Examination of cost overrun in highway projects using Artificial Neural Networks in Kerala

Shemi,S,N
M.,Tech.Mechanical Engineering
College of Engineering,Trivandrum
Kerala, India
shemingr786@gmail.com

Asok Kumar N
Professor, Dept of Mechanical Engineering
College of Engineering, Trivandrum
Kerala, India
asok_akn@yahoo.com

***Abstract*— The trend of cost overrun in highway projects is one of the major challenge faced by the government since time immemorial. There has been many factors which results in providing a positive as well as negative impact to this phenomenon. The purpose of this paper is to determine the causal causes of cost escalation of highway projects in Kerala and in turn develop Regression models and neural network models for analysis of cost overrun for the future projects.**

Keywords*— *Cost overrun, Regression, Artificial Neural Network (ANN)

1. INTRODUCTION

Cost control is a challenging task in project budget. It gives cost related details and data for taking choices to finish the project in a characterized amount of materials, in time, and inside the budget. Sometimes expenses acquired in excess of budgeted amount because of an underestimation of the real expense during budgeting. Which is known as cost overrun. The term cost overrun is frequently alluded to as a budget increment, cost increment, or cost development. Cost overrun are basic in infrastructure, building, and innovation projects. Progressively normal in infrastructure extends particularly, increasingly regular in road development works.

Kerala has smaller highways contrasted with different parts of India. The explanations behind this are trouble and lack of engagement in land acquisition. Numerous roadway projects extending and maintenance had quit working because of the fights from ideological groups. and furthermore road repair works are continued pending which brings about the development of potholes prompts problematic journey. This prompts the poor condition of roads in Kerala. Subsequently the task costs increase day by day. Cost variation happen because of progress of material costs, economy, work profitability, and deferral in repair and maintenance and so forth.

A few reasons cause cost overrun. A typical purpose behind cost overrun is the incorrectness of cost estimate. Because of step by step increment in material costs and work costs may prompt increment in development cost than evaluated cost. Such cost overrun are expected either to off base appraisals or to changed conditions in the commercial center. Survey can accomplish for cost estimate before putting requests to distinguish mistakes or changed conditions. A general audit may find that increments in certain areas are remunerated by diminishes in others. Customers can modify prerequisites to decrease expenses or search out lower-cost providers. Informing the entrepreneurs or supervisors regarding conceivable greater expenses at this stage gives them the choice of making changes and keeping up their financial limit.

This examination points a quantitative methodology in recognizing and surveying the different components influencing the cost overrun in highway project in Kerala. The study begins by creating a model for anticipate the cost overrun dependent on questionnaire survey. This paper is further comprised of the accompanying areas: literature review, research methodology, result and discussion, validation of result and conclusion.

2. LITERATURE REVIEW

2.1 Review on causes of cost overrun

Cost overrun can happen because of a wide range of causes on different kinds of projects. In the event that undertaking expenses surpass their arranged targets, customer fulfillment would be undermined. Globally different investigations have been directed to recognize the reasons for time overrun in highway development projects.

1. Alhomidan (2013)[2] investigated the 41 prime factors which were causing the overrun of cost in road projects, and conducted a survey to find out the most influencing factors. These factors were

classified into various groups according to their origin of delay these are project, managerial factors, consultant, external factors, constructional commodities, and financial. Frequency and severity of each factor is determined by the contractors. The study came to the conclusion that delay in payments, internal administrative and organizational problems, poor mode of communication between the project parties, delays in taking appropriate decisions were the most deciding factors for cost overrun.

2. Mahamid(2013)[7] investigated the 45 factors that might cause delays in construction in western bank of Palestine from contractors. By means of questionnaire survey it was found that financial status of contractors, payment delays by the owner, political situation and segmentation of western bank, lack of interaction between people associated with the project, absence of equipment efficiency, increasing rate of competition in bidings were the primal reasons for the time overruns in Palestine.

3. Meeampol *et al*. (2006)[8] In Thailand investigated the expense and time execution of roadway project from the perspective of the public owner. The consequences of the investigation uncovered quality administration, budget management, human resource management, owner involment and group connections obstruct time execution. While, scheduled management and human resource management executives repress cost execution.

