

**PUBLIC BUILDING COST PREDICTION MODEL IN OROMIA REGION  
USING ARTIFICIAL NEURAL NETWORK**



**BURHAN MUSA**

**A THESIS SUBMITTED TO  
THE DEPARTMENT OF CIVIL ENGINEERING  
SCHOOL OF CIVIL ENGINEERING AND ARCHITECTURE**

**PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE  
DEGREE OF MASTER'S IN CONSTRUCTION ENGINEERING MANAGEMENT**

**OFFICE OF GRADUATE STUDIES  
ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY**

**MAY, 2024**

**ADAMA, ETHIOPIA**

**Public Building Cost Prediction Model in Oromia Region  
Using Artificial Neural Network**

**Burhan Musa Adem**

**Advisor: Bahiru Bewket (Phd)**

**A Thesis Submitted to  
The Department of Civil Engineering  
School of Civil Engineering and Architecture**

**Presented in Partial Fulfillment of The Requirement for The Degree of Master's in  
Construction Engineering Management**

**Office of Graduate Studies  
Adama Science and Technology University**

**May, 2024**

**Adama, Ethiopia**

### Approval Page of M.Sc. Thesis

I/we, the advisors of the thesis entitled “Public Building Cost Prediction Model in Oromia Region Using Artificial Neural Network” and developed by Burhan Musa hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

_____	_____	_____
Advisor	Signature	Date

We, the undersigned, members of the Board of Examiners of the thesis by Burhan Musa have read and evaluated the thesis entitled “Public Building Cost Prediction Model in Oromia Region Using Artificial Neural Network” and examined the candidate during open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Master of Science in Construction Engineering Management.

_____	_____	_____
Chairperson	Signature	Date

_____	_____	_____
Internal Examiner	Signature	Date

_____	_____	_____
Eternal Examiner	Signature	Date

Finally, approval and acceptance of the thesis is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

_____	_____	_____
Department Head	Signature	Date

_____	_____	_____
School Dean	Signature	Date

_____	_____	_____
Office of Postgraduate Studies, Dean	Signature	Date

### **Declaration**

I hereby declare that this Master Thesis entitled “Public Building Cost Prediction Model in Oromia Region Using Artificial Neural Network” is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation.

---

Name of student

---

Signature

---

Date

### **Recommendation**

I, the advisor of this thesis, hereby certify that I have read the revised version of the thesis entitled “Public Building Cost Prediction Model in Oromia Region Using Artificial Neural Network” prepared under my guidance by Burhan Musa submitted in partial fulfillment of the requirements for the degree of Master’s of Science in Construction Engineering Management.

Therefore, I recommend the submission of revised version of the thesis to the department following the applicable procedures.

---

Advisor

Signature

Date

## ACKNOWLEDGEMENT

I would like to express my deepest gratitude to those who have supported me throughout the journey of completing this master's thesis. Their guidance, encouragement, and unwavering support have been invaluable.

First and foremost, I extend my heartfelt thanks to my adviser, Bahiru Bewket (Phd). Your expertise, insight, and patience have been instrumental in guiding my research. Your constructive feedback and constant encouragement have significantly contributed to the quality of this work. I am truly fortunate to have had the opportunity to work under your mentorship.

I also wish to acknowledge the support of the Department of Civil Engineering at Adama Science and Technology University and staff have provided an inspiring and conducive environment for academic growth.

To my beloved wife, Shamsiya your love, patience, and understanding have been my anchor. Your unwavering support has given me the strength to persevere through the challenges of this journey. I am deeply grateful for your sacrifices and constant encouragement.

To my wonderful sons, Ehsan and Emran, thank you for your patience and for being a source of joy and motivation. Your smiles and unconditional love have been a tremendous source of strength.

To my father, mother, brother and sisters, your unwavering belief in me and your constant encouragement have been a cornerstone of my success. Thank you for your love, support, and for always being there for me.

I am also immensely grateful to my friends, Astro, who have been a pillar of support. Your words of encouragement, understanding, and camaraderie have been invaluable throughout this journey. Thank you for always being there for me.

Lastly, I would like to acknowledge everyone who has supported me in one way or another during the course of this research. Your contributions, no matter how small, have made a significant impact.

Thank you all for your unwavering support and belief in me.

Sincerely,

Burhan Musa

## Table of Contents

LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
Abstract .....	xi
1.INTRODUCTION .....	1
1.1 Background .....	1
1.2 Objectives .....	2
1.2.1 General Objective: - .....	2
1.2.2 Specific Objective: - .....	2
1.3 Problem Statement .....	2
1.4 Scope.....	2
1.5 Significance of the Study .....	3
1.5 Delimitation and Limitation.....	3
2. LITERATURE REVIEW .....	4
2.1 Traditional Cost Estimation Methods .....	4
2.2 Artificial Neural Network (ANN).....	6
2.2.1 Nodes (Neurons): - .....	7
2.2.2 Connections (Weights): - .....	8
2.2.3 Layers: - .....	8
2.2.4 Activation Function: - .....	8
2.2.5 Feed forward and Back propagation: - .....	10
2.2.6 Loss Function: - .....	11
2.2.7 Optimization Algorithms: - .....	11
2.2.8 Hyper parameters: - .....	11
2.3 Role of Artificial Neural Networks (ANN) in Construction Management....	11
2.4 Role of Artificial Neural Networks (ANN) in Construction Cost Prediction	16
3. METHODOLOGY .....	25
3.1 Research Design.....	25
3.2 Data Collection .....	26
3.2.1 Population of Study: - .....	26

3.2.2	Sample and Sampling Technique: - .....	27
3.2.3	Sample Size: - .....	27
3.2.4	Research Location: - .....	28
3.3	Study Population Characteristics .....	29
3.3.1	Types of company: -.....	29
3.3.2	Job Title: - .....	29
3.3.3	Years of experience in construction projects: -.....	30
3.4	Independent Variables .....	30
3.5	Data Encoding.....	32
4.	RESULTS AND DISCUSSIONS .....	35
4.1	Influential Factors .....	35
4.1.1	Market Condition: -.....	36
4.1.2	Project Duration: -.....	36
4.1.3	Availability of Local Materials and Skilled Labors: - .....	37
4.1.4	Management and Safety Conditions: -.....	39
4.1.5	Location of Project: - .....	41
4.1.6	External Social Condition and Risk Conditions: - .....	42
4.1.7	Types and Complexity of Building: -.....	44
4.1.8	Contractor: - .....	45
4.2	Developing ANN Model.....	46
4.1.1	Data preprocessing and clearing: -.....	47
4.1.2	Architecture Design: - .....	47
4.1.3	Input and Output Layers: - .....	48
4.1.4	Number of Hidden Layers and Neurons: - .....	49
4.1.5	Activation Functions: -.....	49
4.1.6	Loss Function: -.....	50
4.1.7	Batch Size & Number of Epochs: -.....	50
4.1.8	Learning Rate: -.....	51
4.1.9	Regularization Parameter: -.....	51
4.1.10	Training and Validation .....	52

4.2 Results.....	53
4.2.1 Model Evaluation: - .....	53
4.2.1.1 Mean Absolute Error (MAE): - .....	53
4.2.1.2 Mean Squared Error (MSE): - .....	53
4.2.1.3 R-Squared (Coefficient of Determination): - .....	54
4.2.2 Analysis of Residuals: - .....	55
4.2.2.1 Validation Performance: - .....	55
4.2.2.2 Training State: - .....	56
4.2.2.3 Error Histogram: - .....	57
4.2.2.4 Comparison with Actual Cost: - .....	58
5. CONCLUSION AND RECOMMENDATION.....	63
5.1 Conclusion .....	63
5.2 Recommendations.....	65
6. References.....	66
APPENDECIS .....	
Appendix 1: Survey on Influential Factor on Construction Cost Estimation .....	
Appendix 3B: MATLAB code for Best Hidden Layer Size .....	
Appendix 3C MATLAB code for Activation Function.....	
Appendix 3E MATLAB code for Batch Size and Epochs.....	
Appendix 3F MATLAB code for Learning Rate .....	
Appendix 3G MATLAB code for Regularization parameter .....	

## LIST OF TABLES

Table 1 Application ANN in different areas of Construction management .....	16
Table 2 Number of samples used in different papers .....	28
Table 3 Summary of the Research methodology .....	28
Table 4 Distribution of questionnaire according to Company type.....	29
Table 5 Distribution of questionnaire according to job title.....	30
Table 6 Distribution of the questionnaire based on the number of years of experience.	30
Table 7 Independent variables (selected influential factors) .....	31
Table 8 Excluded variables.....	31
Table 9 Coding.....	32
Table 10 The most influential factors that were adopted in this study .....	35
Table 11 Market condition.....	36
Table 12 Project Duration.....	37
Table 13 Local Material, Skilled Labor, Social, Risk Conditions and Complexity .....	38
Table 14 Management and Safety conditions.....	40
Table 15 Project location .....	42
Table 16 Types of building.....	45
Table 17 Grade or Capacity of Contractor.....	46
Table 18 Results of neural network model at testing phase .....	59

## LIST OF FIGURES

Figure 1 Schematic diagram of processing element (Mohaghegh, 2000) .....	8
Figure 2 structure of ANN, which consists of three basic layers, input, hidden and output layer .....	8
Figure 3 Shapes of some commonly used activation functions (Junxi Feng, 2019) .....	10
Figure 4 Schematic representation of feedforward- and backpropagation .....	11
Figure 5 Research Design .....	25
Figure 6 Population frame .....	27
Figure 7 Original and cleaned data .....	47
Figure 8 Network Architecture.....	48
Figure 9 Selected network architecture .....	49
Figure 10 Training algorism for robust generalization .....	51
Figure 11 Validation and test data .....	52
Figure 12 Higher R-Squared values.....	54
Figure 13 Best Validation performance graph.....	55
Figure 14 Training state graph.....	56
Figure 15 Error histogram .....	58

## ACRONYMS

CCIs;Construction Cost Indicators .....	17
CEM, Cost Estimation Model .....	20
DNN;Deep Neural Networks.....	16
GA-BPNN;Genetic Algorithm Back Propagation Neural Network.....	17
GUI;Graphical User Interface.....	23
MAPE;Mean Absolute Percentage Error .....	22
MLP;Multilayer Perceptron.....	22
OH;Overhead .....	21
PERT;Program Evaluation and Review Technique .....	5
PPB;Project Portfolio Benefit.....	17
RBF;Radial Base Function .....	22
ReLU;Rectified Linear Unit .....	8
RMSE;Root Mean Square Error .....	18
SHAP;Shapley Additive Explanations .....	16
tanh;Hyperbolic Tangent .....	8

## **Abstract**

*The construction industry is a critical component of infrastructure development, and the ability to accurately and reliably estimate project costs is essential for the successful execution of public projects. This research delves into the application of Artificial Neural Networks (ANN) as an innovative approach to cost estimation for public projects within the Oromia Region. The study aims to address the limitations of traditional cost estimation methods by leveraging the adaptability and pattern recognition capabilities inherent in ANN technology. The comprehensive methodology employed in this research includes an in-depth literature review, surveys and interviews with industry professionals, and the collection of historical project data from public records, government agencies, and construction firms located within the Oromia Region. The collected data undergoes rigorous preprocessing before being used to develop, train, and validate the ANN model. The performance of the ANN model is then meticulously evaluated using well-established metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared ( $R^2$ ), with the results thoroughly analyzed to assess the efficacy of the ANN approach. Furthermore, the study explores the various factors that influence construction cost estimation and provides practical, implementation-focused recommendations for integrating ANN-based cost estimation within the context of public projects. The findings of this research contribute significantly to the advancement of construction cost estimation practices, offering valuable insights into the potential of ANN for enhancing accuracy and reliability in the specific domain of public projects within the Oromia Region. The outcomes of this study are expected to benefit all parties involved in construction projects, including owners, contractors, consultants, and others, by enabling them to obtain cost estimates at the early stages of projects with limited available information, while maintaining high accuracy and adhering to feasible timelines. Additionally, the study aims to guide industry professionals and researchers in the adoption of innovative technologies, such as ANN, to achieve more effective and reliable cost estimation in construction projects.*

*Keywords: Cost Prediction, Artificial Neural Networks, Public Projects, Model*

---

# 1. INTRODUCTION

## 1.1 Background

The construction industry serves as a cornerstone for economic development and societal progress, contributing significantly to infrastructure enhancement. Central to the success of construction projects is the accurate estimation of costs, a critical aspect that influences project viability, financial planning, and overall project management. Accurate cost estimation plays a crucial role in enabling stakeholders and decision-makers to conduct comprehensive feasibility studies, determine appropriate financial scales during the bidding stage, and effectively monitor cash flows throughout the construction phase (Elmou salami, H. H., 2020); (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019); (Wang, 2022).

Traditional methods of cost estimation in the construction industry often relied on historical data and expert judgment. However, these approaches faced inherent limitations in adapting to the dynamic and complex nature of contemporary construction projects. Public projects, in particular, involved various stakeholders, intricate regulatory frameworks, and diverse geographical and environmental considerations, adding layers of complexity to accurate cost estimation.

The advent of Artificial Neural Networks (ANN) presented a promising avenue for revolutionizing cost estimation in construction. Inspired by the human brain's neural structure, ANNs exhibited exceptional pattern recognition capabilities and adaptability to nonlinear relationships within data. This research aimed to leverage the power of ANN to enhance the accuracy and reliability of cost estimation specifically for public projects in the Oromia Region.

The Oromia Region, characterized by its diversity and expansive geographic scope, hosted a wide array of public projects ranging from infrastructure development to community facilities. Accurate cost estimation in this context was essential not only for fiscal responsibility but also for ensuring that the benefits of these projects reached the communities they served.

This study responded to the limitations of existing cost estimation methods by exploring the potential of ANN in providing more accurate, adaptive and context-specific cost estimates for public projects in the Oromia Region. By delving into this innovative approach, the research endeavored to contribute valuable insights that could significantly impact the efficiency, transparency, and success of public projects, ultimately fostering sustainable development and improving the quality of life for the residents of the Oromia.

---

## 1.2 Objectives

### 1.2.1 General Objective: -

The objective of this study was to develop a highly accurate Artificial Neural Network (ANN) model for early cost estimation of public building construction projects in the Oromia Region.

### 1.2.2 Specific Objective: -

The specific objective of the study was

- ✚ To Identify Factors Influencing ANN Model Performance
- ✚ To design best NN Architecture, Train and Validate the ANN Model
- ✚ To Evaluate the Performance of the ANN Model

## 1.3 Problem Statement

The construction industry in the Oromia Region faced a critical challenge in accurately estimating costs for public projects, a fundamental aspect that directly impacted project success and financial viability. Traditional cost estimation methods, relying on historical data and expert judgment, exhibited limitations in adapting to the dynamic and complex nature of public construction initiatives in this region. The inadequacies in these methods contributed to cost overruns, inefficient resource allocation, and project delays, jeopardizing the timely completion of vital infrastructure and community development projects. Additionally, the underutilization of advanced technologies, particularly Artificial Neural Networks (ANN), in construction cost estimation further exacerbated the problem, hindering the industry's potential for innovation and improved accuracy. This study addressed these pressing issues by investigating the most influential factors of cost estimation in the application of ANN models to enhance the precision and adaptability of cost estimation for public projects in the Oromia Region. Its aim was to provide a transformative solution to the industry's current challenges.

## 1.4 Scope

This study focused on public construction projects within the geographical bounds of the Oromia Region, delving into the unique challenges and opportunities presented by the diverse nature of these projects. The scope encompassed various types of public construction initiatives, including infrastructure development and community facilities, with a temporal focus on historical project data from 2009 E.C up to the present date. The primary objective was to apply and assess the effectiveness of Artificial Neural Networks (ANN) for improving the accuracy of cost estimation in this specific context. The study

---

involved the development, training, and evaluation of an ANN model tailored to the intricacies of public projects in the Oromia Region, providing a localized perspective on construction cost estimation. Additionally, the research aimed to identify factors influencing the cost estimation and offer practical recommendations for the implementation of ANN in the construction industry within this geographical framework. Ethical considerations regarding data privacy and confidentiality were integral to the study's scope, ensuring responsible research practices throughout the investigation.

### 1.5 Significance of the Study

This study held significant importance as it addressed the critical need for accurate construction cost estimation in public projects within the Oromia Region. By applying Artificial Neural Networks (ANN) to this context, the research sought to revolutionize traditional methods, offering a transformative solution to the industry's longstanding challenges. The outcomes were anticipated to contribute to improved project management, efficient resource allocation, and timely project completion, thereby fostering sustainable development. The study's findings had the potential to elevate the competitiveness of the construction sector in the region by introducing innovative technologies, paving the way for increased accuracy, adaptability, and reliability in cost projections. Moreover, the provision of practical implementation recommendations and ethical data handling protocols ensured that the study not only advanced knowledge in construction cost estimation but also served as a valuable guide for industry professionals, promoting responsible and technologically advanced practices in the field.

