sq.m.	
Plinth Area	Quantitative	sq.m.	
Built-Up Area	Quantitative	sq.m.	
Number of Stories	Quantitative	Number	
Number of Basement	Quantitative	Number	
Principal Structural Material	Categorical	Number	1-Masonry 2-R.C.C 3-Steel
Type of foundation	Categorical	Number	1-Isolated 2-Pile Foundation 3-Raft Foundation
Level of Design Complexity			
Modular design	Categorical		
Market Conditions	Categorical	Number	1-Good 2-Fair 3-Bad
Construction Conditions	Categorical	Number	1-Good 2-Fair 3-Bad
Risk factor based on relevance	Categorical	Number	1-Low Relevance 2-Average Relevance 3-High relevance 4-Extremely High Relevance
Risk factor based on probability of unknown circumstances	Categorical	Number	1-Low Risk 2-Medium Risk 3-High Risk
Estimated duration of work	Quantitative	Number (in months)	
Specification			1-Poor 2-Average 3-Rich
Quality of Work			
Detailed Cost Estimate of Project**			

*The Risk quantification model estimates the probable consequences in terms of cost, schedule and quality. It is measured by rating the level of relevance with respect to the consequences.

6. Data Set and Training of Artificial Neural Network

A dataset was prepared based on the problem formulation sheet as input for the neural network (Figure-5) considering 10 construction projects which includes residential, commercial, institutional, healthcare and retail buildings.

PARAMETERS	Project A	Project B	Project C	Project D	Project E
TYPE OF BUILDING	8	8	1	3	6
LOCATION	1	2	2	2	2
SEISMIC ZONE	4	3	4	4	3
PROJECT COMPLEXITY	3	2	1	2	2
GROUND CONDITION	1	1	1	1	1
PLINTH AREA	9104.85	1150	1800	5040	5027
BUILT-UP AREA	108849	3467	4900	13100	10054
NUMBER OF STORIES	9	3	3	3	2
NUMBER OF BASEMENTS	3	0	0	1	0
PRINCIPAL STRUCTURAL MATERIAL	2	2	2	2	3
TYPE OF FOUNDATION	3	2	1	1	1
LEVEL OF DESIGN COMPLEXITY	3	2	1	2	1
MARKET CONDITIONS	1	1	1	1	2
CONSTRUCTION CONDITIONS	2	2	2	2	1
RISK FACTOR	2	2	2	2	2
IMPACT OF RISK FACTOR	3	3	1	2	1
ESTIMATED DURATION OF WORK	24	12	8	16	6
SPECIFICATION	3	3	2	3	2
QUALITY OF WORK	3	3	2	3	3
COST ESTIMATE	2363006505	407032054	563540744	936459256	69616000

PARAMETERS	Project F	Project G	Project H	Project I	Project J
TYPE OF BUILDING	1		4	1	1
LOCATION	3		1	3	3
SEISMIC ZONE	3		2	3	3
PROJECT COMPLEXITY	1		2	2	1
GROUND CONDITION	1		1	1	2
PLINTH AREA	166	2668	7343.92	110	2832
BUILT-UP AREA	720	8403	56953.46	271.55	34215
NUMBER OF STORIES	4	3	6	3	11
NUMBER OF BASEMENTS	0		2	0	0
PRINCIPAL STRUCTURAL MATERIAL	2		2	2	2
TYPE OF FOUNDATION	1		1	1	2
LEVEL OF DESIGN COMPLEXITY	1		2	2	2
MARKET CONDITIONS	1		1	1	2
CONSTRUCTION CONDITIONS	1		2	1	2
RISK FACTOR	1		1	1	1
IMPACT OF RISK FACTOR	1		2	1	1
ESTIMATED DURATION OF WORK	18		25	9	28
SPECIFICATION	1		3	2	2
QUALITY OF WORK	2		2	2	2
COST ESTIMATE	11495407	65402135.87	2796053034	9960000	696913416