4. Kaliba *et al*. (2008)[6] featured that cost overrun of road development extends in Zambia are brought about by elements, for example, postponed installments, money related procedures and troubles with respect to contractors and customers, contract adjustment, monetary issues and materials acquisition. While, terrible or harsh climate because of substantial rains and floods, scope changes, natural assurance and relief costs, plan postponement, strikes and specialized difficulties are the significant reasons for cost acceleration in Zambia's road development project.

5. Nasir et al. (2011)[9] in Pakistan so as to examine key components causing cost and time overrun in highway project through writing audit and poll overview. The consequences of the examination demonstrated that variables like postponement in progress installments to temporary workers, land acquisition process, value heightening of significant development materials, improper planning, contractual worker's lack of ability to carry out the responsibility, delay in giving over the site to the contract based worker, extra work/scope changes and improper government strategies and needs were distinguished to had a severe effect on the length and cost of a roadway project in Pakistan.

2.2 Review on modeling of cost overrun

A few examinations have endeavored to give techniques to take care of the issue of anticipating development cost and time and cost overrun of construction project.

1. Mahamid (2013)[7] utilized multiple regression analysis methods to create models for foreseeing the cost deviation in street extends in the West Bank – Palestine as for the undertakings' physical attributes, for example, project size, road length, road width etc. The examination results indicated that project size has a decent connection with cost deviation

2. Choudhury et al. (2004)[3] in this study a stepwise regression model with a forward choice strategy was utilized to associate time overrun in business development project with 9 distinguished autonomous factors. The outcomes demonstrated that only three autonomous factors have associations with time overrun in business development project at a degree of centrality of <0.20.

3. El-Kholy (El-Kholy, 2013; El-Kholy (2015)[4,5] directed two examinations to demonstrate development postponement and cost overrun of construction by utilizing Fuzzy methodology and Case-based thinking strategy separately. In these two investigations the most critical makes that lead postponement and cost overrun in the Egyptian development project are distinguished as the free factors of the proposed model.

4. Ahmed Yousry Akal (2017)[1] The fundamental commitment of this examination is foreseeing the rates of calendar overwhelm and cost acceleration of expressway extends in Egypt. Through the utilization of the linear regression analysis method and statistical fuzzy theory, four prescient models have been created and it has been noticed that the linear regression based model shows expectation exactness better than measurable fuzzy based model in anticipating rates of cost overrun and cost acceleration.

5. Yasser Elfahham (2018)[10] In this investigation Estimating construction costs and foreseeing value acceleration are significant advances. The construction cost index (CCI) has been broadly used to figure project costs. Neural Networks, Linear Regression, and Autoregressive Time Series are then used to conjecture the CCI.

3. OBJECTIVE

1. Identifying the critical factors causing cost overrun in highway project of Kerala

2. Modeling the percentages of cost overrun of highway projects in Kerala using multiple regression analysis method and neural network method.

4. RESAERCH METHODOLOGY

So as to accomplish the examination targets, the investigation system includes different stages. The succession of such stages is introduced in Figure 1.

A questionnaire was then drawn up and was separated into two segments. Segment A tried to know the general points of interest of the respondents while segment B was centered around the impacts of cost overrun. The questionnaire was done among three gatherings of development experts which are identified with architects, engineers and contractors related to road construction. In the planned questionnaire the respondents need to pick one out of the four potential answers speaking to fluctuating degrees of impact on a size of 1 to 4.