### 1.5 Delimitation and Limitation

When using Artificial Neural Networks (ANNs) for construction cost prediction, limitations and delimitations referred to the constraints and boundaries within which the predictive model operated. ANNs were often considered black-box models, making it challenging to interpret how the model arrived at its predictions. This lack of transparency limited the understanding of the factors driving cost predictions. ANNs struggled to generalize well to unseen data if the training data did not sufficiently represent the diversity of situations or contexts in which cost predictions were needed.

The predictions made by the ANN were limited to the variables included in the training dataset. Variables not considered during training were not accurately predicted, and the predictions were delimited by the time period and geographical region covered by the training data. Predictions were less reliable for projects outside these boundaries.

---

## 2. LITERATURE REVIEW

### 2.1 Traditional Cost Estimation Methods

Traditional cost estimation methods in construction have been widely used for many years and are often based on historical data, expert judgment, and established rules of thumb. While these methods have their merits, they also come with inherent limitations. Some of the commonly employed traditional cost estimation methods include:

(Len Holm, 2021) and (Jinisha, R., & Jothi, R., 2019) estimate types often correlate with project design phases and are labeled as Conceptual design budget estimate, Schematic design budget estimate, Design development estimate and Construction document estimate (Nicholas, J. M., & Steyn, H. , 2017), (Marker, 2017) Expert judgment consists of cost estimates made by cost experts. In the estimating process, the experts use their cost knowledge of prior similar projects to draw comparisons between past projects and the new project to make and adjust the new cost estimate. Usually, expert judgments are only made in the initiative phase and for projects that are poorly defined or if there are no prior similar projects to compare the new project with.

Analogous estimating is known as a comparative estimating method. The cost estimate is developed by reviewing cost from previous similar projects and considering the differences with similar projects. An analogous estimate is a relatively fast estimate (Nicholas, J. M., & Steyn, H. , 2017); (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019). If there are minor deviations in the cost data used from previous projects, the estimate can be quite accurate. However, it is very difficult to select the appropriate projects to compare the new project with as this method requires extant information about prior projects. Therefore, companies gather cost data and store it into databases that classify costs according to type of project, work package, element and so on. Parametric estimating is a statistical estimating technique that uses multiple regression analysis to relate the construction costs to a model of prediction parameters (Ramos, 2017). A parametric estimate is derived from an empirical, mathematical and statistical relationship between historical project cost and cost parameters of the project (U.S. Government Accountability Office, 2009; (Nicholas, J. M., & Steyn, H. , 2017). It is important that the cost parameters used in a parametric estimate are the cost drivers of the project (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019). Mostly these cost parameters are physical features of a project such as area, volume, weight, or capacity (Nicholas, J. M., & Steyn, H. , 2017). To determine cost drivers and relevant CERs, a parametric estimate requires historical data (Ramos, 2017). Parametric

---

estimates can be conducted quickly and they can be easily replicated (NASA Executive Cost Analysis Steering Group, 2015), as such they are useful early in the design phase where estimates are needed quickly (Nicholas, J. M., & Steyn, H. , 2017). Because parametric estimates use data of actual observations, the reliance upon opinion such as expert judgement is eliminated. In order to assure that parametric estimates stay in line with actual cost relationships between project attributes and costs, the CERS must be continually revised. This is a disadvantage of the parametric estimating technique (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019).

Detailed estimating, also known as cost engineering, is estimating the costs of a project from a detailed determination of the costs of each cost category, work package and element of the project, including the duration of the work packages (Ramos, 2017); NASA Executive Cost Analysis Steering Group, 2015). It is a bottom-up estimating method (Nicholas, J. M., & Steyn, H. , 2017); (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019) that provides the most accurate estimates of all estimating techniques (Nicholas, J. M., & Steyn, H. , 2017), but the method is very time consuming and is associated with high costs (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019). Detailed estimations are mostly made later on in the design phase because the estimating technique requires detailed project and design information which is often not available in the early phases of a project (Nicholas, J. M., & Steyn, H. , 2017). Estimation in detail has the advantage that there is the ability to determine exactly what is included in the estimate and check if nothing was overlooked (U.S. Government Accountability Office, 2009). Another advantage is that the detailed estimation method provides insight into the cost drivers of the project. The method has the disadvantage that for every new project a new detailed estimate must be made. This is time consuming and very costly. However, to speed up the estimating process, estimates of certain activities that reoccur in new projects can be taken from previous projects but must be integrated into the context of the new estimate (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019).

The Program Evaluation and Review Technique (PERT) is often used for complex projects. It incorporates a range of potential costs, addressing uncertainty. Useful for risk management but may not capture all project uncertainties. Relies on subjective input for optimistic, most likely, and pessimistic estimates. In probabilistic estimating, probability distributions are used as input for the cost estimate of one or more cost parameters (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019). This

---

estimating technique focuses more on the risks and an uncertainty involved in a project and tries to quantify the cost variability of the project. A probabilistic estimate gives insight into the change of exceeding the cost of a particular aspect in the range of possible costs, how much costs could overrun, and how uncertainties drive costs (Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019).

Finally, Traditional cost estimation methods in construction have been widely utilized for many years, typically relying on historical data, expert judgment, and established rules of thumb. These methods encompass various approaches, including conceptual design budget estimates, schematic design budget estimates, design development estimates, and construction document estimates, as highlighted by Len Holm (2021) and Jinisha and Jothi (2019). Expert judgment, based on cost experts' insights from prior similar projects, provides another avenue for estimation, especially for poorly defined projects or those lacking comparable precedents, as noted by Nicholas and Steyn (2017) and Marker (2017). Analogous estimating, a comparative method, draws on cost data from similar past projects, offering relatively quick estimates but requiring extensive historical information, as outlined by Nicholas and Steyn (2017) and Matel et al. (2019). Parametric estimating, a statistical technique, establishes cost predictions based on empirical relationships between project attributes and costs, allowing for rapid estimation but necessitating continuous refinement of cost estimation relationships, according to Ramos (2017) and NASA Executive Cost Analysis Steering Group (2015). Detailed estimating, or cost engineering, offers the most accurate estimates by meticulously determining costs for each project category, work package, and element, though it is time-consuming and costly, as highlighted by Nicholas and Steyn (2017) and Matel et al. (2019). PERT and probabilistic estimating address uncertainties in complex projects by incorporating a range of potential costs or probability distributions, respectively, aiding in risk management but relying on subjective input and requiring detailed data analysis, as discussed by Matel et al. (2019). While each traditional method has its strengths and weaknesses, their combined use is common to enhance overall estimation accuracy, contingent upon project characteristics, data availability, and project development stage.

## 2.2 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) is computational models inspired by the structure and functioning of the human brain. They fall under the broader category of machine learning and are designed to recognize patterns, learn from data, and make predictions or decisions.

---

ANNs consist of interconnected nodes, often organized into layers, each layer serving a specific purpose. The key components of an ANN include:

(Shehatto, 2013) & (LeCun, Y., Bengio, Y., & Hinton, G., 2015) states that the term "neural" was adopted in the context of artificial neural networks due to historical factors, as early researchers predominantly had backgrounds in biology or psychology rather than engineering or computer science.

According to (Haykin, 1999) the first model of an artificial neuron was proposed by Warren McCulloch and Walter Pitts in 1943. The model was based on the structure and function of a biological neuron. In the 1980s, the backpropagation algorithm was introduced, which made it possible to train ANN with multiple layers. This breakthrough allowed for the development of deep neural networks, which are now widely used in many applications (LeCun, Y., Bengio, Y., & Hinton, G., 2015).

The use of ANN in cost estimation began in the early 1990s and has since gained popularity due to their ability to handle non-linear problems and learn from examples (Karim, A., & Bubshait, A., 2013). Today, ANN continues to evolve and advance, with new architectures, algorithms, and applications being developed constantly (LeCun, Y., Bengio, Y., & Hinton, G., 2015).

Artificial neural networks, commonly known as ANN, are computational models that draw inspiration from the functioning of the nervous system, specifically the behavior of neurons (Alcineide, P., Gean, S., Luiz, M. F. M., Felipe, C. A., & Débora, de G. S. , 2021). ANN can be described as mathematical structures and their corresponding implementations, which encompass both hardware and software, and are designed to mimic and derive insights from the observed behavior of natural nervous systems (Juszczuk, M., Leśniak, A., & Zima, K., 2018).

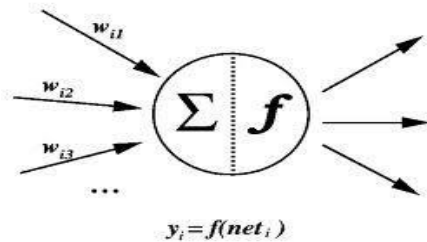
(Najafi, A., & Tiong, R., 2015) , Describes Artificial Neural Network as a type of computer processor that consists of many interconnected processing units (neurons). These neurons can store and recall experiential knowledge through a complex system of weighted connections. As a result, ANNs are capable of processing information in a manner that is similar to the human brain and can learn and generalize from past experiences to make more accurate predictions.

#### 2.2.1 Nodes (Neurons): -

Nodes, also known as neurons, are the basic computational units in an ANN. Each node receives input, processes it using a mathematical function (activation function), and produces an output. Neurons are organized into layers: an input layer, one or more hidden

layers, and an output layer.

Figure 1 Schematic diagram of processing<sup>1</sup> element (Mohagheh, 2000)



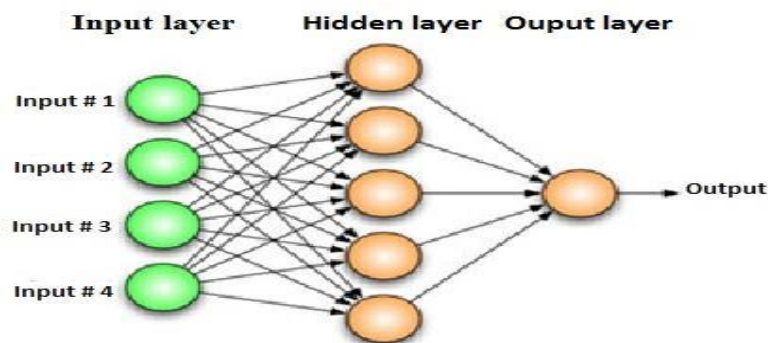
### 2.2.2 Connections (Weights): -

Connections between nodes have associated weights that modulate the input signal. These weights are adjusted during the training process, allowing the network to learn and adapt to patterns in the data. The strength of the connection (weight) determines the impact of a particular input on the node's output. These weights are responsible for determining how much influence each input has on the output produced by a neuron. (Chakraverty, S., Deepti, ., Sahoo, M., Nisha, ., & Mahato, R. , 2019).

### 2.2.3 Layers: -

ANNs typically consist of three types of layers: Input Layer (Receives the initial input data), Hidden Layers: (Intermediate layers between the input and output layers, where complex patterns are learned) and Output Layer:(Produces the final output or prediction).

Figure 2 Structure of ANN, which consists of three basic layers, input, hidden and output layer



### 2.2.4 Activation Function: -

The activation function of a node determines its output based on the weighted sum of inputs. Common activation functions include the sigmoid, hyperbolic tangent (tanh), and rectified linear unit (ReLU). The choice of activation function influences the network's ability to capture non-linear relationships in the data. The activation function introduces non-linearity to the network, enabling it to learn and approximate complex relationships in the data (Aggarwal, 2018). Different types of activation functions, including sigmoid, tanh, ReLU, and linear, are used to achieve this mapping of non-linearity (Wang, 2022).

Activation functions are nonlinear transformations of the linear combination of the inputs.

---

The activation function in a neural network is one of the most critical aspects in determining whether or not a neuron was activated and passed to the next layer. This simply implies that it will decide whether or not the neuron's input to the network is significant or not the process of making prediction. The activation function utilized has a substantial impact on the neural network's capabilities and performance, and multiple activation functions may be used in different parts of the model. Although networks are meant to employ the same activation function for all nodes in a layer, technically, the activation function is applied within or after the internal processing of each node in the network. Activation functions are often differentiable, which means that the first-order derivative can be determined for a given input value. This is essential because neural networks are often trained using the backpropagation of error procedure, which requires the derivative of prediction error to update the model's weights. These functions are not directly observed but learned during the network training (James, 2021). There are a few activation functions commonly used in neural networks, depending on the problem at hand. For examples,

**Sigmoid activation:** This is commonly used in logistic regression to convert a linear function into probabilities between zero and one. The larger the input (more positive), the closer the output to 1, while the smaller the input (more negative), the closer the output to 0.

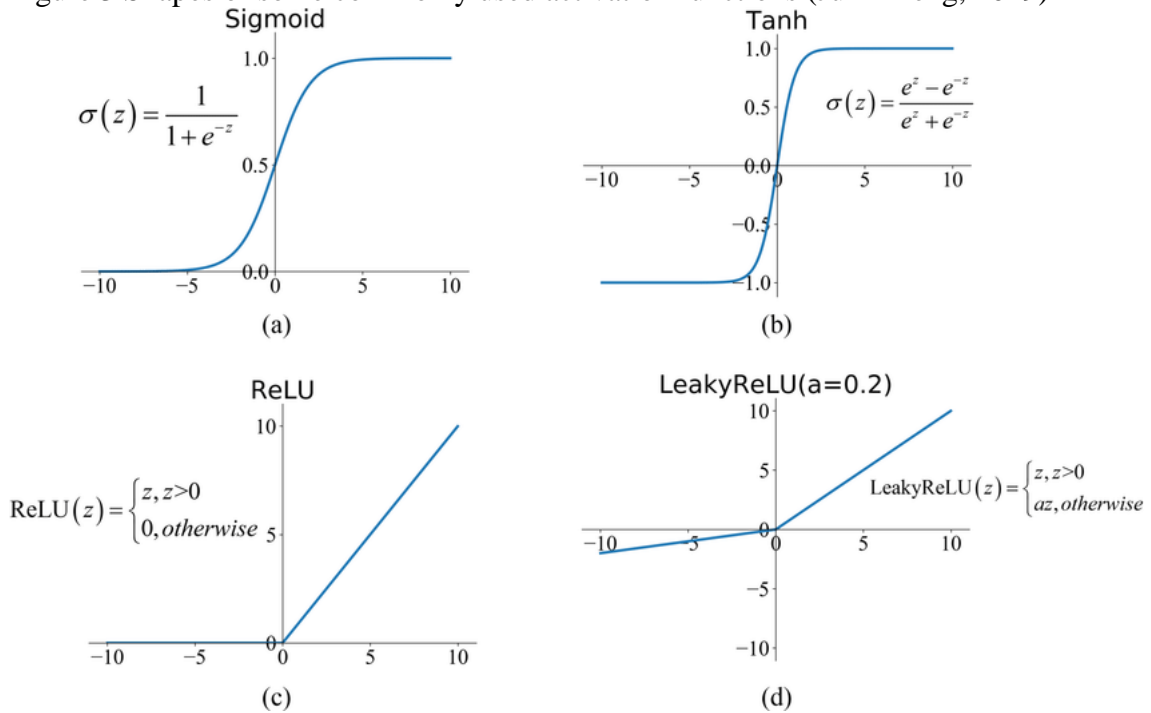
**ReLU (Rectified linear unit) activation:** This is the most preferred choice in modern neural networks because of its ability to compute and store more efficiently than the sigmoid activation function. It is less vulnerable to vanishing gradients, which prevents deep models from being trained, but it might suffer from other issues such as saturated or "dead" units.

**Tanh activation:** The hyperbolic tangent activation function is another name for the Tanh function. It is quite similar to and even has the same *S - shape* as the sigmoid activation function. The function accepts any real value as input and returns values ranging from -1 to 1. The larger the input (more positive), the closer the output to 1, while the smaller the input (more negative), the closer the output to -1.

**SoftMax activation:** The SoftMax function returns a vector of values that add to 1 and can be regarded as class membership probability. It is similar to the *argmax* function, which returns 0 for all options and 1 for the selected option. SoftMax is a "softer" form of *argmax* that allows a winner-take-all function to provide a probability-like result. As a result, the function's input is a vector of real numbers, and the output is a vector of the

same length with values that add to 1, similar to probabilities.

Figure 3 Shapes of some commonly used activation functions (Junxi Feng, 2019)

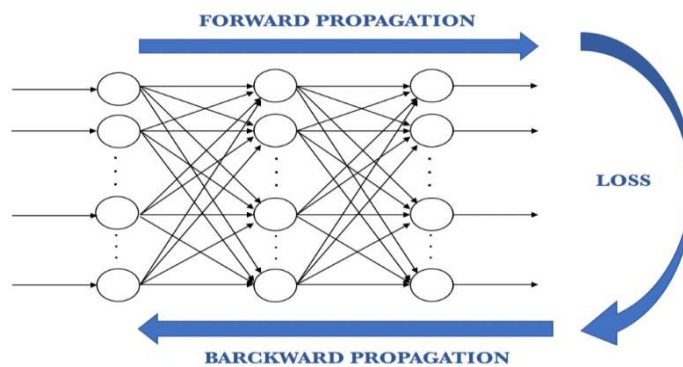


ReLU, Nowadays, the Rectified Linear Unit (ReLU) is the most used activation function (Sharma, 2017), (Walia, 2017) . ReLU is used in almost all the convolutional NNs or deep learning. The range of ReLU is between 0 and infinity. The function has the disadvantage that all the negative values become zero, thereby immediately decreasing the ability of the NN model to fit or train the data properly. Therefore, the ReLU activation function is most often used in cases where the data only concerns positive values.