Figure 2. Prepared data-set used as input in the Matlabs software

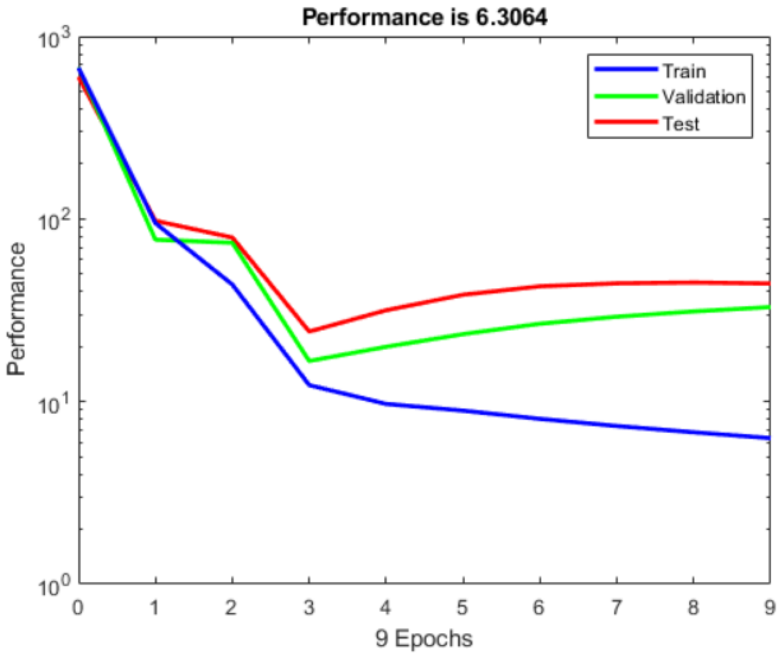

Figure 3. Mean Squared Error curve for trained data-set

The trained neural network shows a mean squared error of 6.0364. MSE is one of the best ways to check the performance of the trained neural network. While performing the testing of the network, the network with the least MSE was selected as the trained network.

MSE can be defined mathematically as (highlighted in Eq. 2:

- sum_i 1/n (desired_output(i) - model_output(i))^2

$$MSE = \frac{\sqrt{\sum_{i=1}^{n} (x_i - E(i))^2}}{n} \tag{2}$$

- A well trained neural network should have a very low value of MSE.
- The mean squared error is defined as the average of squares of the ‘errors’. It is the difference between the attribute which is to be estimated and the estimator.
- The above figure indicates that the training has shown a positive outcome since the validation and the test curves are nearly similar (as seen in Fig. 3).