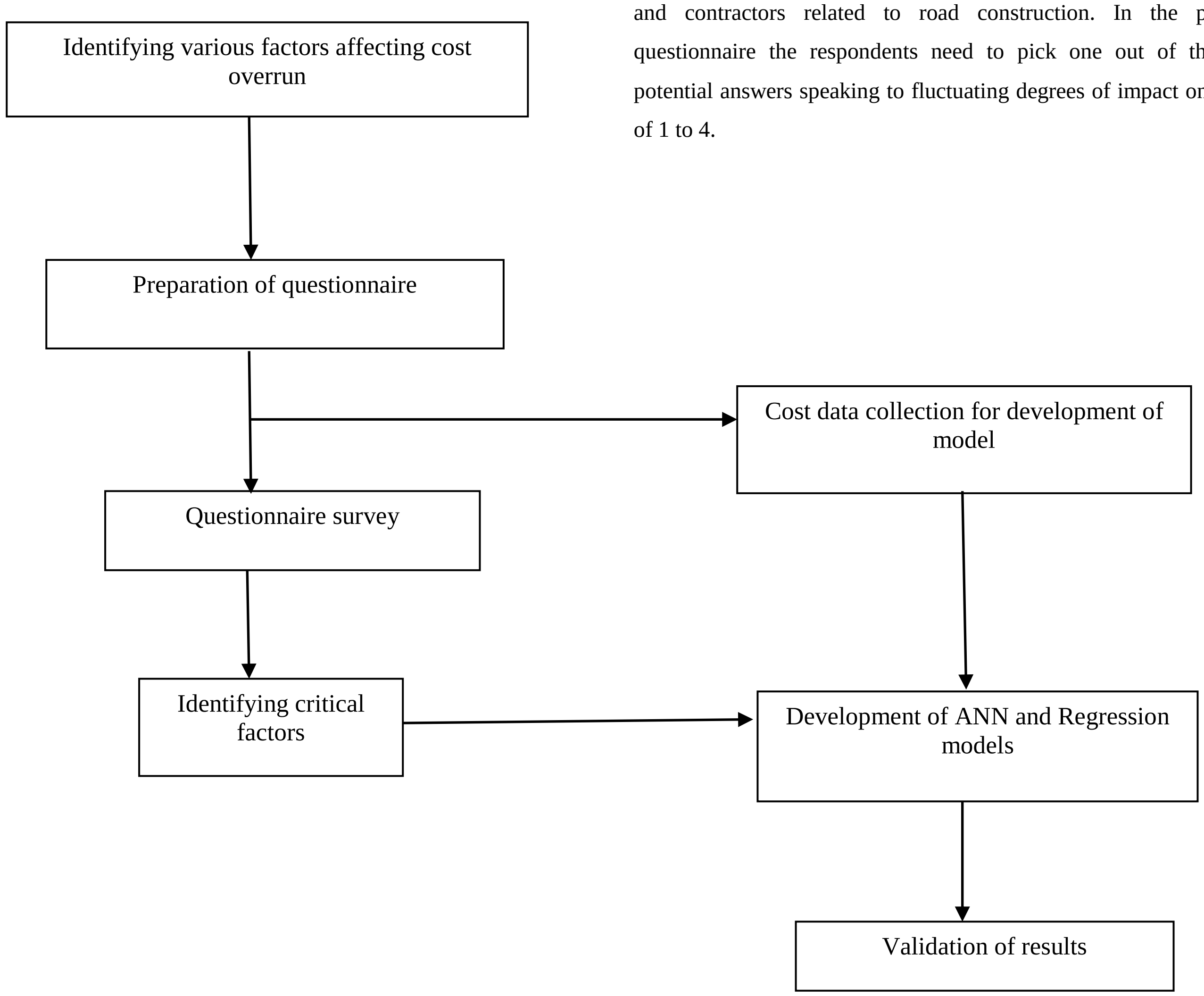


Fig.1 Overview of methodology

4.1 Identification of various factors affecting cost overrun

In the current study 40 variables have been recognized as reasons for cost overrun in highway construction projects, which have been accumulated from the previous research papers. After the variables which may cause cost overrun in highway construction projects have about the viability level of the identified factors.

A reaction of 1 means the factor is ineffective, 2 means the factor has medium effect, 3 means the factor is effective and 4 means the factor is very effective.

4.2 Identification of critical factors

In order to assess the viability level of the distinguished elements, Relative Importance Index (RII) has been recommended.