### 2.2.5 Feed forward and Back propagation: -

During the training process, ANNs use a feed forward approach to make predictions based on the input data. The predicted output is then compared to the actual output, and the network adjusts its weights using back propagation. This iterative process continues until the network achieves satisfactory performance. (Tijanić, 2020) Outlines that backpropagation, the squared error between the output values and the desired values is minimized using gradient descent. The error signals obtained from this process are utilized to compute the weight updates (Tijanić, 2020). The Back propagation algorithm iteratively reduces the error between the model's output and the target output. Backpropagation is based on the error correction learning rule and consists of two phases: forward propagation and backpropagation (Alcineide, P., Gean, S., Luiz, M. F. M., Felipe, C. A., & Débora, de G. S., 2021). An epoch refers to a single pass through the entire training dataset during the learning process.

Figure 4 Schematic representation of feedforward- and backpropagation



#### 2.2.6 Loss Function: -

The loss function measures the difference between the predicted output and the actual output. The goal during training is to minimize this loss, indicating that the network's predictions align closely with the true values.

The choice of the loss function must match the specific predictive modelling problem the model is concerned with, such as classification or regression. Also, the output layer of the NN model must be configured appropriately to be compatible with the loss function. According to Brownlee (Brownlee, 2020) there are three types of loss functions:

#### 2.2.7 Optimization Algorithms: -

Optimization algorithms, such as gradient descent, are employed to adjust the weights systematically and minimize the loss function. These algorithms determine how much to update the weights based on the calculated gradient of the loss function.

#### 2.2.8 Hyper parameters: -

ANNs have hyper parameters, such as the learning rate, the number of hidden layers, and the number of neurons in each layer, which need to be set before training. The selection of appropriate hyper parameters influences the network's performance.

### 2.3 Role of Artificial Neural Networks (ANN) in Construction Management

(Kaviya, 2019) used ANN for forecasting the compressive strength of high-performance concrete from 446 concrete data samples, using 326 data. The model had 8 Input Layers and 1 Hidden Layer and 30 nodes were used. The results presented that ANN was suitable compared to the multiple linear regression model, which can reduce the errors in the concrete industry in any safety issue.

(Singh, 2019) Feasibility of Artificial Neural Network in Civil Engineering, it is clear that ANNs have been successfully applied to many civil engineering areas like prediction, decision-making, risk analysis, resources optimization, classification and selections etc. Based on the results, it is that ANNs perform better than other conventional methods. In civil engineering many problems are very complex and not well understood. ANNs are

---

based on only input and output data by which model can be trained easily. ANNs can always be updated to obtain better output by showing new training examples as new data become available

(Elevado, K. J. T., Galupino, J. G. & Gallardo, R. S., 2018) discusses using flying ash as a substitute for Portland cement type 1 and replacing gravel as ground aggregates with ceramic waste tiles. Also, specimens were subjected to different curing days to test their improvement of 33strength. ANN was considered because a large range of data was available. The goal of this study was to provide an ANN model that would be able to forecast the concrete's compressive strength while using waste ceramic tiles as a replacement for coarse aggregates while varying the quantity of fly ash as a partial cement replacement. The model architecture consisted of three input variables, which represent the days of compressive strength 7 days, 28 days, and 56 days, and also two output variables, which are the percentage of ceramic and the percentage of fly ash. The ANN model was validated and tested to ensure accuracy.

(Alaloul, 2018) , develops an ANN model to assess the impact of coordination factors on the performance of construction projects. The most successful coordinating factors that influence the performance of construction projects were identified. These factors were simplified to five main factors, which are Resource management, planning and scheduling, value engineering, contract implementation, and documentation. Based on these factors, a questionnaire was designed and sent out of 610 questionnaires, and only 325 were received. All the previous were considered as a database to develop the ANN model. The model architecture consisted of three hidden layers and developed with feed-forward back propagation algorithms. The model was trained, validated, tested, and showed MSE 0.0231 and determination of correlation coefficient  $R^2 = 0.77, 0.75, \text{ and } 0.76$  for the cost, quality, and time respectively.

(Raghd, A., Hanafy, H. & Tarek, A., 2018) Developed a model based on ANN that supports construction companies to evaluate and predict the outcomes of their current projects. This is what improves the performance of the construction company and enhance it is the capability to compete in the local and international market. The proposed ANN model included public construction projects only. Twelve factors were identified from the previous literature review: project scope, external constraints, time urgency, interdependency between elements, project resources, project budget, project parties, resource availability, level of interface, project client, and several elements. After analyzing these factors, time urgency and external constraints had the highest impact while project

---

client and number of elements had the lowest impact. Data from 30 previous project projects developed in Egypt, the Gulf, and North Africa were collected. These collected projects included malls, hotels, airports, educational, offices, and airports. The ANN model was developed using IBM SPSS software. Twenty-five projects were used for training the 34 network using a back propagation algorithm, and five projects were kept for testing the model. The model predicted four projects correctly, which is equal to 80% accuracy.

(ShehuWaziri, 2017) Artificial Neural Networks has gained considerable application in construction engineering and management in recent time. Over 100 resources published in refereed journals and conference proceedings were screened and reviewed with the view to exploring the trend and new directions of the applications of different ANN algorithms. The study revealed successful applications of ANNs in cost prediction, optimization and scheduling, risk assessment, claims and dispute resolution outcomes and decision making. It was observed that ANN has been applied to problems that are difficult to solve with traditional mathematical and statistical methods. The integration of ANN with other soft computing methods like Genetic Algorithm, Fuzzy Logic, Ant Colony Optimization, Artificial Bee Colony and Particle Swarm Optimization were also explored which generally indicated better results when compared with conventional ANNs. The study provides comprehensive repute of ANN in construction engineering and management for application in different areas for improved accuracy and reliable predictions.

(Jose, 2017) Quality is the fitness for use, conformance to requirements, predictable degree of uniformity and dependability, at low cost and suited to market. Cost of quality is an essential element of the total cost of any construction project. The main objective of this paper is to develop a neural network model that will enable the construction firms to access cost of quality for any future building project. The different sequences of the model development was investigated. Moreover, the validity of the proposed model was evaluated using case study applications. The main 29 factors affecting the expected cost of quality were identified. A questionnaire survey was carried out which was conducted among 60 experts in the construction domain to determine the importance of these twenty-nine factors. By using Microsoft Excel, relative importance index of the factors is obtained and the factors with important index more than 70% was selected for a second stage questionnaire survey. Second stage questionnaire survey was conducted among construction experts from different construction industries and the result obtained was used as the input parameters of the proposed cost of quality model.

(Al-saadi, 2017) The aim of this study to predicted the duration of road projects in republic

---

of Iraq. Historical data was adopted for (99) projects for interval between 2000 to 2017 from Roads and Bridges Directorate (RBD). Artificial Neural Network (ANN) model used to estimate the duration using six variables (length of road, No. of lane, No. of intersection, volume of earth, type of pavement and furniture level). The methodology used in this study included two important parts, the first part, reviewing the literature of the subject (estimating the duration of the road projects), and the second part, used of a program neuroma v.4 to build the models of neural networks to estimate the duration of road project. The results showed strong correlation between actual duration and predict duration by (90.6%), minimizes testing error (3.2%) and training error (4.9%). The MAPE and Average Accuracy Percentage generated by ANN model were found to be (25.73 %) and (74.27%) respectively. Therefore, it can be concluded that ANNs model show very good agreement with the actual measurements

(P.S.Kulkarni, 2017), reviews application of ANNs in construction activities related to prediction of costs, risk and safety, tender bids, as well as labor and equipment productivity. The review suggests that the ANN's had been highly beneficial in correctly interpreting an inadequate input information. The review confirms the usefulness of ANNs in carrying out a variety of prediction, classification, optimization and modelling related tasks in areas of CM. ANN thus has significant benefits that make it a powerful tool for solving many problems in the field of CM

(Fatima, A., Prasad, B. S. R. K. & Sekhar, S. T., 2017), developed an ANN model using MATLAB software to minimize the disputes in the construction industry, which automatically decreases the 32 construction costs. The data collection to build the model was done through a survey sent to several construction companies with a rating scale from 1 to 5. The respondents must give a rating for each factor to identify the weight of the factors and their influence on the disputes. SPSS software was used to analyze and identify the collected 20 factors which were used as input variables for the ANN model.

(Nani, G., Mensah, I. & Kumi, T. A., 2017) created a model for predicting the duration of prefabricated steel Bridge construction in Ghana using ANN. Data were provided by the Department of Feeder Roads for eighteen bridge construction projects. The collected data had all the information that contained the quantity and duration of all work items per project. Another questionnaire was used to collect information concerning the pre-manufactured steel bridge components of each project chosen. The factors were first reduced to a lesser number, which is the formwork and the span of the bridge. These two factors were used as the input variables for the model to predict the output, which is the

---

duration of the project. The findings showed that the duration required to construct a bridge is closely linked to the formwork and the bridge span. The model developed provided precise results with  $R^2 = 0.998$  and accuracy of 95.95% on average during validation, which proves the efficiency of the ANN model in predicting the duration of bridge construction projects.

(Sharmila, S., Lekha, G. & Kaushik, S., 2016) developed an ANN model to predict the bearing capacity of the soil, which is considered a critical factor in the foundation's design. The input variables for the model were several factors such as depth of foundation, soil type, and unit weight of the soil. The ANN model was trained by a feed-forward backpropagation algorithm then tested. The model output showed accurate results on the bearing capacity value of the soil with high correlation and less time and cost than the traditional way.

(Aswed, 2016) developed an ANN model to estimate the productivity rate of the labors using thirty factors as input variables. These factors are site security, mortar type, site conditions, and the thickness of the wall, height of the wall, length of the wall, availability of materials, salaries, labor health, labor, age, labor experience, and labor number.

The data set was collected based on these factors from different project types. The model was trained and tested with an accuracy of 86.28%, which proves that the model can estimate the productivity rate of the labors in different types of projects such as residential, educational, and commercial through using these 30 factors as input for the model.

(Goa, 2016), Developed an ANN model for safety evaluation for construction projects. The model consisted of expert scoring as input variables and a class of security as a target variable. The model classified some construction companies as class 2, which was compatible with the actual situation of the construction companies. Research indicated that ANN technology has an excellent memory, and the role of the association and the digital properties of sample data are expressed. The ANN is convenient, simple, accurate, fair, and suitable to be used for the evaluation of safety for construction projects. However, this ability is influenced by the difference between samples and the capacity of memory.

Table 1 Application ANN in different areas of Construction management

S. No	Author	Application Area
1	(Kaviya, 2019)	ANN model forecasting the compressive strength of high-performance concrete
2	(Elevado, K. J. T., Galupino, J. G. & Gallardo, R. S., 2018)	ANN model that would be able to forecast the concrete's compressive strength while using waste ceramic tiles as a replacement
3	(Raghd, A., Hanafy, H. & Tarek, A., 2018)	ANN Model that supports construction companies to evaluate and predict the outcomes of their current projects.
4	(Alaloul, 2018)	ANN model to assess the impact of coordination factors on the performance of construction projects
5	(Al-saadi, 2017)	ANN predicted the duration of road projects
6	(Nani, G., Mensah, I. & Kumi, T. A., 2017)	ANN model for predicting the duration of prefabricated steel Bridge construction
7	(Fatima, A., Prasad, B. S. R. K. & Sekhar, S. T., 2017)	ANN model using MATLAB software to minimize the disputes in the construction industry
8	(Sharmila, S., Lekha, G. & Kaushik, S., 2016)	ANN model to predict the bearing capacity of the soil
9	(Goa, 2016)	ANN model for safety evaluation for construction projects
10	(Aswed, 2016)	ANN model to estimate the productivity rate of the labors

#### 2.4 Role of Artificial Neural Networks (ANN) in Construction Cost Prediction

(Wang, 2022), conducted their study on adopting the data on 98 public school projects in Hong Kong SAR. Their point of focus was to address certain limitations in construction cost estimation models. Existing models tend to focus solely on project characteristics and overlook the influence of external economic factors. The study aims to quantitatively explore the effects of economic factors on construction cost estimation by using Deep Neural Networks (DNN) as an estimator and Shapley Additive explanations (SHAP) as a model interpreter. The analysis utilized data set and included a comparison analysis with other popular machine learning models used in construction cost estimation

(Yuanyuan & Libiao, 2022), developed a genetic algorithm-back propagation neural network (GA-BPNN) model to quantitatively predict project portfolio benefit (PPB). The model developed in the study showed higher prediction accuracy than conventional BPNNs.

(Kim, Choi, Shahandashti, & Ryu, 2022), proposed a hybrid ARIMA-ANN model to

---

predict construction costs and city-level construction cost indicators (CCIs), which could provide more accurate predictions than ARIMA or ANN.

(Lee, J.G.; Lee, H.S.; Park, M.; Seo, J.O. , 2021) Found that the ensemble model can be utilized in ANN-based conceptual cost estimation models to improve prediction accuracy and stability when the amount of project data is not large.

According to (Elmou salami, H. H., 2020) the first model of ANNs were proposed in 1943 by Warren McCulloch. It was not until 1982 when Hopfield interconnected these neurons to construct a network that gave rise to modern artificial neural networks (ANNs). (Shehatto, 2013), Outlines the first application in construction was in early 1980's. According to (Heravi, 2015), ANN application has been successful in construction engineering and management. They are used to estimate productivity for use in excavation hauling operations, concrete formwork elements, pouring concrete, and concrete finishing tasks. Heravi, further outlines that ANN facilitates the generation of precise labor production rates (labor/unit) for various industrial construction activities.

(Khalaf, 2020), Applied PSO to estimate the initial construction cost and construction period of 60 construction works. What was inferred in this study is that PSO was well performed with high accurate results and faced with parameters with extensive variability. Strength of this model is that it is more reliable than projects based on existing projects and based on judgments and experimental cases. This paper investigated a model with a wider range of parameters and tried to apply it to eco-friendly buildings.

(Jiang Q. , 2019), Studied the application of ANN in cost estimation of construction projects, and as a result of comparing the results with the RBFNN method, ANN showed excellent performance. In addition, research was conducted in a way that optimizes model accuracy and applies it to other types of projects to use different methods for cost factor selection. Finally, the study was conducted in a way that optimizes model accuracy, applies it to different types of projects, and uses different methods for cost factor screening.

(Chandanshive, 2019), studied the capabilities of multi-layered feed anterior neural network working model through a back propagation learning algorithm for estimating the cost of 78 building projects in India, along with testing the effectiveness of the Early Stopping and Bayesian normalization approaches. Along with this, over fitting errors were also studied. He also implemented fuzzy logic to predict the cost of building projects. Because their models are not dynamic in response to market prices, the need for a more agile model is felt. He also used the integration of BP neural networks and genetic algorithms to estimate the cost of residential buildings. The role of GA in the study was to prevent ANN from

---

falling to the local maximum and to improve ANN performance by increasing the convergence rate.

(Chandanshive, V. B., & Kam (Chandanshive, 2019) studied a dataset of 78 building projects located in regional area of Mumbai, India. The dataset was taken from the same regional area to reduce different parameters. The objective of the study is to increase the accuracy of cost estimation. The authors chose to focus on structural cost as they considered it the most influential design parameter. The structural cost of building was assigned as an input and the total structural skeleton cost was the ANN model output. The authors used training and testing models to improve the accuracy of the model and the study and avoid over fitting. The authors used multiple training sets and performance measures and succeeded in reaching a correlation coefficient (R) of almost 1 which indicates the perfect fit. They concluded that a trained neural network can successfully predict early-stage project costs.

In the civil engineering aspect, there is the utilization of ANN to forecast or assist in some Processes. For instance, (Abd, A. M., Jasim, N. A., & Naseef, F. S., 2019) used ANN to estimate the final cost of the Iraqi construction projects, from 501 data sets since 2005-2015. There were 25 Input Layers, and a Hidden Layer was created by using 2 Nodes obtained through several experiments. The result value obtained from this model was  $R^2 = 0.987$  which made ANN has been proved for its accuracy of the least Root Mean Square Error (RMSE) from the trial-and-error process.

(Juszczuk, 2019) Construction site overhead costs are key components of cost estimation in construction projects. The estimates are expected to be accurate, but there is a growing demand to shorten the time necessary to deliver cost estimates. The balancing (symmetry) between time of calculation and satisfaction of reliable estimation was the reason for developing a new model for cost estimation in construction. This paper reports some results from the authors' broad research on the modeling processes in engineering related to estimation of construction costs using artificial intelligence tools. The aim of this work was to develop a model capable of predicting a construction site cost index that would benefit from combining several artificial neural networks into an ensemble. Combining selected neural networks and forming the ensemble-based models compromised their strengths and weaknesses.