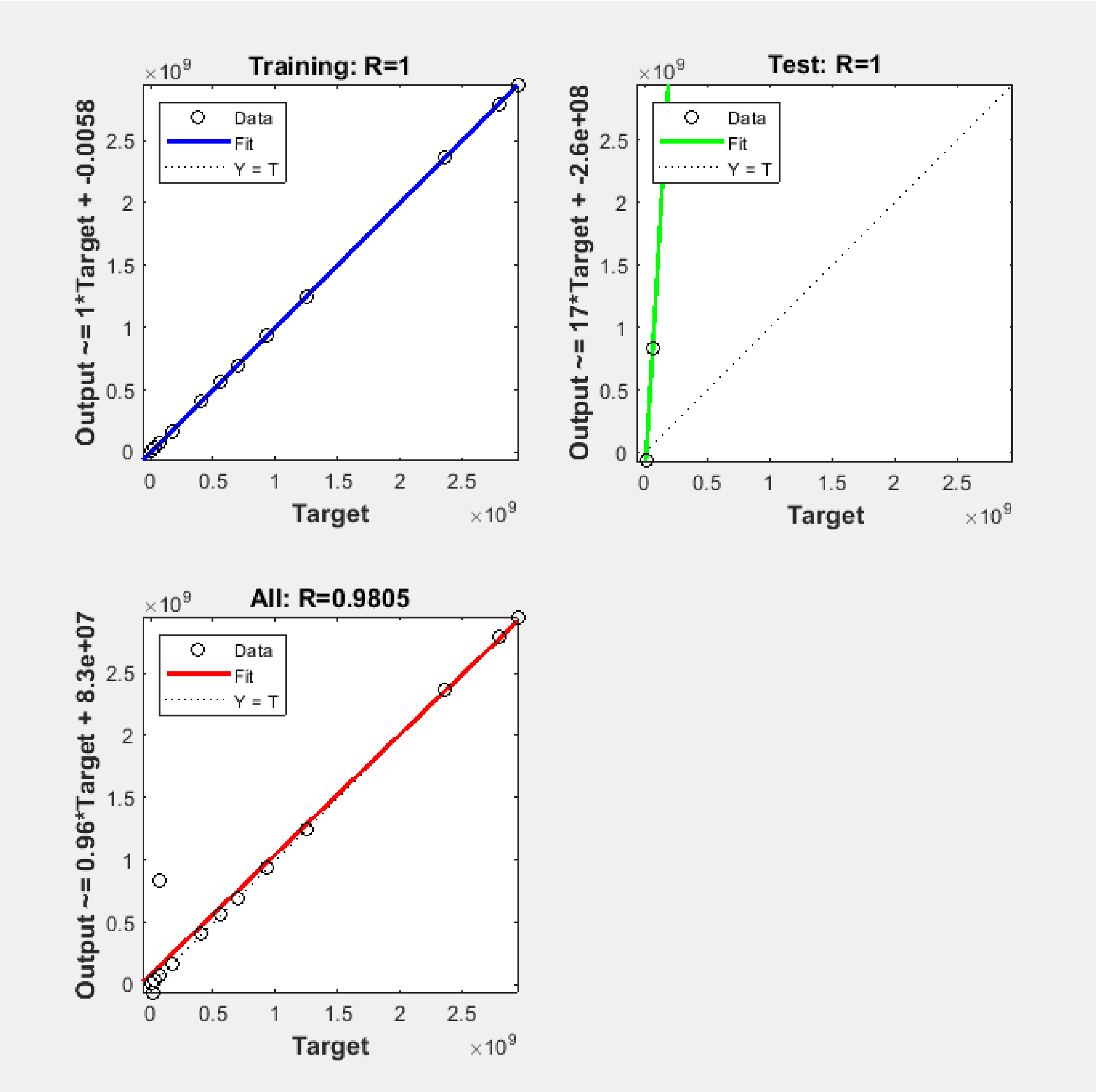

Figure 4. Regression Results for Training, Test and Target Data.

The coefficient of correlation or R values indicate whether the training data is a good or bad fit. It measures the direction and strength of the linear relationship between actual and predicted value.
An agreement can be seen between the actual and predicted lines, especially in the training and overall regression line, as indicated by the concentration of predicted values around the 45 ° line (Ongpeng, 2014).
An R value of 1 means a close relationship while 0 means a random relationship. A well trained neural network must have an R value of more than 0.8 (as seen in Fig. 4 the trained neural network shows an R values of 0.9805).

7. Prediction of Cost by The Trained ANN Model

Theoretically, testing the artificial neural network shall be same as training it. With the introduction of new data, the network needs to generalize the results from the test data and forecast the results. This evaluates the performance of the network after the training process.
For the purpose of testing the following new project dataset was taken:

TEST PROJECT PARAMETERS	AIIMS JAMMU-MEDICAL COLLEGE
TYPE OF BUILDING	3
LOCATION	2
SEISMIC ZONE	4
PROJECT COMPLEXITY	1
GROUND CONDITION	2
PLINTH AREA	3865
BUILT-UP AREA	17416
NUMBER OF STORIES	6
NUMBER OF BASEMENTS	0
PRINCIPAL STRUCTURAL MATERIAL	2
TYPE OF FOUNDATION	2
LEVEL OF DESIGN COMPLEXITY	2
MARKET CONDITIONS	2
CONSTRUCTION CONDITIONS	2
RISK FACTOR	1
IMPACT OF RISK FACTOR	2
ESTIMATED DURATION OF WORK	18
SPECIFICATION	1
QUALITY OF WORK	2

Figure 5. The test project dataset considered in the Matlabs software

For Fig. 5 (which shows the data of the test project for the trained ANN in Matlabs), the following results were obtained):

- Predicted value of Cost using the trained Artificial Neural Network: Rs. 691920130.2
- Actual Estimated Value of the project as per the BOQ: Rs. 614330223
- Percentage Difference in Estimated Value= +12.63%
- Therefore on the basis of the input data set and training of the ANN, the results are 87% accurate.

[Note: Only one project has been used as a test case. This is because the data size is low. Increasing the data size would have improved the accuracy of the result.]

7.1 Remarks for the Artificial Neural Network model in Matlabs

A three layer function (Fig. 6) has been used for the designed artificial neural network: It consisted of the input layer, hidden layer and output layer.