Eq. (1): Relative Importance index (RRI) under consideration

$$RII = \frac{\sum W}{(H \times N)} \qquad (1)$$

Where

$\sum W$ = total weight given to each factor by the respondents

H = highest ranking available

N = total number of respondents that answered the question

In view of the RII value the components were ranked and distinguish the most basic elements. The ten basic elements were chosen as contribution for both multiple regression and ANN models for forecast of cost overrun. Table 1 shows the ranking of factors based on the RII value.

Table 1. RII value

Effective factor	RII	Rank
Late issuing of approval documents	81.92358	1
Poor workmanship	80.95238	2
Poor communication between construction parties	80.95238	2
Late submission of nominated materials	80.95238	2
Mistake in design	79.04762	3
Delays in decision making	79.04762	3
Payments delay	77.14286	4
Availability of materials	77.12078	5
Limited construction area	76.19048	6
Design changes	76.19048	6
Undefined scope of working	74.28571	7
Resource management	74.28571	7
Strike	73.33333	8
Inappropriate design	73.33333	8
Financial status of contractor	73.33333	8
Construction season	72.38095	9
Postponement of project	71.42857	10
Pavement thickness	69.52381	11
Lack of facilities	69.52381	11
Traffic maintenance	68.57143	12
Unreasonable project cost frame	68.57143	12
Improper construction method	68.57143	12
Late land hand-over	67.61905	13
Internal administrative problems	67.61905	13
Late approval	63.80952	14
Delay in commencement	61.90476	15
Changes in management ways	60.95238	16
Type of pavement	60	17
High competition in bids	60	17
Resurfacing	59.04762	18
Late documentation	59.04762	18
Incapable inspectors	58.09524	19
Equipment failure	57.14286	19
Hauling distance	56.19048	20
Traffic volume	56.19048	20
Late inspection	55.2381	21
Insufficient equipment	53.33333	22
Disturbance to public activities	49.52381	23
Late design works	46.66667	24
Inflation	42.85714	25

4.3 Multiple regression model

Multiple regression analysis is an incredible procedure utilized for anticipating the obscure estimation of a variable from the known estimation of at least two variables. By utilizing the multiple regression analysis the cost overrun model is obtained. Table 2 shows the regression coefficients.

Table 2 regression coefficient

Coefficients					
Model	Unstandardized		Standardized	t	Sig.
	B	Std. error	Beta		
constant	-30.27	18.321		-1.43	0.12
LIAD	3.242	0.845	0.786	3.23	0.02
PW	-10.32	3.345	-1.423	-1.52	0.10
PCBP	-0.292	3.907	-0.042	-0.07	0.82
LSNM	2.134	1.345	0.456	1.112	0.241
MD	1.786	1.700	0.345	1.354	0.234
DDM	6.542	1.213	1.012	2.45	0.089
PD	1.256	1.654	0.012	0.412	0.234
AM	1.346	1.678	0.034	0.756	0.456
LCA	3.976	1.098	0.762	3.671	0.012
DC	3.924	2.345	0.876	1.456	0.123

Utilizing the unstandardized coefficients got for each factor from the model, the model equation acquired was shown indicated in Eg. (2):

$$Y = -30.27 + 3.242 \text{ LIAD} - 10..32 \text{ PW} - 0.292 \text{ PCBP} + 2.134 \text{ LSNM} + 1.786 \text{ MD} + 6.542 \text{ DDM} + 1.256 \text{ PD} + 1.346 \text{ AM} + 3.976 \text{ LCA} + 3.924 \text{ DC} \quad (2)$$

Where

LIAD = Late issuing of approval documents

PW = Poor workmanship

PCBP = Poor communication between construction parties

LSNM = Late submission of nominated materials

MD = Mistake in design

DDM = Delays in decision making

PD = Payments delay

AM = Availability of materials

LCA = Limited construction area

DC = Design changes

The table 3 shows the model summary of the regression. The basic static related with regression analysis is R^2 which is the deviation of the genuine estimation of the reliant variable from the regression line. The sum of squared deviations about the regression line is a proportion of the degree to which the regression neglects to clarify the depended variable. Thus, the R^2 measurement is a proportion of the degree to which the all out variety of the dependent variable is clarified by the regression.