Traditional methods of estimating project costs do not attempt to assess the magnitude of the variation inherent in the estimate. As a result, there is a risk that decisions on strategy selection were based on a high degree of uncertainty. The traditional approach to cost

---

estimating is to derive a best estimate from knowledge of existing conditions based on current rates and prices in similar situations with adjustments to reflect anticipated differences in ground conditions, site accessibility, and other factors (KabirYaqubi, 2019). Several estimation methods are used in construction practice and the suitability of any particular method is usually dependent on the purpose it is used for, the amount of information available at the time of estimation, and the party using it. Despite the reliance of clients and contractors on available cost estimation and forecasting methods, the actual final costs of construction projects still considerably deviate from their original estimates (Juszczuk, 2019).

(KabirYaqubi, 2019)The importance of decision-making in cost estimation for the building design process signifies the requirement for both designers and project managers. This article examines the usefulness of neural network methodology to address costs assessment at early stages of building design. The data is collected from 10 educational projects constructed in India. An automatic cost estimation GUI is prepared in MATLAB simulator to calculate the total cost of a construction project. Artificial intelligence is used to train the system on the basis of the data collected from 10 different projects and cost has been analyzed with higher accuracy. From the experiment it is observed that the use of artificial neural network helps the project participants to analyze total cost in a very small time.

When it comes to the problem of construction site overhead costs, it is worth noting the quite new research from Poland from (M. Juszczuk and A. Leśniak,, 2019)for a few reasons. (The authors claim that the “Construction site overhead costs are key components of cost estimation in construction projects. (The estimates are expected to be accurate, but there is a growing demand to shorten the time necessary to deliver cost estimates.” After considering and then combining several types of neural networks, in order to select the members of the ensemble, the authors developed three models intending to predict a construction site overhead cost index. It was proved that proposed models offer better cost prediction than those based on single neural networks (M. Juszczuk and A. Leśniak,, 2019).

(Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H., 2019) stated that the ANN method was inspired by the process in which the human brain works. (Polat, 2012) said that ANN finds the correlation weight of the hidden layer in the relationship between the input layer and the output layer and predicts the result value through the calculated weight. ANN algorithms are of great help in solving problems involving variables with dynamic properties. Through the learning process, the pattern of variables

---

can be found, and this can also be confirmed in the form of a learning curve.

(Pal, 2018) Cost estimating is one of the most important and challenging activities during project planning, which occurs at the early stages of a project life where limited information is available and many unknown factors affecting the project costs. ANN is a new advent that is used in cost estimation, which is able to lucubrate from experience and examples and deal with non-linear problems. It can perform tasks involving deficient data sets, fuzzy or insufficient information and for highly complex problems. The objective of this study is to review cost estimation models which used Artificial Neural Network (ANN) tool and to suggest the most effectual algorithm for cost prediction and the factors predominantly affecting the total construction costs of building projects. To build CEM, the most effective factors affecting cost in construction projects were identified based on a comprehensive survey among a collected sample of construction relevant model studies. The developed neural network model examines the data set into distinct cases classified on the basis of hidden layers and each of them containing the independent input neurons, hidden layers and a dependent output neuron. The results of the trained models indicated that neural network reasonably succeeded in estimating the Total construction cost of building projects at the planning stage itself. The average error of test dataset for the adapted model was largely acceptable and can perform as a good indicator regarding the ability of the proposed model to predict the total construction cost of any future construction project at an appreciated degree of accuracy. This paper gives a clear review of implementing the ANN tool in prediction of total cost of building construction projects and the relevant factors affecting it.

(Mistry, Jignesh, Mukeshchandra, 2018) Construction projects globally are considered to be multifaceted in nature. This complex work involves estimates of labor production rates during the planning as well as in the execution phase of the project. These estimates are habitually carried out by experienced personnel based on his/her experience, may sometimes not have the means to discern the controlling factor factors affecting the production rates. There are various trends in the soft computing techniques for identification of labor production rate in construction, and one being associated with Artificial Neural Networks (ANN). The current study emphasizes on critical literature review on factors likely to affect the productivity rates for concreting activities like reinforcement installation, formwork installation and concreting placement and also some of the other industrial activities.

(Barros, 2018) This paper focuses on the development of a more accurate estimation

---

technique for construction highway projects using ANN. This preliminary study showed the feasibility the tool applied to projects in Brazil and may be used by public agencies in the future. For the performance, authors have found results in accordance with this paper and the MAPE was 2.82% for the construction estimation, whilst found the MAPE equals to 5.84% for the construction material quantities estimation.

(Alshahethi, 2018) In this study investigation, two stages were carried out to achieve the objectives. Data collection and analysis, and developing ANN model have been done. Fifteen NNs models were built to predict the cost of the project by using neural network Tool Box software by MATLAB program Through five attributes were taken as predictor variables namely; collect data, preprocessing data, architecture the network, training the model, and testing the model using excel sheet and MATLAB. RMSE, MSE, MAPE, and R2 were calculated and compared for all 15 models to show the best model. It is observed the error from Bayesian Regularization- back propagation shown the best convergence towards minimum error compared to other algorithms. Among those models is 13- 17- 1 model as its percentage of error is 6% which is the least mean absolute percentage error and its coefficient of determination is 0.9998 for models that have already been tested. The findings clearly provide a good indicator for predicting the construction costs in the future with high degree of accuracy by using artificial neural network method.

(Lesniak, A. & Juszczak, M., 2018), Developed an ANN regression model that can predict the percentage of site OH cost for the contractors in Poland. Quantitative studies were carried out with regard to the proposed factors that affected the site OH cost in the construction industry. The following factors have been taken into account: the amount of work done by them and the amount done by subcontractors, work times, the complexity of the project, and difficulties in wintertime and localization conditions. The database collected to build the model was done through a questionnaire that was sent to 400 contractors in the polish construction market, and only 158 were received. These 158 projects were used as the database of the model. These data were encoded and used as an input variable. The ANN model was trained by doing fifty-four trials using different activation functions and a different number of neurons. The model, which has the lowest RMSE value, was selected and tested.

(M. Juszczak, A. Leńniak, and K. Zima., 2018) on a sample of 129 sports field construction projects that have been implemented in Poland in recent years. (The possibility and justification of the application of the NN for the assessment of total construction costs for sports' fields were explored. As one of the research tasks was to establish a set of cost

---

predictors, 7 predictors regarding the technical and functional characteristics were established. After that, the data were analyzed using two neural network types: multilayer perceptron networks (MLP) and radial base function networks (RBF). By applying Pearson's correlation coefficient between real and predicted values of construction costs and by using the root mean square error (RMSE) as the measure of prediction errors, satisfactory results were established for MLP networks. In the next step, the analysis for a group of 5 MLP networks was performed and the results were compared. As a comparison measure, Pearson's correlation coefficient was used between the actual and predicted construction cost and the root mean square error (RMSE) as the measure of the prediction errors. The accuracy of the estimation was tested using mean absolute percentage error (MAPE). The best results for all assessors were established for one network. In conclusion, this type of network can be recommended for estimating the sports field construction costs. (Abhijit N. Bhirud, 2017) This project is highlighting the study of Application of ANN for pre design cost estimation of building projects to investigate and overcome problems caused in estimating project cost at pre-design stage of building projects. As well as to develop & test a GUI model of cost estimating for building projects in the early design phase using MATLAB software. The accuracy performance of the adopted model for "Grand Total of the Project" recorded 86.11% where average percentage for all other parameters is 77.55% between the output obtained from model and the actual budget value of the project.

(Alshemosi, 2017), Reviewed the cost estimation of residential buildings using multi-factor linear regression, which eventually reached an accuracy of about 92%. This study recommended comparing the results with studies implementing neural network technology to identify differences. In fact, this study recommended using a cost estimation model instead of the existing method for construction projects.

(Ambrule, 2017) This project is highlighting the study of Application of Artificial neural network (ANN) for pre design cost estimation of building projects to investigate and overcome problems caused in estimating project cost at pre design stage of building projects. As well as to develop & test a Graphical User Interface (GUI) model of cost estimating for building projects in the early design phase using MATLAB software. In this study, a detailed analysis of data and results was presented and elaborated. The data used in this study was collected from 12 project estimates in the past three years. A data sheet was prepared and used to extract all useful information from each project. The accuracy performance of the adopted model for "Grand Total of the Project" recorded 86.11% where

---

average percentage for all other parameters is 77.55% between the output obtained from model and the actual budget value of the project.

(Magdum, 2017) Construction cost prediction is important for construction firms to compete and grow in the industry. Accurate construction cost prediction in the early stage of project is important for project feasibility studies and successful completion. There are many factors that affect the cost prediction. This paper presents construction cost prediction as multiple regression models with cost of six materials as independent variables. The objective of this paper is to develop neural networks and multilayer perceptron-based model for construction cost prediction. Different models of ANN and MLP are developed with varying hidden layer size and hidden nodes. Four artificial neural network models and twelve multilayer perceptron models are compared and NN give better results than statistical regression method. As compared to ANN, MLP works better on training dataset but fails on testing dataset. Five activation functions are tested to identify suitable function for the problem. 'Relu' transfer function gives better results than other transfer function.

(Lee, D., Kim, S. & Kim, S., 2016), developed a model to predict the cost and quantity of waste in the early construction phase in the early construction phase, implemented the hybrid model for measuring waste quantity and cost. The ant colony optimization method was used to optimize the ANN parameter collection. The model proposed can be utilized to enhance the waste management process in the construction industry and to reduce the cost overruns in the early construction phase.

(Shrestha, 2016), developed an ANN model to estimate the contingency cost of a road maintenance contract in Kenya. The model forecasts based on the historical change orders data and was validated using the change orders cost data from road maintenance contracts. The interfaces used to develop the model are the Microsoft Visual Studio Professional 2012, and the collected data were coded using the Visual C++ platform. To store the data and retrieval process, a database was designed using the Microsoft Access program. In this model, the factors that were used as input variables for the contingency cost estimation were the region name, site accessibility, weather condition, road condition, contract award cost, road surface type, and work category. The model proved it is accurate, and for every maintenance activity in the contract, a contingency cost based on changing order data was calculated. Also, this tool has an option to provide an adjusted contingency value. According to the cost weightage value of each maintenance activity, the overall contingency value for the project could be determined. Furthermore, this Model offers a modified contingency value, and the contingency cost of a project is estimated based on the

---

weight for each activity.

(Mirahadi, F. & Zayed, T., 2016), proposed to improve the accuracy of estimating the productivity for building operations with a hybrid model. The proposed paradigm model impacts the output of the productivity rate of both the quantitative, qualitative, and variables and the model's complex structure based on inherent data characteristics. A new model of the ANN driven fuzzy framework was created for this purpose<sup>31</sup> using Fuzzy reasoning. The dataset of 131 data items was divided into 117 samples for training the model fourteen samples Validation and testing. With respect to MSE, the proposed ANN model showed 83% of accuracy. This study allows researchers to build estimating models using the data features.

(Alqahtani, A., & Whyte, A., 2016), undertook a study involving a sample size of 20 building project to compare the performance of regression and artificial neural network in order to improve accuracy. The study is piloted over the previously compiled data from Al-Hajj (1991). The research concluded that further research is appropriate to determine additional cost drivers affecting cost estimation through detailed analysis of various project parameters such as location, inflation rate, and project design flexibility. The authors further highlighted regression models had no clearly defined model to help estimators select best design model, in contrary to ANN that seemed to perform easily across various designs. They recommended ANN can accept a larger number of independent variables than regression.

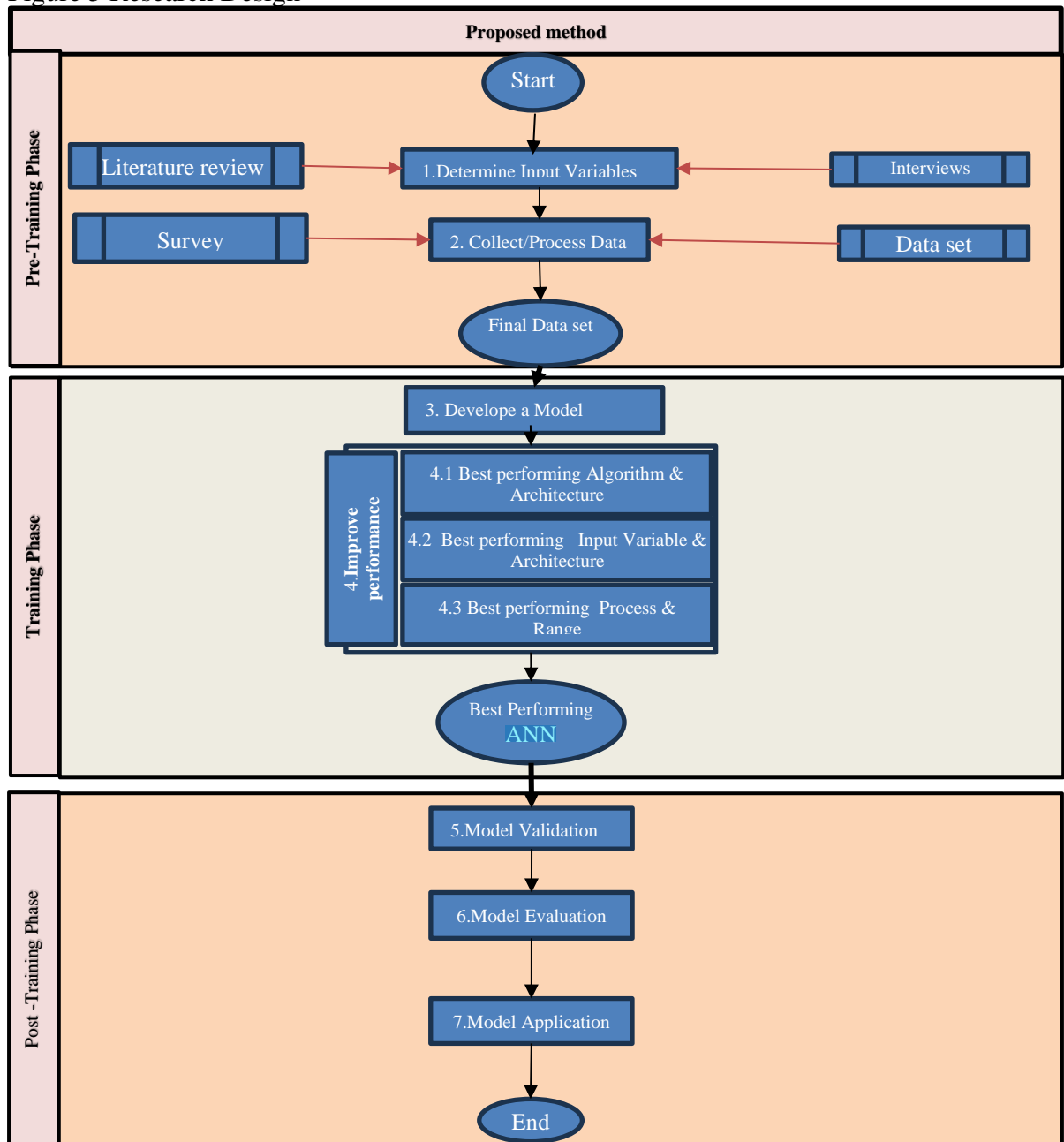
Although there are various methods for cost estimation, ANN is an accurate and widely known method for cost estimation. Previously, various studies have been conducted to solve the cost estimation problem, and these studies confirm that ANN's performance is better than that of a simple regression method and has a great effect on accuracy depending on the composition of neural networks in cost estimation (Fachrurrazi, Saiful, & Mubarak, 2016).

### 3. METHODOLOGY

#### 3.1 Research Design

The objective of the study was to identify crucial factors for early-stage cost estimation and subsequently develop an optimized Artificial Neural Network (ANN) model as an end result. The subsequent phase of the study concentrated on enhancing the accuracy of early-stage estimation in building projects through improvement of input factors that could be used in the implementation of the ANN model. The figure below depicted the designed procedure of every step of the research.

Figure 5 Research Design



---

## 3.2 Data Collection

The data collection process was a crucial phase in any research study, and for construction cost estimation using Artificial Neural Networks (ANN), it involved gathering relevant and representative data to train and validate the model. It consisted of two steps: - The first step involved preliminary data collection after an intensive literature review through interviews with experts using open and closed-ended questionnaires to identify the parameters for developing the cost prediction model. The second step of data collection entailed gathering final data to serve as a historical basis for parametric estimation for the selected parameters in step one.

### 3.2.1 Population of Study: -

Population constituents could be categorized along the lines of client types, the type of building projects, procurement type adopted, as well as project cost range. In the context of this study, the type of client referred to any of the following individual clients: local (Woreda) government, state (Regional) government, federal (National, international government organization). A project was described as one that could be procured through the traditional method of (Design, Bid, and Build). **Similarly, concerning project cost (initial budget cost), cost ranges were set to distinguish Small and Micro Enterprise (SME) project from Contractors from Grade-5 and above er, such as more than 50 million Birr.** Building types were also categorized along the lines of client type in population classification. Buildings could be residential, office, academic, recreational facilities, health facilities, or special buildings. Against this background, the population constituents for the purpose of this research were categorized along the lines of clients and building project types. The population frame for the study included public building projects used for public purposes, reinforced concrete, and in-situ concrete structures, completed within the last (1) to (7) years in Oromia Regional State. The initial cost (Bill of Quantities value) and As-built cost (final cost) were extracted from project documents of around 105 building projects and used as modeling parameters for the Neural network-based model. The targeted population for the surveys of input factors was chosen based on their profiles and not distributed randomly to reduce misleading results. The population consisted of organizations ranging from owners, consultants, and contractors to view results from all sides. The population itself was from various job positions but primarily focused on managerial positions with high experience for optimum results.