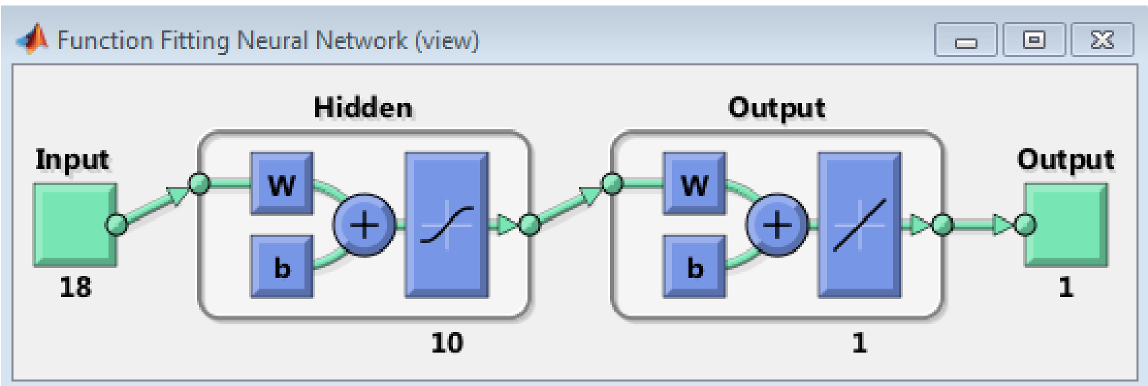

Figure 6. Three layer function analysis.

For the same set of data, the network has been trained multiple times till we reach the least MSE (Mean Squares Error) and Maximum R value (Coefficient of Correlation). A feed forward back propagation type of the network has been designed as it is the simple and most commonly used to generate the best model (Kulkarni, 2017). The backpropagation algorithm gradually reduces the error between the model output and the target output by minimizing the mean square error (MSE) over a set of training set (Murat Günaydın, 2004). Each time we initiate a feed-forward network, the network parameters become different, due to assignment of different weights.
The MSE is a good overall measure of the success of the training process. If the network is still not trained properly then we can try increasing the number of hidden neurons. This is because the larger number of hidden neurons give the network more flexibility as there are more parameters to optimize. However, if the hidden layers are too large then the problem itself might become under-characterized.
Two different types of training algorithms were tried for this project: Bayesian and Lavenberg-Marquardt. The final neural network was trained using the Bayesian regularization method. The Bayesian function (trainbr) is used for difficult, small or noisy datasets. The algorithm requires more time to execute than the Lavenberg-Marquardt (trainlm) function but shows better results. For this ANN the Lavenberg-Marquardt function showed the best coefficient of correlation result as 0.8385 whereas the Bayesian function has shown a much better coefficient of correlation result- 0.9805 indicating a better fit.
The ANN showed a better result when the training data set was increased. Providing additional data helps in producing a network that can generalize well to new data and shows more accurate prediction values.
Too much input data is also not good for the training of an artificial neural network as it can result in a common problem known as overfitting. In this case the error in the training set is driven to a small value because of the large data-set provided. But when a new data is added to the network, the error margin increases because the network has memorized the training samples but it has not learnt to generalize to new situations.

8. Conclusions

When different typologies of buildings across India were used for training, validation and testing of an Artificial Neural Network (taking a data-set with 15 samples or case studies) the following conclusions were drawn:

- The neural networks can successfully predict the early stage cost of construction for a building project.
- The accuracy of the predicted cost increases with the data size.
- With the increase in the sample size, the results showed a higher coefficient of correlation (R) and a lower Mean Squared Error (MSE).
- The neural network shows different training results and prediction outcomes if any of the parameters of the training data are altered or there is a change in weight from the original training. The training needs to be repeated till the best results are obtained.
- After repeated training iterations for this exercise, the training model with 10 hidden nodes gave the best results.
- For the training of construction cost data set in ANN, the Bayesian function (trainbr) shows better training results as compared to the Lavenberg-Marquardt (trainlm) function but the process takes more time to run (by a couple of seconds). The optimal neural network models involve a good combination of input data set, network architecture and training method.
- Apart from cost estimating in the decision making phase of the project, the ANN can also be used to find the lowest estimated bid of a supplier or the cost of insurance.