Table 3 Model summary

Model	R	R^2	Adjusted R^2	Std, Error of the estimate
	0.934	0.925	0.830	1.2134

The all out variation is characterized by all the factors are 83% that is the project is possible from information perspective with just 17% loss of information. Right now, = 0.934 which is almost equivalents to 1, henceforth the relationship can be discovered with less measure of misfortune in information.

The regression models acquired is approved utilizing ten project. The plot demonstrating the real cost overrun versus anticipated cost overrun shows a R^2 value of 0.910. The rate of error for anticipated and actual cost overrun lies between 2% to 35 %.

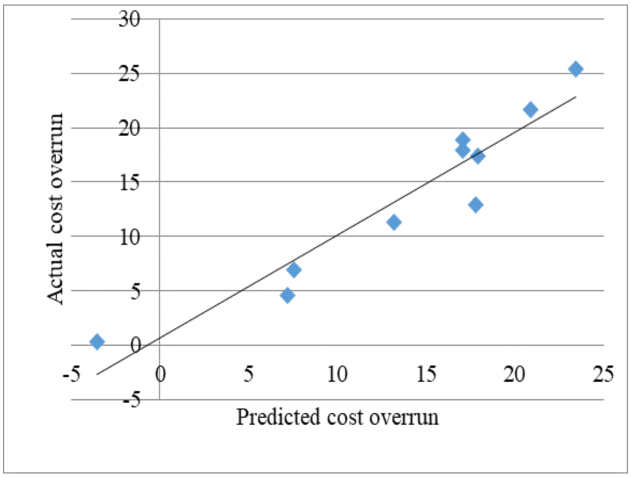


Fig. 2 actual cost vs predicted cost overrun

4.4 Artificial Neural Network model

Neural network models were created using MATLAB R2011b software. The models were created in an experimentation way to recognize ideal network parameters and network execution. A few systems were prepared utilizing the info factors, for example, chose ten basic components and the yield factor which is the cost overrun. In light of a fittingly arranged ANN model and an adequate arrangement of past finished roadway projects, an ANN model had the option to show up at exact estimates of the expense of another construction highway projects. Figure 3 shows steps used in ANN.

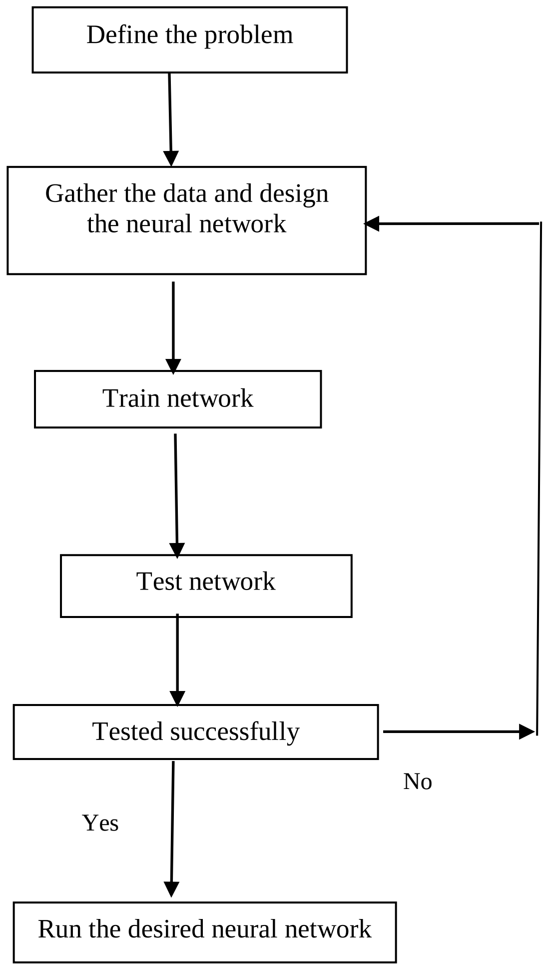


Fig. 3 Flow chart for neural network

4.4.1 Training the network

The objective of training neural network is to get a network that performs best on unseen data through training many networks on a training set. The following results were used to compare the network problems or errors on a predetermined set. Therefore, several network parameters such as number of hidden layers, number of hidden nodes, transfer functions and learning rules were trained multiple times to produce the best weights for the model. In training feed forward back propagation algorithm is used as network architecture and TRAINLM as the learning function.