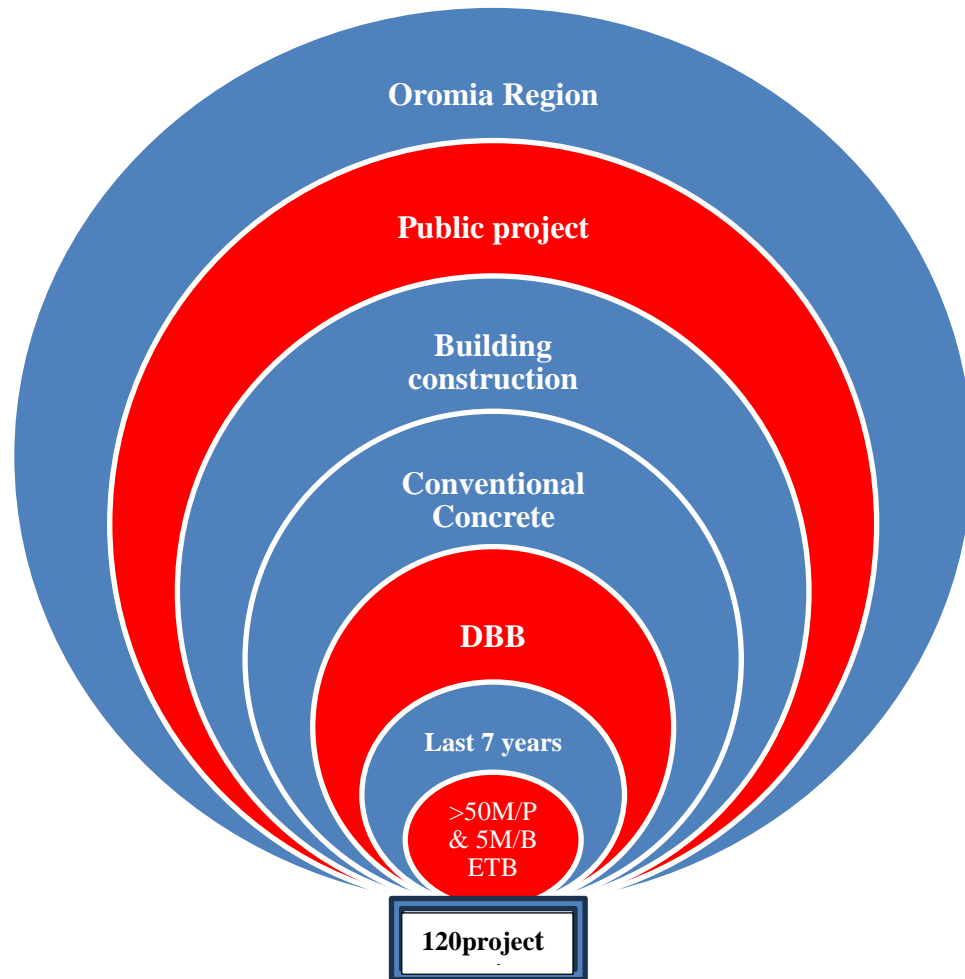


Figure 6 Population frame

### 3.2.2 Sample and Sampling Technique: -

Probability sampling technique was used in this research work while random sampling method was deployed in sample selection for model development. The samples for this work were categorized into two: The first category is sample for data training for designing a suitable network algorithm, while the second category is sample for model validation.

### 3.2.3 Sample Size: -

As the sample space (construction projects) is ample and unknown, the sample size can be computed using Equation bellow (Badawy, 2022).

$$\text{Sample Size} = \frac{Z^2 p(1-p)}{C^2}$$

Where Z is a value corresponding to a 95% confidence level and is equal to 1.96 and p represents the probability choice of 0.5. C is the confidence interval, which should be less than 0.2 (Badawy, 2022). Therefore, the minimum sample size for a confidence level of 95% is 44, about 120 samples were collected.

Table 2 Number of samples used in different papers

S. No	Author	Number of Sample
1	(Abdullah M. Alsugair, 2023)	135
2	(Abdulmohsen S. Almohsen, 2023)	94
3	(Yun D. P., 2023)	78
4	(Wang, 2022)	98
5	(Khalaf, 2020)	60
6	(Chandanshive, 2019)	78
7	(Abd, A. M., Jasim, N. A., & Naseef, F. S., 2019)	501
8	(Yaqubi, 2019)	10
9	(M. Juszczuk, A. Le´śniak, and K. Zima., 2018)	129
10	(Ambrule, 2017)	12

3.2.4 Research Location: -

This research work was conducted on several construction sites in the Oromia Region of the Federal Democratic Republic of Ethiopia. The data for the study was gathered from Contractors, Consultants and key governmental offices at the regional, zonal, Woreda, and municipal levels, such as Construction, Education, and Health Bureaus

Table 3 Summary of the Research methodology

S. No	Description	Proposed
1	Research Type	Mixed
2	Independent variables	Influential Factors affecting cost =10
3	Dependent variables	Total cost
4	Population	Oromia, Public, Building, completed in last 7 Years, Project Cost more than 50 million.
5	Sample	120 Project Data & 150 Professionals' interview
6	Data collection method	Survey Method
7	Tools/Instrument	MATLAB
8	Project Phase	Pre-construction
9	Delivery System	DBB

---

### 3.3 Study Population Characteristics

A total of one hundred fifty (150) questionnaires were distributed, with fifty being distributed online and one hundred distributed manually. From these, one hundred forty-three (143) questionnaires were returned, representing a response rate of 95.33% of the total number of questionnaires distributed. Among these, fifty-two (52) were from the online distribution and ninety-one (91) were from the manual distribution. These returned questionnaires underwent a cleaning process, during which some were omitted due to incomplete or inaccurate data.

The characteristics of the study population included the type of company, job title, and years of experience, and experience in using ANN for cost estimation in building projects.

#### 3.3.1 Types of company: -

Table 4 presented that (40.56%) of respondents were owners of the project, while (28.67%) were contractor companies, and (25.17%) were consulting firms. This indicated that all types of experiences were included, with the highest percentage being those who designed and estimated the initial cost of projects, and for the companies that implemented and assessed the actual cost of projects. However, contractors recorded the higher percentage of respondents due to the greater number of contractors compared to consulting companies in the Oromia Region.

Table 4 Distribution of questionnaire according to Company type

No.	Description	Percent%	No.
1	Owner (Municipality, Ministry,)	40.56%	58
2	Contractor (GC/BC-1 to GC/BC-5)	28.67%	41
3	Consulting Company	25.17%	36
4	Others	5.60%	8
Total Respondent		100%	143

#### 3.3.2 Job Title: -

According to Table 5, it was found that the survey encompassed various job levels of engineers, with (34.27%) being project managers, (23.08%) resident engineers, (18.18%) site engineers, and (13.6%) being owners and general managers. The highest proportions of job titles were project managers and resident engineers, indicating they possessed a comprehensive understanding of all project cost details.

Table 5 Distribution of questionnaire according to job title

No.	Description	Percent	Frequency
1	Owner /General Manager	13.99%	20
2	Construction manager	34.27%	49
3	Resident Engineer	23.08%	33
4	Site engineer	18.18%	26
5	Quantity Surveyor	3.50%	5
5	Others	6.99%	10
Total Respondent		100%	143

### 3.3.3 Years of experience in construction projects: -

Table 6 presented the work experience of participants, with (28.67%) of them having high experience of 12-15 years, while (27.97%) had 9-11 years, and only (10.49%) had less than 5 years of experience. This indicated that the greatest percentage of respondents possessed suitable experience in the construction field, enabling them to determine the critical factors affecting the cost of projects.

Table 6 Distribution of the questionnaire based on the number of years of experience.

No.	Description	Percent	Frequency
1	1 – 2 years	3.50%	5
2	3 – 5 years	6.99%	10
3	6- 8 years	10.49%	15
4	9-11 years	27.97%	40
5	12-15 years	28.67%	41
6	15-20 years	18.18%	26
7	Over 20 years	4.20%	6
Total Respondent		100%	143

Regarding to the experience of the respondents, it was found that high percent of them have good experience in construction field, and have advanced positions in job level, which gives more logic and reality of surveyed results to some extent.

### 3.4 Independent Variables

In the construction of an Artificial Neural Network (ANN) model for cost estimation in public projects within the Oromia Region, the research considered the top fifteen selected independent variables outlined in Table 7. These variables served as inputs to the ANN model, exerting influence on its predictions within specified ranges as indicated in Table

7. Additionally, parameters specified in Table 8 were excluded from consideration.

Table 7 Independent variables (selected influential factors)

Dependent variable	Independent variables	Range
Cost	Market Condition	Volatile to Stable
	Duration	4 Month (Short) to 24 Month
	Local Materials	Very Low-Very High
	Number of Floors	1-8 Stores
	Availability of Labor	Very Low-Very High
	Management Condition	Poor – Excellent
	Location of Project	Remote - Urban
	Floor Area	200m2 -2000m2
	Floor Height	2.5m-9m
	External Social Condition	Very Low- High
	Safety condition	Poor – Excellent
	Risk condition	Very Low- High
	Complexity of Building	Low- High
	Types of building	All public
Contractor	Grade-5 to Grade-1 GC/BC	

Table 8 Excluded variables

Variable	Excluded input	Included input
Project type	Mosque, Church	All public
Interior decoration	Luxury	Basic
Sanitary Works	Luxury	Basic, Normal
Type of external finishing	Luxury	Normal plaster, stones
Type of HVAC	Central	None, window, split
Type of Tiles	Luxury	Normal, Ceramic, granite
Type of Electricity Works	Luxury	Basic
Type of Mechanical Works	Luxury	Basic
Basement	Exists	Does not exist
Duration	More than 2 years	3-24 month

### 3.5 Data Encoding

Artificial networks only deal with numeric input data. Therefore, the raw data must often be converted from the external environment to numeric form (Kshirsagar & Rathod, 2012). This may be challenging because there are many ways to do it and unfortunately, some are better than others are for neural network learning (Principe, et al., 2010).

Categorical encoding was a crucial step in preparing categorical data for machine learning algorithms, including artificial neural networks (ANNs). It involved converting categorical variables (variables that could take on a limited, fixed number of values) into a numerical format that could be fed into machine learning models. Some common categorical encoding techniques included one-hot encoding, label encoding, and ordinal encoding. The data was converted to numeric form, as shown in the table below.

Table 9 Coding

No.	Input Factors	Types of Information	Unit	Encode	Code
1	Project types	Categorical	Number	Residential	1
				Educational	2
				Institutional	3
				Health care	4
				Industrial	5
				Assembly	6
				commercial	7
				Others	8
2	Market condition	Categorical	Number	Poor / Volatile	1
				Fair	2
				Good/Stable	3
3	Project Duration	Categorical	Number	Very Short/1-4 Months	1
				Short/5-9Months	2
				Medium/10-15 months	3
				Long /16-24Months	4

No.	Input Factors	Types of Information	Unit	Encode	Code
4	Area of typical floors	Categorical	Number	201-300m2	1
				301-400m2	2
				401-500m2	3
				501-700m2	4
				701-900m2	5
				901-1100m2	6
				1101-1300m2	7
				1301-1500m2	8
				1501-1700m2	9
5	Skilled labor	Categorical	Number	Very Low	1
				Low	2
				Medium	3
				High	4
				Very High	5
6	Local material	Categorical	Number	Very Low	1
				Low	2
				Medium	3
				High	4
				Very High	5
7	Management condition	Categorical	Number	Poor	1
				Fair	2
				Good	3
				Very Good	4
				Excellent	5
8	Complexity of building	Categorical	Number	Low	1
				Medium	2
				High	3
9	Project Location	Categorical	Number	Remote	1
				Rural	2
				Suburban	3
				Urban	4

No.	Input Factors	Types of Information	Unit	Encode	Code
10	Risk condition	Categorical	Number	Very Low	1
				Low	2
				Medium	3
				High	4
11	Floor height	Categorical	Number in (m)	2.5-3.2m	1
				3.3-4.0m	2
				4.1-5.0m	3
				5.1-6.0m	4
				6.1-7.0m	5
				7.1m-9m	6
12	Number of floors	Categorical	Number in (m)	G+0	1
				G+1	2
				G+2	3
				G+3	4
				G+4	5
				G+5	6
				G+6	7
13	Safety conditions	Categorical	Number	Poor	1
				Fair	2
				Good	3
				Very Good	4
				Excellent	5
14	External Social Condition	Categorical	Number	Very Low	1
				Low	2
				Medium	3
				High	4
15	Capacity of Contractor	Categorical	Number	Grade-1	1
				Grade-2	2
				Grade-3	3
				Grade-4	4
				Grade-5	5

#### 4. RESULTS AND DISCUSSIONS

##### 4.1 Influential Factors

Literature studies was the first process in determining the key parameters, then the questionnaire was designed according to these literatures for twenty-five selected parameters that acquired a final identification of most influential factors on building projects cost.

Table 10 The most influential factors that were adopted in this study

No.	Factors	Percent			Total	Rank
		Moderate	High	Very High		
1	Market Condition	17.48%	34.26%	48.25%	100%	1
2	Duration	15.38%	55.94%	23.07%	94.41%	2
3	Local Materials	20.97%	37.76%	34.27%	93.01%	3
4	Availability of Labor	30.07%	35.66%	25.17%	90.91%	4
5	Location of Project	46.15%	30.11%	12.58%	89.51%	5
6	Management Condition	39.86%	39.86%	9.09%	88.81%	6
7	Number of Floors	30.07%	41.96%	15.38%	87.41%	7
8	Floor Area	13.98%	47.55%	25.87%	87.41%	8
9	Floor Height	33.56%	49.65%	2.09%	85.31%	9
10	External Social Condition	23.08%	32.17%	28.67%	83.92%	10
11	Risk Condition	48.25%	28.67%	4.89%	81.82%	11
12	Safety Condition	39.86%	39.16%	2.09%	81.12%	12
13	Types of Building	40.56%	31.47%	8.39%	80.42%	13
14	Complexity of Building	37.76%	30.77%	9.79%	78.32%	14
15	Contractor Grade	39.16%	25.17%	11.89%	76.22%	15
16	Types of Structural Materials	32.87%	34.97%	6.29%	74.13%	16
17	Types of Contracts	33.57%	32.17%	3.5%	69.23%	17
18	Types of Foundation	40.56%	25.17%	3.49%	69.23%	18
19	Land Scape	43.35%	20.27%	4.19%	67.83%	19
20	Types of Slabs	44.75%	15.38%	2.09%	62.24%	20
21	Types of Soil	34.26%	25.17%	0.69%	60.14%	21
22	Open Space	34.96%	11.18%	0%	46.15%	22
23	Earth quack	28.67%	14.68%	2.09%	45.45%	23
24	Roof Area	34.96%	8.39%	0%	43.36%	24
25	Weather Condition	23.08%	6.99%	0.69%	30.77%	25

---

#### 4.1.1 Market Condition: -

Market conditions encompass the prevailing economic, financial, and business environment that influences the behavior of buyers and sellers in a specific market. These conditions significantly impact various aspects of business operations, investment choices, and consumer behavior. In the construction industry, market conditions play a crucial role in cost estimation, affected by fluctuations in material and labor costs, interest rates, regulations, inflation, supply chain disruptions, currency values, and economic uncertainty. These factors contribute to increased expenses, delays, and challenges in estimating construction project costs.

Approximately 70.48% of projects were impacted by volatile market conditions, characterized by highly unpredictable and unstable market environments. During such periods, prices and market indicators undergo significant fluctuations and rapid changes. Volatility can stem from economic uncertainty, geopolitical tensions, natural disasters, or sudden shifts in supply and demand. In volatile markets, projects encounter greater risks and challenges in making accurate predictions or decisions due to the unpredictable nature of price movements.

Table 11 Market condition

<b>Market Condition</b>	<b>No. of projects</b>	<b>Percentage (%)</b>
Poor /Volatile	74	70.48 %
Fair	24	22.86%
Good/Stable	7	6.67%
<b>Total</b>	<b>105</b>	<b>100</b>

#### 4.1.2 Project Duration: -

The duration of the project significantly impacted cost estimation at the early stages of construction due to various factors. Longer project durations resulted in higher labor costs as extended timelines required more labor hours, leading to increased expenses for wages, salaries, and benefits. Cost estimation had to consider the anticipated duration of labor requirements. Additionally, longer project durations often necessitated the rental or purchase of construction equipment for an extended period, resulting in higher equipment costs due to increased rental fees, maintenance expenses, and depreciation over time. Prolonged project durations also led to higher material costs due to factors like inflation, storage expenses, and material wastage, requiring cost estimation to consider potential fluctuations in material prices over the project timeline. Moreover, delayed completion

of construction projects could result in opportunity costs for project stakeholders, including lost revenue, missed market opportunities, and delayed return on investment, which cost estimation had to consider. Furthermore, longer project durations might lead to additional contractual obligations, such as extended warranties, penalty clauses for delays, or liquidated damages provisions, requiring cost estimation to account for the potential financial implications of these obligations. Lastly, extended project durations could affect the availability and allocation of resources, including labor, equipment, and materials, necessitating cost estimation to consider the impact of resource constraints on project costs and schedule.