9. Limitations and Future Scope

The artificial neural network has a wide range of applications in the analogous field of cost estimation. It shows accurate results when the data sets, comprising of a large number of independent variables, have a non-linear structure or functional discontinuities.

Given the accurate results after successful training, the ANN can be considered as an alternative for parametric cost estimating especially during the earlier stages of the project. It is a great tool if we have a reliable historical data set. However the most tedious and expensive task in this process is to gather the right data for training and testing.
Artificial Neural Networks are yet to be used for live projects and a standard benchmark needs to be developed in order to determine the accuracy of the predicted results through ANN. They cannot be used where little or no information is available and they cannot be used for precise and arithmetic calculations (Mishra, 2018).
Apart from the early stage cost estimation of projects, the ANN can also be used to find the lowest estimated bid of a supplier or the cost of insurance. (Egemen, 2007).

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interests

The authors declare no conflict of interest.

References

Abraham, A. (2005). Artificial Neural Networks. Handbook of Measuring System Design. doi:10.1002/0471497398.mm421

Aziz Amen, M. (2022). The effects of buildings' physical characteristics on urban network centrality. Ain Shams Engineering Journal, 13(6), 101765. https://doi.org/10.1016/j.asej.2022.101765

Arafa, M. &. (2010). Early Stage Cost Estimation of Buildings Construction Projects using Artificial Neural Networks. *Journal of Artificial Intelligence, 4(1)*, 63–75. doi:10.3923/jai.2011.63.75

Chan. S, P. M. (2005). Project cost estimation using principal component regression. *Construction Management and Economics*, 295-304. doi:10.1080/01446190500039812

Egemen, M. &. (2007). A framework for contractors to reach strategically correct bid/no bid and mark-up size decisions. *Building and Environment, 42(3)*, 1373–1385. doi:10.1016/j.buildenv.2005.11.016

Ferry, D. B. (1999). Cost planning of buildings.

Herrera, R. F. (2020). Cost Overrun Causative Factors in Road Infrastructure Projects: A Frequency and Importance Analysis. *Applied Sciences, 10(16)*, 5506. doi:10.3390/app10165506

Huang, Y. (2009). Advances in Artificial Neural Networks – Methodological Development and Application. *Algorithms, 2(3)*, 973–1007. doi:10.3390/algor2030973

Jain, A. K. (1996). Artificial neural networks: a tutorial. *Computer, 29(3)*, 31–44. doi:10.1109/2.485891

Ji S.H, A. J. (2019). Cost Estimation Model Using Modified Parameters for Construction Projects. *Advances in Civil Engineering*. doi:10.1155/2019/8290935

Kulkarni, P. L. (2017). 'Artificial Neural Networks for Construction Management: A Review. *Soft Computing in Civil Engineering*. doi:10.22115/SCCE.2017.49580

Mishra, V. e. (2018). Comprehensive and comparative analysis of neural network. *International Journal of* Computer Application. doi:10.26808/rs.ca.i8v2.15

Murat Günaydın, H. &. (2004). A neural network approach for early cost estimation of structural systems of buildings. *International Journal of Project Management, 22(7)*, 595–602. doi:10.1016/j.ijproman.2004.04.002

Olotuah, A. O. (2002). Recourse to earth for low-cost housing in Nigeria. *Building and Environment, 37(1)*, 123-129. doi:10.1016/s0360-1323(00)00081-0

Ongpeng, J. &. (2014). An Artificial Neural Network Approach to Structural Cost Estimation of Building Projects in the Philippines. *DLSU Research Congress*.

Stoy, C. &. (2007). Residential Building Projects: Building Cost Indicators and Drivers. *Journal of Construction Engineering and Management, 133(2)*, 139–145. doi:10.1061/(asce)0733-9364(2007)133:2(139)

Takase, T. O. (2018). Effective neural network training with adaptive learning rate based on training loss. *Neural Networks*, 101, 68–78. doi:10.1016/j.neunet.2018.01.016

Toh, T. T. (2012). Critical Cost Factors of Building Construction Projects in Malaysia. *Procedia - Social and Behavioral Sciences, 57*, 360–367. doi:10.1016/j.sbspro.2012.09.1198

"What really drives construction cost?" http://businessbarbados.com/industries/construction/what-really-drives-construction-cost/