The training of network architecture was carried out using 75% of the data. Neural network models for different hidden layers in training phase are shown in Table 4. The training was continued for 7 iterations with a gradient of 0.0045 and learning rate of 0.0444. The best architecture for training was obtained as 10-13-1 consists of thirteen hidden layers with a correlation value of 0.9999 which indicates the best training performance.

Table 4 ANN training

Architecture	Gradient	Mu	Correlation R
10-1-1	0.0113	1.00e-06	0.9684
10-2-1	0.00174	1.00e-07	0.964
10-3-1	0.00348	1.00e+10	0.7985
10-4-1	0.00066	1.00e+11	0.9941
10-5-1	0.00187	1.00e-06	0.8733
10-6-1	0.00101	1.00e+11	0.9810
10-7-1	0.00646	1.00e-07	0.9827
10-8-1	1.62e-06	1.00e-06	0.9763
10-9-1	6.06e-07	1.00e-07	0.9911
10-10-1	0.0010	1.00e+10	0.9920
10-11-1	0.00248	1.00e+03	0.9981
10-12-1	4.84e-06	1.00e-07	0.9942
10-13-1	**0.00045**	**1.00e+10**	**0.9999**
10-14-1	3.38e-10	1.00e-07	0.9995
10-15-1	0.00013	1.00e+10	0.9961
10-16-1	4.52e-08	1.00e-06	0.9983
10-17-1	0.00246	1.00e+10	0.8281
10-18-1	6.80e-09	1.00e-06	0.9267
10-19-1	3.78e-09	1.00e-06	0.9993
10-20-1	1.38e-05	1.00e+10	0.9918

The plot shown in Fig. 4 shows the training performance of the network with thirteen hidden layers

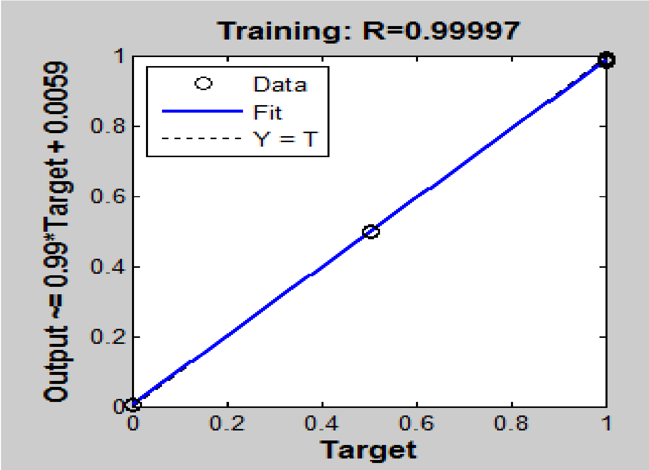


Fig. 4 Training performance

The R value provides an evidence for the realtion between the solutions and the objectives. If R = 1, is the evidence for a linear relation between the objective and the solutions. As R approaches zero, represents that there is no relationship. Here the training data indicates a good fit. The output cost overrun approximately equal to 0.99 target value where 0.99 is the slope of the fit.

4.4.2 Testing the network

Testing the system is basically equivalent to training it, then again, actually the system is demonstrated realities it has never observed, and no corrections are made. At the point when the system isn't right, it is essential to assess the presentation of the system after the preparation procedure. In the event that the outcomes are acceptable, the network is prepared to utilize. If not, this implies it needs more or better information or even re-structures the network.