In the data conducted for this paper, it was found that out of the 105 projects considered, 37.14% fell into the category of medium duration contracts. These projects spanned a range of durations, indicating a significant portion of the sample fell within this timeframe. Moreover, 23.4% of the projects were classified as long duration contracts, specifically lasting between 19 to 24 months. This subset represented a substantial portion of the sample as well; indicating a considerable number of projects had longer timelines. Conversely, only 14.29% of the projects were categorized as short duration contracts. This minority subset suggests that a smaller proportion of the sample had relatively brief durations in terms of contract length. These findings provide insights into the distribution of project durations within the dataset, highlighting the varying timelines observed across the projects analyzed.

Table 12 Project Duration

<b>Project Duration</b>	<b>No. of projects</b>	<b>Percentage (%)</b>
3-6month (Very Short)	15	14.29 %
7-12month (Short)	26	24.76%
13-18(month)	39	37.14%
19-24(long)	25	23.81%
<b>Total</b>	<b>105</b>	<b>100</b>

#### 4.1.3 Availability of Local Materials and Skilled Labors: -

The availability of local materials significantly impacted cost estimation during the early stages of construction. Local materials, often cheaper than imported ones due to reduced transportation costs and tariffs, offered cost savings in procurement, thus lowering overall project expenses. Their shorter transportation distances minimized associated costs,

which were factored into cost estimation. Moreover, local materials' shorter lead times facilitated quicker procurement and construction, reducing delays and positively affecting cost estimation. Access to them ensured better quality control, reducing inspection expenses, while their consistent availability minimized supply chain disruptions and material shortages, factors integrated into cost estimation to mitigate project risks. Additionally, local materials' adaptability met project requirements efficiently, while their usage contributed to the local economy, fostering community support and reducing environmental impact, aspects considered in comprehensive cost estimation.

In the dataset analyzed, it was observed that only 31.41% of the projects fell within areas where local materials were readily available. Within this subset, 25.71% of the projects were situated in regions characterized by high availability of local materials, while 5.71% were in areas where local material availability was very high. This suggests that a relatively small proportion of the projects benefited from ample access to locally sourced materials, with a majority facing varying degrees of scarcity in this regard. Additionally, it was noted that 25.72% of the projects were impacted by a shortage of local material availability. This indicates that a significant portion of the analyzed projects encountered challenges related to the procurement of locally sourced materials, potentially affecting project timelines and costs. These findings underscore the importance of considering the availability of local materials in cost estimation and project planning, particularly given their impact on construction project feasibility and economics.

Table 13 Local Material, Skilled Labor, Social, Risk Conditions and Complexity

Influencing Factors	Responses in Percentage (%)				
	Very Low	Low	Medium	High	Very High
Availability of Local Materials	2.86%	22.86%	42.86%	25.71%	5.71%
Availability of Skilled Labor	11.43%	32.38%	22.86%	20.95%	12.38%
External Social Condition	14.29%	41.90%	38.10%	5.71%	0%
Risk Condition	8.51%	50.48%	32.51%	8.51%	0%
Complexity of Building	0%	39.04%	32.38%	28.57%	0%

The availability of skilled labor played a crucial role in influencing cost estimation during the initial phases of construction. Skilled labor, requiring specialized knowledge and training, typically commanded higher wages compared to unskilled or semi-skilled labor. Shortages in skilled labor often resulted in increased wages and higher labor expenses.

---

However, skilled labor's proficiency and efficiency in construction tasks typically translated to higher productivity, potentially reducing project durations, labor hours, and overall labor costs. The presence of skilled labor directly impacted project timelines and efficiency, ensuring workmanship and construction quality met specified standards. Availability of skilled craftsmen minimized rework, defects, and associated costs, thereby reducing the risk of construction delays, errors, and safety incidents. Skilled labor's ability to anticipate and mitigate risks contributed to lower overall project costs, efficient resource allocation, and adherence to budgetary constraints.

According to Table 13, it was observed that 43.81% of the projects analyzed were impacted by a shortage of low-skilled labor. Conversely, only 20.95% of the projects were situated in areas characterized by high availability of skilled labor, while 12.38% were in regions where the availability of skilled labor was classified as very high. These findings shed light on the distribution of skilled labor availability across the projects under consideration. The prevalence of low-skilled labor shortages in a significant portion of the projects highlights a potential challenge in sourcing labor with the necessary expertise and qualifications. Conversely, the relatively smaller percentages of projects benefiting from high or very high levels of skilled labor availability suggest varying degrees of accessibility to skilled workforce across different project locations. Such insights are crucial for understanding the labor dynamics within the construction industry and its implications for project planning and execution.

#### 4.1.4 Management and Safety Conditions: -

Management conditions encompass various factors related to project management practices, organizational capabilities, and leadership styles that influenced cost estimation during the early stages of construction. Effective project planning and coordination contributed to efficient resource allocation, streamlined workflows, and timely task completion. Well-defined project plans and schedules aided in more accurate cost estimation by identifying milestones, resource needs, and potential risks early on. Open and transparent communication fostered collaboration among stakeholders, facilitating information exchange, clarification of requirements, and issue resolution, thereby enhancing cost estimation accuracy. Efficient resource allocation ensured optimal utilization of labor, materials, and equipment, minimizing waste. Cost estimators considered management practices promoting resource efficiency to avoid misestimating of requirements. Strong supplier and contractor relationships resulted in favorable pricing, timely deliveries, and improved project performance, factors assessed when

estimating procurement and subcontracting costs. The table provided illustrated that approximately 18.00% of the projects fell within the poor to fair range, indicating a relatively smaller proportion characterized by lower performance or quality levels. Conversely, the vast majority, accounting for 82.00% of the projects, were situated in the good, very good, and excellent ranges. This suggests that the predominant share of projects exhibited higher levels of performance or quality, reflecting satisfactory to exceptional standards. These findings provide insights into the distribution of project performance across the dataset, highlighting the prevalence of projects meeting or exceeding expected standards. Such information is valuable for assessing overall project quality and identifying areas for improvement in future endeavors.

Table 14 Management and Safety conditions

Influencing Factors	Responses in Percentage (%)				
	Very Low	Low	Medium	High	Very High
Management Conditions	3.81%	14.29%	38.10%	34.29%	9.52%
Safety Conditions	9.52%	30.48%	37.14%	17.14%	5.71%

Safety conditions played a critical role in cost estimation during the early stages of construction due to their direct and indirect impact on project expenses. Implementing safety measures and providing necessary equipment for construction workers incurred additional costs, which needed to be factored into cost estimation. This encompassed expenses for personal protective equipment (PPE), safety signage, fall protection systems, and safety training programs. Poor safety records or hazardous working conditions could lead to higher insurance premiums, while meeting regulatory requirements involved compliance costs such as safety inspections and permits. Investing in safety training and education programs was crucial for fostering a culture of safety and reducing accidents. Safety-related incidents on construction sites incurred both direct and indirect costs, including medical expenses, legal fees, and productivity losses. Safe working conditions not only enhanced worker productivity and efficiency but also minimized project delays and disruptions. Proactively managing safety risks and incorporating contingency allowances for safety-related uncertainties were essential for effective risk management. Maintaining a strong reputation for safety practices was vital for attracting clients and skilled workers, underscoring the long-term benefits of investing in safety measures and promoting a positive safety culture in the construction industry.

---

According to Table 14, it was evident that 40.00% of the projects analyzed exhibited poor to fair safety conditions, indicating a significant portion characterized by suboptimal safety standards. In contrast, 37.14% of the projects were classified as having excellent safety conditions, showcasing a substantial percentage adhering to high safety standards. Additionally, 17.14% of the projects fell within the category of good safety conditions, reflecting a notable proportion meeting acceptable safety criteria. Conversely, only 5.71% of the projects were rated as having excellent safety conditions, suggesting a relatively smaller percentage with exemplary safety practices. These findings provided valuable insights into the distribution of safety conditions across the projects, highlighting variations in safety standards and their potential impact on project outcomes. Understanding these dynamics was crucial for assessing the overall safety landscape within the construction industry and its implications for cost estimation and project management.

#### 4.1.5 Location of Project: -

The location of a construction project exerted a significant influence on cost estimation during the early stages. Variations in construction costs across regions stemmed from disparities in labor rates, material prices, land expenses, regulatory demands, and market conditions. Estimators had to incorporate these regional cost differences into their calculations. Site-specific factors like topography, soil conditions, climate, and environmental regulations influenced construction methods, material needs, and project timelines. Sites with challenging conditions often required additional preparatory work or environmental measures, impacting project costs. Estimators adjusted estimates to accommodate such site characteristics. Accessibility of utilities, including water, electricity, and telecommunications infrastructure, varied by location, necessitating additional investments in remote or underdeveloped areas, which estimators considers when estimating project costs.

According to Table 15, it was observed that 10.48% of the projects were constructed in remote areas, indicating a relatively small proportion situated in locations far removed from urban centers or major infrastructure. Conversely, 29.52% of the projects were situated in rural areas, reflecting a larger share in less densely populated regions characterized by agricultural or natural landscapes. Furthermore, 42.86% of the projects were located in suburban areas, signifying a significant portion situated on the outskirts of urban centers with moderate population density. In contrast, only 17.14% of the projects were in urban areas, indicating a comparatively smaller proportion situated in

densely populated urban centers or metropolitan areas. These findings provide insights into the distribution of construction projects across different types of locations, highlighting variations in project settings and potential implications for cost estimation and project management.

Table 15 Project location

<b>Project Location</b>	<b>No. of projects</b>	<b>Percentage (%)</b>
Remote	11	10.48%
Rural	31	29.52 %
Suburban	45	42.86%
Urban	18	17.14%
<b>Total</b>	<b>105</b>	<b>100</b>

#### 4.1.6 External Social Condition and Risk Conditions: -

External social conditions had a significant impact on cost estimation during the early stages of construction in several ways. Social factors, including population demographics, immigration policies, and labor unionization, influenced the availability and cost of construction labor. Changes in labor market dynamics, such as labor shortages or increases in minimum wage laws, affected labor costs and project timelines, thus influencing cost estimation. Societal trends towards diversity and inclusion impacted workforce composition and labor practices in the construction industry. Additionally, the emphasis on corporate social responsibility and ethical business practices influenced construction project requirements, specifications, and procurement processes. Positive community relations and stakeholder engagement were crucial for obtaining necessary approvals, permits, and local support for construction projects. Concerns for worker health and safety influenced industry regulations, standards, and practices, adding costs such as safety training and protective equipment, which were factored into cost estimation. Cultural factors influenced project design, aesthetics, and construction practices, particularly in culturally diverse or heritage-sensitive environments. Public perception of construction projects also played a role in project planning, execution, and outcomes. Changing social trends and preferences, such as urbanization, sustainability, and lifestyle choices, impacted industry practices and market demand for specific types of projects.

---

The table 13 presented above indicated that 56.19% of the projects were influenced by low external social effects, suggesting that the majority of projects were subject to relatively minimal social influence from external factors. Conversely, only 5.71% of the projects were affected by high external social effects, indicating a much smaller proportion facing significant social influence. These findings provided insights into the varying degrees of social impact on the projects analyzed, highlighting the prevalence of projects with lower levels of external social influence compared to those with higher levels. Understanding these dynamics was essential for assessing the broader social context surrounding construction projects and its potential implications for cost estimation and project management.

Risk conditions had a significant impact on cost estimation during the early stages of construction. Identifying potential risks associated with the project was crucial for ensuring accurate cost estimates. These risks encompassed a range of factors, including design changes, material shortages, labor issues, weather disruptions, regulatory shifts, site conditions, and unforeseen events. After identifying these risks, cost estimators assessed their likelihood and potential impact on the project. Risks with higher probabilities and significant consequences received greater attention, leading to the inclusion of contingency allowances in the project budget to mitigate their impact. Contingency allowances represented additional funds set aside to cover unforeseen expenses or project delays resulting from risks. The size of these allowances depended on the level of risk exposure and uncertainty associated with the project. Implementing risk mitigation strategies, such as transferring risk to third parties or obtaining insurance coverage, helped reduce the likelihood or severity of identified risks. Cost estimators also assessed insurance costs associated with various coverage options to protect against financial losses from unforeseen events. Alongside cost contingencies, schedule contingencies were necessary to address risks that might impact project timelines. Throughout the project lifecycle, cost estimators monitored and controlled identified risks to minimize their impact on project costs, ensuring continuous risk management and timely interventions to address emerging risks.

The table 13 indicated that 8.57% of the projects analyzed were categorized as being in high-risk conditions, highlighting a relatively small proportion facing significant risk levels. Conversely, the vast majority of projects, totaling 91.43%, were constructed in conditions ranging from very low to medium risk, underscoring a prevailing trend of projects operating within relatively lower risk thresholds. These findings provided

---

valuable insights into the distribution of risk conditions across the projects, emphasizing the prevalence of projects with lower risk exposure compared to those with higher risk levels. Understanding these dynamics was crucial for assessing the overall risk landscape within the construction industry and its implications for cost estimation and project management.

#### 4.1.7 Types and Complexity of Building: -

The type of building under construction played a significant role in cost estimation during the early stages. Each building type presented distinct requirements, specifications, and complexities that directly influenced project costs. For instance, multi-family residential buildings incurred higher construction costs per unit due to shared amenities and common areas. Office buildings' cost estimation factored in considerations such as floor plan layout, interior finishes, and building systems. Retail buildings required attention to store layouts, display areas, and customer amenities, while industrial buildings focused on clear spans and floor load capacities. Educational buildings needed spaces for classrooms and laboratories, while healthcare buildings required facilities for patient care and medical equipment. Government buildings incorporated features for offices and courtrooms, while cultural buildings had unique design elements for exhibits and performance spaces. These factors highlighted the diverse range of considerations involved in cost estimation for different building types, emphasizing the importance of understanding each type's specific requirements to develop accurate cost estimates.

The utilization of building factors comprised six subtypes, as illustrated in Table 16. Each subtype exhibited a well-distributed representation of projects within the dataset. Among these subtypes, residential buildings accounted for 8.57% of the projects, while educational buildings constituted the highest proportion at 14.29%. Schools followed closely behind at 13.33%, while healthcare facilities represented 15.24% of the projects. Industrial buildings comprised 20% of the dataset, indicating a notable presence. Additionally, other building types collectively represented 13.33% of the projects. These findings provided insights into the distribution of project types based on their utilization factors, showcasing the varied landscape of construction projects across different sectors. Understanding these distributions was essential for comprehensively assessing the factors influencing cost estimation and project management in diverse building contexts.

Table 16 Types of building

Types of building	No. of projects	Percentage (%)
Residential	9	8.57 %
Educational	15	14.29%
Institutional	14	13.33%
Health care	16	15.24 %
Industrial	21	20.00 %
Assembly	8	7.62 %
Commercial	8	7.62 %
Others	14	13.33 %
<b>Total</b>	<b>105</b>	<b>100</b>

The complexity of building designs had a profound impact on cost estimation during the early stages of construction. Complex designs often necessitated extensive architectural and engineering services to develop detailed plans and specifications, including structural analysis and MEP design. Specialized materials and finishes were often required to meet aesthetic and performance requirements, contributing to higher material costs and longer procurement schedules. Moreover, complex designs often required specialized construction methods such as prefabrication and modular construction, incurring additional costs for labor and expertise. Skilled labor and specialized tradespeople were essential for executing intricate tasks, influencing labor costs and availability. Challenges related to site logistics and constraints further impacted construction activities, requiring effective coordination among various disciplines and subcontractors. Cost estimators had to account for these complexities and coordination efforts when estimating project costs. The table 13 illustrated that only 28.57% of the projects fell into the category of High complexity, while the majority, comprising 71.43%, were classified as medium to low complexity. These findings provided insights into the distribution of project complexities within the dataset, indicating that a significant portion of the projects faced low to medium levels of complexity. Understanding these nuances was crucial for assessing the varying degrees of complexity across construction projects and their potential implications for cost estimation and project management.

#### 4.1.8 Contractor: -

The grade or capacity of a contractor reflected their ability to successfully execute construction projects within specified parameters of time, budget, and quality standards.