Like training, testing was conveyed with various concealed layers for ten sources of input. 15% of in general information were tried. The testing results for various hidden layers were appeared in Table 5

Table 5 ANN testing

Architecture	Gradient	Mu	Correlation R
10-1-1	0.0113	1.00e-06	0.8246
10-2-1	0.00174	1.00e-07	0.862
10-3-1	0.00348	1.00e+10	0.1749
10-4-1	0.00066	1.00e+11	0.9457
10-5-1	0.00187	1.00e-06	0.9874
10-6-1	0.00101	1.00e+11	0.9562
10-7-1	0.00646	1.00e-07	0.8522
10-8-1	1.62e-06	1.00e-06	0.9971
10-9-1	6.06e-07	1.00e-07	0.9067
10-10-1	0.0010	1.00e+10	0.9823
10-11-1	0.00248	1.00e+03	0.8479
10-12-1	4.84e-06	1.00e-07	0.8912
10-13-1	**0.00045**	**1.00e+10**	**0.9967**
10-14-1	3.38e-10	1.00e-07	0.8859
10-15-1	0.00013	1.00e+10	0.7838
10-16-1	4.52e-08	1.00e-06	0.9258
10-17-1	0.00246	1.00e+10	0.8685
10-18-1	6.80e-09	1.00e-06	0.9959
10-19-1	3.78e-09	1.00e-06	0.9715
10-20-1	1.38e-05	1.00e+10	0.9215

During testing stage in which the relationship esteem differs from 0.1 to 0.9. The best training architecture got is connect with thirteen hidden layers with a relationship of 0.9967. Here the result is around equivalent to 0.87 of target esteem.

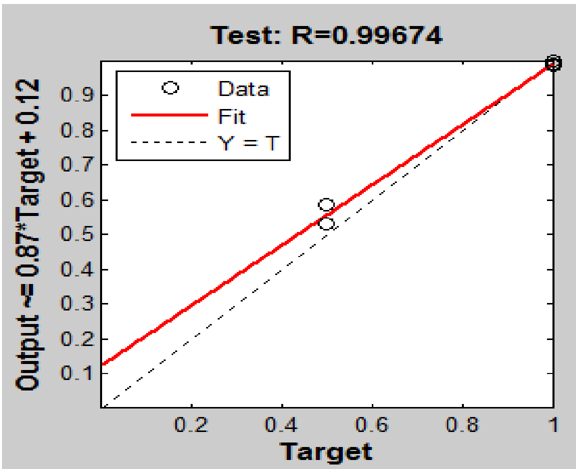


Fig. 5 Testing performance

4.5 Validation

Validation of data was completed by neural network apparatus itself. 10% of data is utilized for validation. During iterative training of a neural network, an age is a solitary go through the whole training set, followed by testing of the validation set. The validation results for various hidden layers are shown in Table 6.

From the table it was noticed that the system with thirteen hidden layers has a correlation of 0.9531 which is more noteworthy than 0.9 which shows entire data was validated.

Table 6 Validation of result

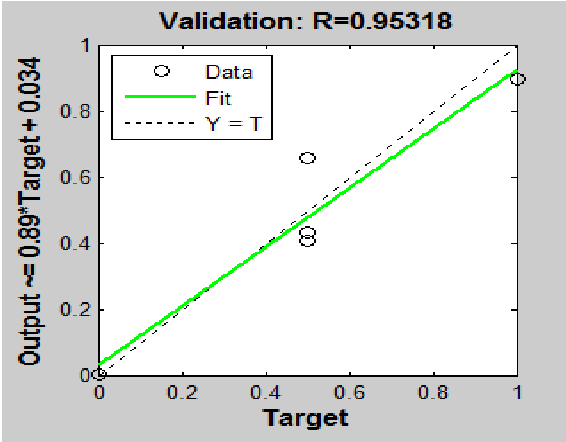
Architecture	Gradient	Mu	Correlation R
10-1-1	0.0113	1.00e-06	0.9980
10-2-1	0.00174	1.00e-07	0.976
10-3-1	0.00348	1.00e+11	0.9702
10-4-1	0.00066	1.00e-06	0.9927
10-5-1	0.00187	1.00e+11	0.9615
10-6-1	0.00101	1.00e-07	0.9846
10-7-1	0.00646	1.00e-07	0.9623
10-8-1	1.62e-06	1.00e-06	0.9802
10-9-1	6.06e-07	1.00e-07	0.9820
10-10-1	0.0010	1.00e+10	0.9664
10-11-1	0.00248	1.00e+03	0.9634
10-12-1	4.84e-06	1.00e-07	0.6950
10-13-1	**0.00045**	**1.00e+10**	**0.9531**
10-14-1	3.38e-10	1.00e-07	0.9285
10-15-1	0.00013	1.00e+10	0.7689
10-16-1	4.52e-08	1.00e-06	0.9352
10-17-1	0.00246	1.00e+10	0.6276
10-18-1	6.80e-09	1.00e-06	0.9343
10-19-1	3.78e-09	1.00e-06	0.6731
10-20-1	1.38e-05	1.00e+10	0.9733