This capacity held significant implications for cost estimation during the early stages of construction. Contractors with higher capacity generally possessed access to a broader array of resources, including skilled labor, equipment, and subcontractors. Their enhanced efficiency and productivity often translated into reduced labor costs and shorter project durations. Moreover, contractors with greater capacity tended to be more competitive in bidding processes, offering favorable pricing and terms to secure contracts. With proven track records, they were adept at managing project risks and maintaining quality assurance measures. Additionally, their robust project management capabilities, financial stability, and extensive industry experience further underscored their reliability and suitability for construction projects. Understanding these contractor capacities was crucial for developing accurate cost estimates and ensuring successful project outcomes. Table 17 displayed a breakdown of contractors based on their grade, revealing distinct proportions within each category. Grade-1 contractors comprised 40% of the total, indicating a significant presence in the dataset. Grade-2 contractors accounted for 10.48%, while Grade-3 contractors represented 27.62%. Grade-4 contractors constituted 11.42% of the dataset, followed closely by Grade-5 contractors at 10.48%. These findings highlighted the varied distribution of contractors across different grades, with Grade-1 contractors being the most prevalent. Understanding these distributions provided valuable insights into the capacity and capabilities of contractors available for construction projects, aiding in informed decision-making during the early stages of project planning and estimation.

Table 17 Grade or Capacity of Contractor

Capacity of Contractor	No. of projects	Percentage (%)
Grade-5	11	10.48%
Grade-4	12	11.42%
Grade-3	29	27.62%
Grade-2	11	10.48%
Grade-1	42	40.00%

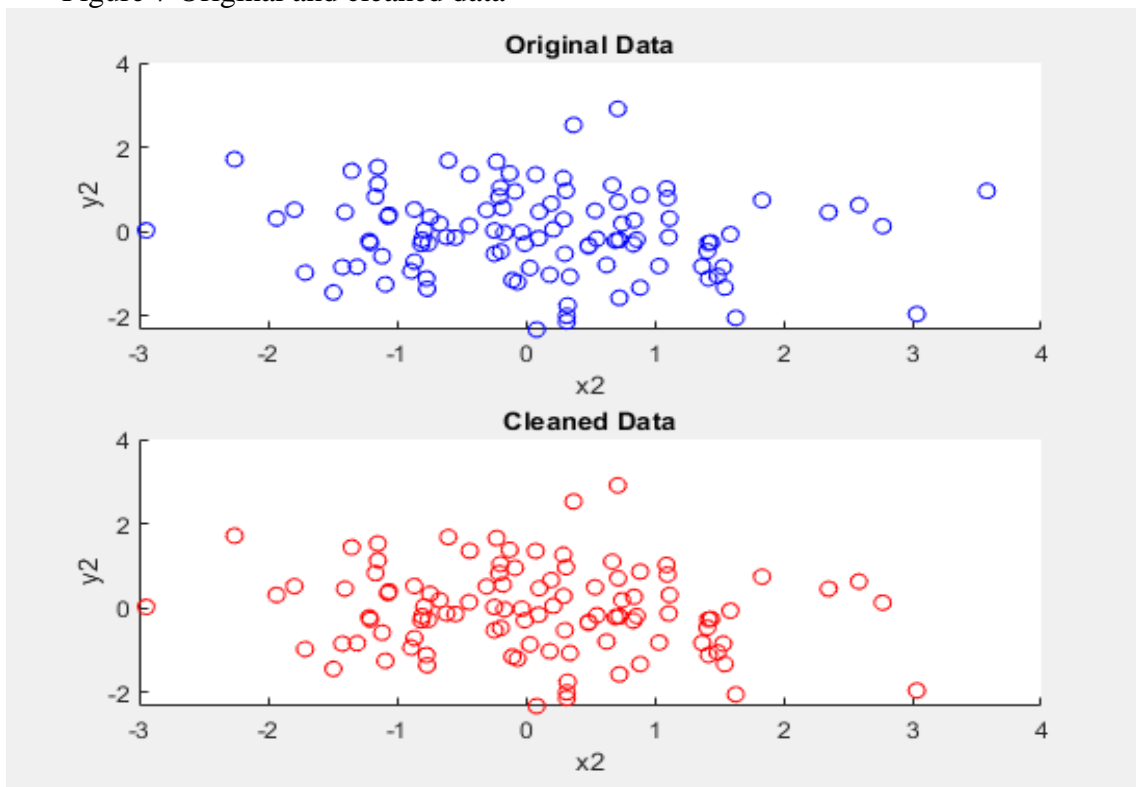
#### 4.2 Developing ANN Model

The development of an Artificial Neural Network (ANN) model for construction cost estimation involves several key steps, including defining the model architecture, preparing the data, training the model, and evaluating its performance.

#### 4.2.1 Data preprocessing and clearing: -

The data underwent thorough preparation for input into the neural network, involving normalization and standardization through log and min-max functions to ensure feature consistency. Cleaning procedures, including outlier detection using a z-score method, were applied to handle missing values, remove duplicates, and address inconsistencies. This meticulous preparation ensured the dataset's integrity and reliability, optimizing its suitability for analysis within the artificial neural network model.

Figure 7 Original and cleaned data



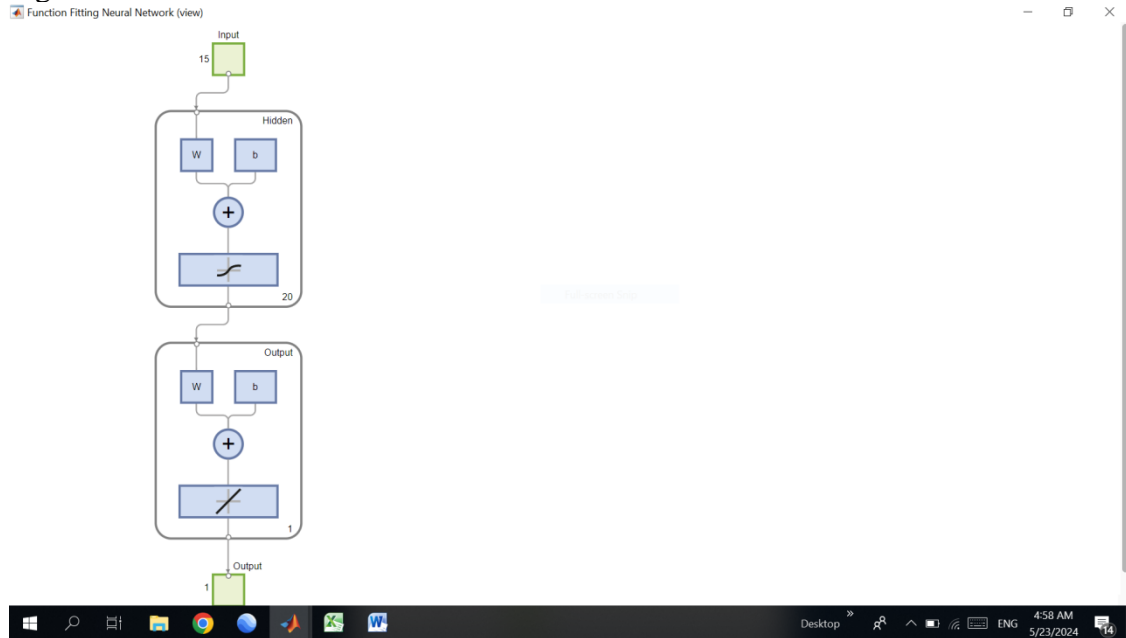
From 106 projects only one project (project-9) was outlier and 105 projects were used for in feeding ANN.

#### 4.2.2 Architecture Design: -

The process of designing the architecture of an artificial neural network (ANN) entailed several crucial steps, including the determination of the number of layers, the allocation of neurons within each layer, and the selection of appropriate activation functions. This architectural framework plays a pivotal role in enabling the network to effectively learn intricate relationships embedded within the dataset. In this particular case, utilizing the 15 independent variables previously selected as the input layer, along with one hidden layer comprising 20 neurons, and designating the dependent variable as the output layer, a neural network was defined. This neural network, depicted in a figure (not shown here), was

constructed using MATLAB code as detailed in Appendix-3A. The use of the 'logsig' transfer function further characterized the network's behavior and functionality contributing to its overall performance in analyzing the dataset.

Figure 8 Network Architecture



Designing an effective architecture often involves a degree of trial and error. It's essential to understand the characteristics of data and problem domain and to monitor the model's performance during training and validation. The architecture was adjusted based on the observed behavior and aim for a balance between model complexity and generalization to achieve optimal performance on construction cost estimation

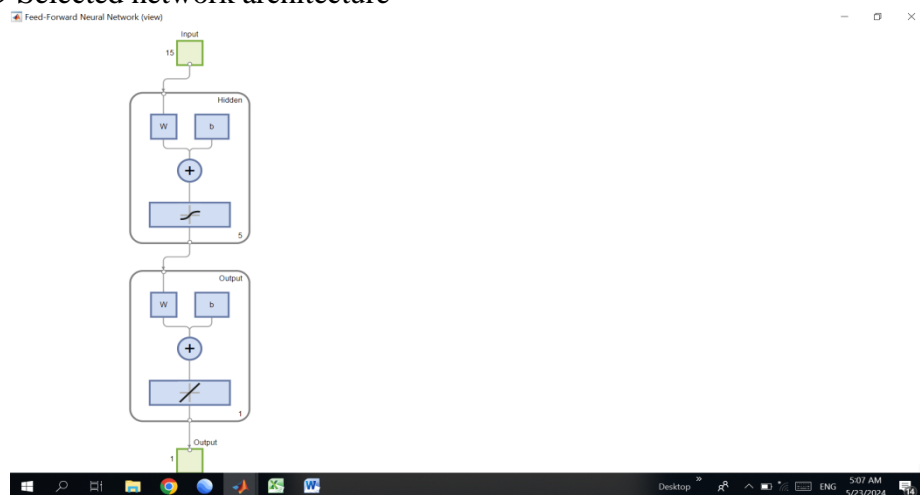
#### 4.2.3 Input and Output Layers: -

The process of selecting suitable parameters for an Artificial Neural Network (ANN) comprised making informed decisions regarding multiple hyper parameters that significantly impact the network's architecture, training, and optimization. These parameters were carefully chosen based on the expertise and preferences of construction professionals, encompassing a range of factors such as project types, market conditions, project duration, floor area, availability of skilled labor, availability of local materials, management conditions, complexity of the building, project location, risk conditions, floor height, number of floors, safety conditions, external social conditions, and the grade or capacity of the contractor. Each of these parameters played a crucial role in shaping the ANN's ability to effectively analyze and model the dataset, thereby enhancing its performance in addressing the complexities inherent in construction projects.

#### 4.2.4 Number of Hidden Layers and Neurons: -

In the quest to determine the optimal number of hidden layers and neurons, various configurations of neural network architecture were defined and evaluated using MATLAB code provided in Appendix 3B. The number of hidden layers plays a pivotal role in shaping the network's performance, influencing its ability to effectively learn complex patterns from the data. By experimenting with different architectures, wherein the number of hidden layers and neurons per layer were systematically varied, researchers aimed to identify the configuration best suited for the given problem. Additionally, the number of neurons in each hidden layer was a critical consideration, as it directly impacted the model's capacity to capture intricate data patterns while avoiding both under fitting and over fitting scenarios. After thorough experimentation, a neural network with a single hidden layer comprising five neurons emerged as the optimal choice, reflecting a carefully balanced architecture poised to yield optimal performance in addressing the complexities inherent in the dataset.

Figure 9 Selected network architecture



#### 4.2.5 Activation Functions: -

In the pursuit of selecting the most effective activation functions for the neural network, a systematic approach akin to previous methodologies was employed, involving the exploration of various activation functions for the hidden layers. This process was facilitated either through a MATLAB code provided in Appendix-3C or through manual trial and error. Activation functions, pivotal in shaping the network's capacity to capture non-linear relationships within the data, encompassed popular options such as Relu (Rectified Linear Unit), sigmoid, tanh, purelin, logsig, and softmax. By iteratively applying each activation function, training the neural network, evaluating its performance on a testing set, and subsequently assessing the resulting accuracy, the optimal activation function was identified. Noteworthy outcomes included a performance of 0.4440 with the

---

tansig function, 0.7157 with logsg, and 0.3361 with purelin. Ultimately, after meticulous evaluation, the purelin activation function emerged as the most suitable choice, boasting a performance metric of 0.3361, indicative of its efficacy in capturing the nuances of the dataset.

#### 4.2.6 Loss Function: -

Loss during the training phase of the neural network, the definition of loss functions was a critical step, facilitated by a MATLAB code provided in Appendix 3D. Utilizing this code, a neural network was trained with the specified architecture and training parameters, incorporating a custom loss function. Subsequently, upon completion of the training process, the model's performance was assessed by computing the loss on the testing set using the same custom loss function. This iterative approach allowed for the adjustment of not only the loss function but also the network's architecture and training parameters to cater to specific requirements. Ultimately, from the three evaluated loss functions—'mse', 'mae', and 'crossentropy'—the best-performing loss function was determined to be 'mae', with a recorded performance metric of 27.2242, signifying its efficacy in guiding the model's learning process and optimizing its predictive accuracy.

#### 4.2.7 Batch Size & Number of Epochs: -

The optimization of neural network training heavily relies on the selection of crucial hyper parameters, namely the batch size and number of epochs. These parameters represent arrays of different batch sizes and numbers of epochs, respectively, aimed at fine-tuning the training process. Employing a MATLAB code provided in Appendix 3E, the neural network's parameters were configured within a nested loop, systematically iterating over various combinations of batch sizes and numbers of epochs. Following the training of each network configuration, its performance was rigorously evaluated using the mean absolute error (mae) loss function on the testing set. The batch size, influencing the number of samples processed before updating the model's parameters, required a delicate balance between computational efficiency and model performance, where larger batch sizes promised faster convergence but demanded more memory. Similarly, epochs, representing complete passes through the training dataset, played a crucial role in determining the model's exposure to data and potential for over fitting or under fitting. Ultimately, the configuration yielding the lowest loss on the testing set—batchSize\_64\_epochs\_300—was identified as the optimal choice, boasting a performance metric of 21.8356 and showcasing its effectiveness in guiding the network's training process toward enhanced predictive accuracy.

---

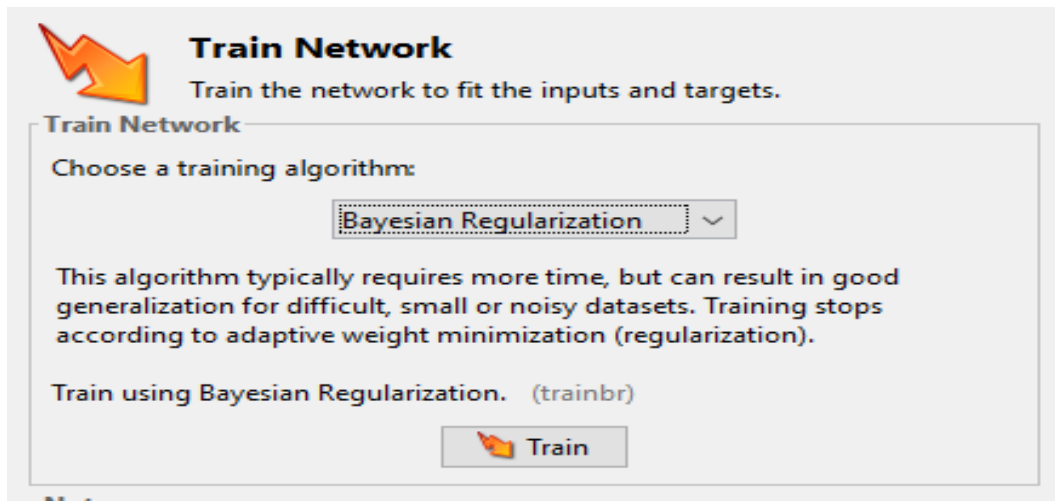
#### 4.2.8 Learning Rate: -

The learning rate, crucial in dictating the pace at which a model's parameters are adjusted during training, played a pivotal role in the optimization of neural network performance. With the potential to either facilitate rapid convergence or risk overshooting the optimal solution, selecting an appropriate learning rate demanded careful consideration. Common strategies for fine-tuning the learning rate encompassed methods such as grid search, random search, or the utilization of adaptive learning rate techniques like Adam, RMSprop, or AdaGrad. Within a MATLAB code provided in Appendix 3F, the `net.trainParam.lr` parameter of the neural network was systematically adjusted within a loop, allowing for the iteration over different learning rates. Subsequently, each network configuration was trained, and its performance meticulously evaluated using the mean absolute error (mae) loss function on the testing set. Ultimately, the learning rate yielding the lowest loss on the testing set—0.001—was identified as the optimal choice, epitomizing its effectiveness in guiding the network's training process towards superior predictive accuracy.

#### 4.2.9 Regularization Parameter: -

Regularization techniques serve as essential safeguards against over fitting within neural networks, a facet extensively explored through a MATLAB code provided in Appendix 3G. These techniques, including L1 regularization, L2 regularization, and dropout, were systematically employed to mitigate the risk of over fitting by either penalizing large parameter values or randomly dropping neurons during training. The optimization of regularization strength emerged as a crucial consideration, striking a delicate balance between averting over fitting tendencies while preserving the model's performance on unseen data. Additionally, the choice of optimization algorithm wielded significant influence over the model's parameter updates during training, with alternatives such as stochastic gradient descent (SGD), Adam, RMSprop, or AdaGrad offering distinct advantages. Through rigorous experimentation, Bayesian Regularization (`trainbr`) emerged as a particularly effective regularization technique, surpassing alternatives like Levenberg-Marquardt (`trainlm`) and scaled conjugate gradient (`trainscg`) in promoting robust generalization performance within the neural network.