Fig. 4 shows the validation plot for network architecture 10-13-1. Here the output value is roughly equivalent to 0.89 of target value. The cost overrun acquired from validation was appeared in Eq. (3)

$$\text{Cost overrun, } Y = 0.89 \text{ Target value} + 0.034 \quad (3)$$

The slope of the line acquired from validation is $\beta = 0.89$ and the intercept value $\alpha = 0.034$.

Fig. 6 Performance of ANN during validation



5. CONCLUSION

The cost overrun in road construction projects in the Kerala is examined in a field review. It considered significance of the causes influencing cost estimating accuracy. 40 reasons for cost overrun were distinguished through research. Followed by developing statistical models using multiple regression and neural network method. The regression model can be used to determine the correlation between the selected factors and cost of the projects and then the results of the model developed using ANN are validated statistically. The all out variety is characterized by all the factors are 83% that is the task is feasible from information perspective with just 17% loss of information. In neural network models the best execution of engineering were gotten was 10-13-1 with 10 inputs and 13 hidden layers with one output which is cost overrun shows the best measurable model. On comparing the outcomes got from the regression and neural network, neural network shows the most probabilistic model as these networks.

The main contribution of this study is to predicting the cost overrun of the future highway projects in Kerala. Recommendations for future works includes using different model parameters such as different factors, training algorithm, activation functions with extension to the feed forward algorithm. Future investigations could be performed for various explicit kinds of development project, for example, railroad development project, building lodging project, viaducts and dam development project, and so on.

References

[1] Ahmed Yousry Akal, (2017), "Modeling schedule overrun and cost escalation percentages of highway projects using fuzzy approach", *Journal of Engineering, Construction and Architectural Management*

[2] Alhomidan, A. (2013), "Factors affecting cost overrun in road construction projects in Saudi Arabia", *International Journal of Civil and Environmental Engineering*, Vol. 113, No.03, pp. 1-4.

[3] Choudhury, I.(2004), "Correlates of time overrun in commercial construction, *Proceedings of the 40th Annual Conference Brigham Young University - Provo*, Utah.

[4] El-Kholy, A. M. (2013), "Modeling delay percentage of construction projects in Egypt using statistical-fuzzy approach", *IOSR Journal of Mechanical and Civil Engineering*, Vol. 7, No. 5, pp. 47-58.

[5] El-Kholy, A. M. (2015), "Predicting cost overrun in construction projects", *International Journal of Construction Engineering and Management*, Vol.4, No.4, pp. 95 105.

[6] Kaliba, C. (2009), "Cost escalation and schedule delays in road construction projects in Zambia", *International Journal of Project Management*, Vol. 27, No. 5, pp. 522-531.

[7] Mahamid, I. (2013), "Frequency of time overrun causes in road construction projects in Palestine: contractors' view", *Organization, Technology and Management: an International Journal*, Vol. 5, No. 1, pp. 720-729.

[8] Meeampol, S. (2006), "Factors affecting cost and time performance on highway construction projects: evidence from Thailand", *Journal of Financial Management of Property and Construction*, Vol. 11, No. 1, pp. 3-20.

[9] Nasir, A. R. (2011). "Cost and time overruns in highway projects of Pakistan", *Proceedings of the 6th International Conference on Construction in the 21st Century*, Kuala Lumpur, Malaysia, pp. 69-76.

[10] Yasser Elfahham (2018), "Estimation and prediction of constructioncost index using neural networks, time series, and regression", *Alexandria Engineering Journal (2019) 58, 499–506*