Figure 10 Training Algorithm for robust generalization

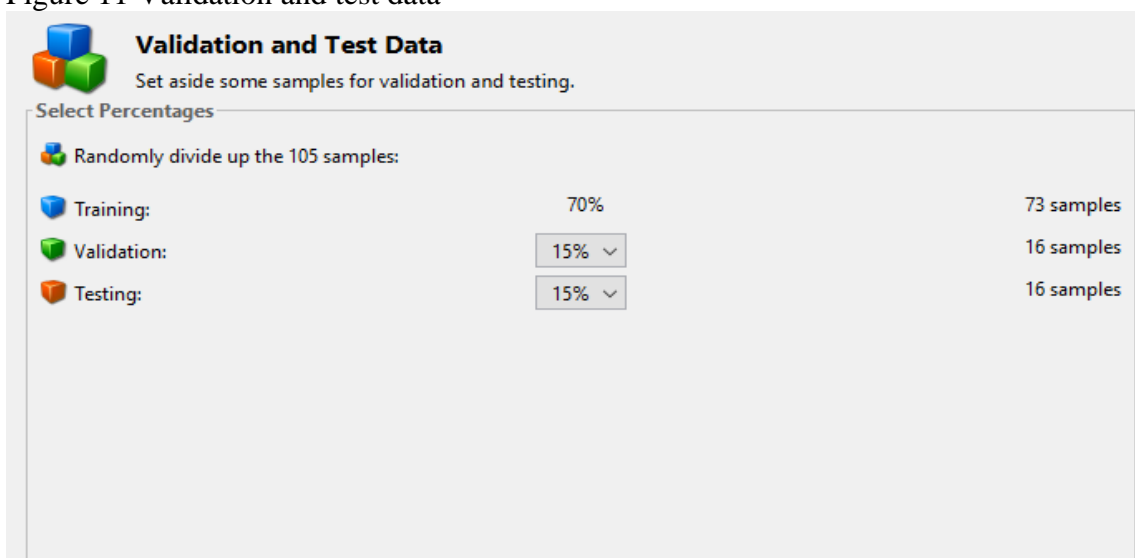


#### 4.2.10 Training and Validation: -

Data splitting and training are critical steps in the development of an Artificial Neural Network (ANN) model for construction cost estimation. Properly dividing the dataset into training and validation sets, along with effective training procedures, is essential for building a model that generalizes well to new, unseen data.

The training dataset comprised pairs of input features alongside their respective target values, forming the foundation for the neural network's learning process. To facilitate effective model training and performance optimization, the dataset underwent a meticulous partitioning, with 70% of the data allocated for training purposes. An additional 15% of the dataset was earmarked for the validation set, utilized to fine-tune hyperparameters and preemptively counteract overfitting tendencies. The remaining 15% of the data served as the evaluation set, pivotal in gauging the final model's performance and assessing its efficacy in real-world predictive scenarios.

Figure 11 Validation and test data



---

## 4.3 Results

### 4.3.1 Model Evaluation: -

Model evaluation is a crucial step in assessing the performance of an Artificial Neural Network (ANN) for construction cost estimation. It involves analyzing how well the trained model generalizes to new, unseen data. Some of the several metrics and techniques that can be employed to evaluate the model's effectiveness are discussed below.

#### 4.3.1.1 Mean Absolute Error (MAE): -

In construction projects, accurate cost estimation plays a pivotal role in budgeting and resource allocation decisions. Mean Absolute Error (MAE) served as a direct quantifier of the average error in cost prediction, offering practical insights into the model's efficacy in real-world scenarios. Monitoring MAE fluctuations throughout model development guided parameter tuning endeavors, encompassing adjustments to the network architecture, regularization techniques, or input features. The iterative minimization of MAE facilitated the enhancement of the model's capacity to grasp the intricacies of construction cost estimation, ultimately leading to improved accuracy in predictions. Lower MAE values not only bolstered client satisfaction by furnishing more dependable cost estimates for construction endeavors but also cultivated greater confidence among stakeholders, fostering positive relationships. However, while the presented MAE of 27.2242 may seem largely acceptable for projects of substantial worth, it's crucial to recognize its limitations, particularly in cases where project costs vary significantly. Despite this, Mean Absolute Error (MAE) remains a valuable performance metric, offering resilience to outliers, interpretability, and practical relevance, thereby proving instrumental in evaluating and refining the accuracy of cost estimation in construction projects.

#### 4.3.1.2 Mean Squared Error (MSE): -

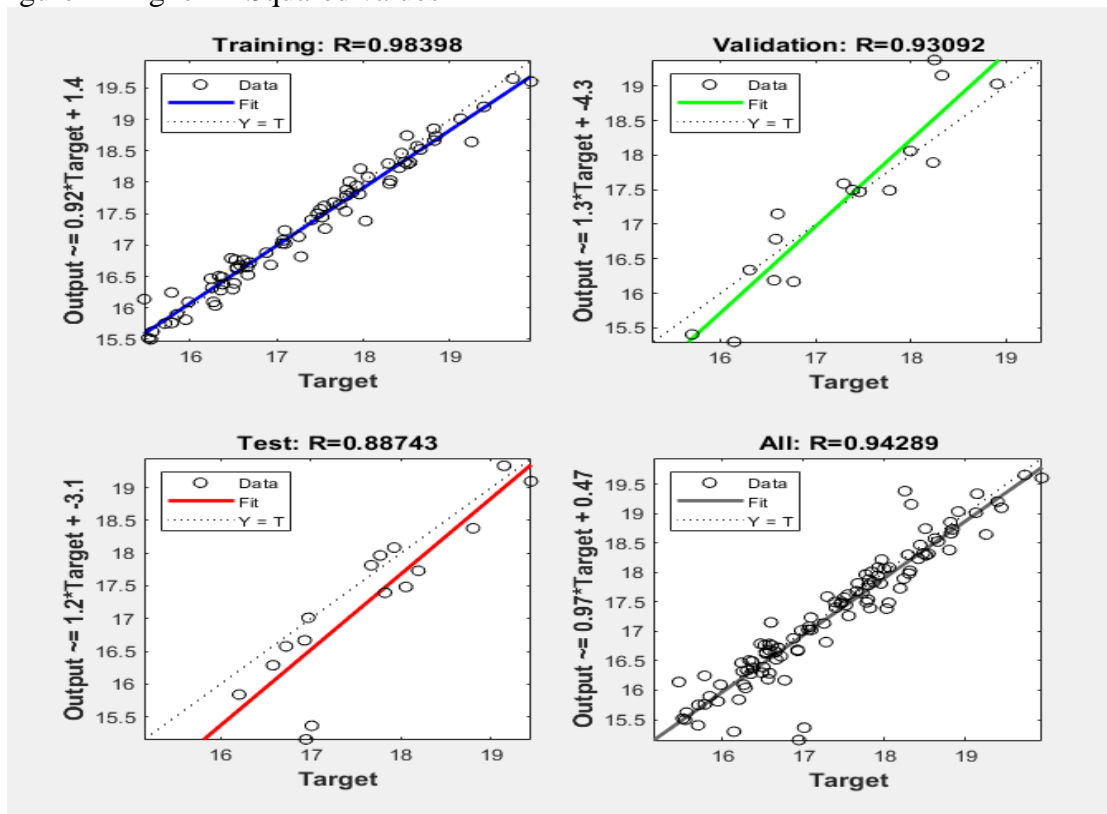
In the realm of construction cost prediction, Mean Squared Error (MSE) served as a crucial metric, quantifying the average squared difference between predicted and actual construction costs across the dataset. A lower MSE denoted closer alignment between model predictions and actual costs, whereas a higher MSE indicated larger disparities. Elevated MSE values could point to outliers or instances of poor predictive performance, prompting adjustments to input data or model architecture to enhance robustness. Monitoring MSE fluctuations during model development informed the selection of optimal model parameters, including the number of hidden layers, neurons, or

optimization algorithms, fostering iterative improvements in predictive capability. In summary, MSE emerged as a fundamental metric guiding the assessment and refinement of ANN models in construction cost prediction, playing a pivotal role in enhancing the reliability of cost estimates in construction projects.

#### 4.3.1.3 R-Squared (Coefficient of Determination): -

In this context, R-Squared served as a fundamental metric for assessing the explanatory power of the artificial neural network (ANN) model in predicting construction costs, quantifying the proportion of variance in the dependent variable (construction costs) explained by the independent variables (features used in the prediction model). Higher R-Squared values indicated a better fit of the model to the construction cost data, suggesting that a larger proportion of the variance in construction costs was captured. Monitoring changes in R-Squared during model development guided efforts aimed at enhancing the ANN model's performance through iterative adjustments to architecture, input features, or training parameters. Higher R-Squared values instilled greater confidence in the accuracy and reliability of cost estimates provided by the ANN model. Additionally, the correlation coefficient (R) derived from linear regression analysis indicated a strong linear relationship ( $R = 0.94289$ ) between the actual and estimated costs, further affirming the model's predictive capabilities during the testing phase, as illustrated graphically below.

Figure 12 higher R-Squared values



In summary, R-Squared is a valuable metric in construction cost prediction using ANN models. Its ability to quantify the proportion of variance explained, assess model fit, guide model improvement, and instill confidence in predictions makes it an essential tool for evaluating and enhancing the accuracy of cost estimation in construction projects.

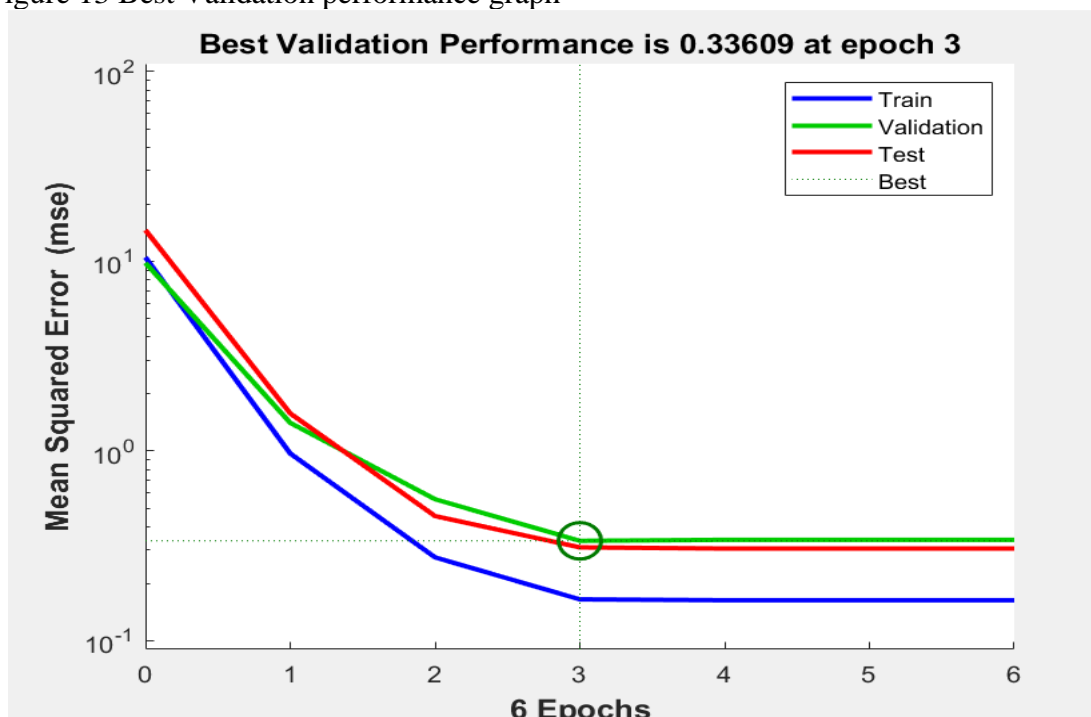
#### 4.3.2 Analysis of Residuals: -

The detailed analysis of residuals within the Artificial Neural Network (ANN) model, a comprehensive examination of the differences between predicted and actual construction costs was conducted to assess the model's performance and identify potential areas for improvement. Residual analysis involved scrutinizing the patterns, distribution, and behavior of residuals across the dataset. Various statistical techniques, such as plotting residual histograms, scatter plots of residuals against predicted values, were employed to elucidate the characteristics of residuals.

##### 4.3.2.1 Validation Performance: -

By visually inspecting these plots, stake holder gained insights into the presence of systematic biases, heteroscedasticity (unequal variance of residuals), or non-normality in residual distributions. Additionally, tests, such as the Validation performance and training state, were utilized to validate findings from graphical analysis. Overall, the detailed analysis of residuals played a pivotal role in validating the ANN model's performance, identifying areas for refinement, and enhancing its reliability for accurate cost estimation in construction projects.

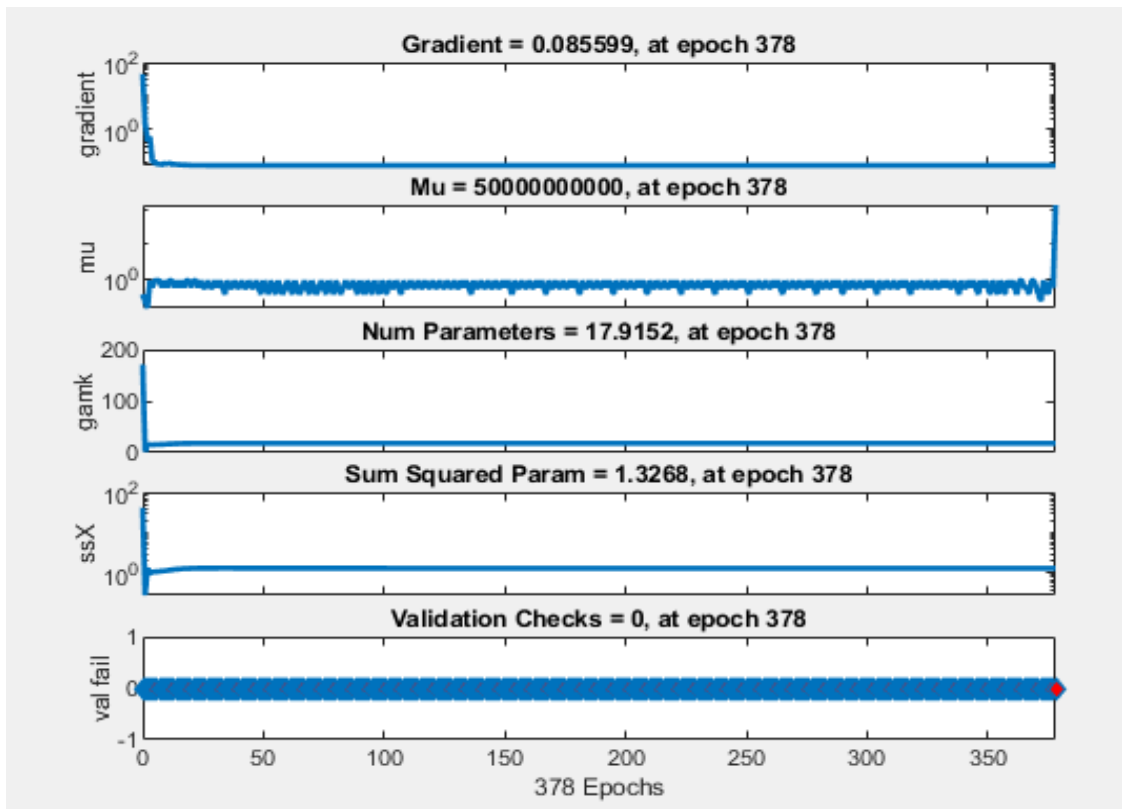
Figure 13 Best Validation performance graph



#### 4.3.2.2 Training State: -

In the detailed analysis of the performance graph within the Artificial Neural Network (ANN) model for construction cost prediction using MATLAB, a comprehensive examination of the model's performance across various training epochs was conducted. The performance graph depicted key performance metrics, such as Mean Squared Error (MSE) or Mean Absolute Error (MAE), plotted against the number of training epochs. By visually inspecting the performance graph, researchers gained insights into the convergence behavior of the ANN model during training. Fluctuations or trends in performance metrics over epochs provided valuable diagnostic information regarding the model's learning dynamics, convergence speed, and stability. Additionally, the performance graph facilitated the identification of optimal stopping criteria for training, helping to prevent overfitting or underfitting of the model. By iteratively analyzing the performance graph and adjusting training parameters, such as learning rate or batch size, we can optimize the ANN model's performance and enhanced its predictive accuracy for construction cost estimation. Overall, the detailed analysis of the performance graph played a crucial role in understanding the training dynamics of the ANN model, guiding parameter tuning efforts, and improving the reliability of cost predictions in construction projects.

Figure 14 training state graph



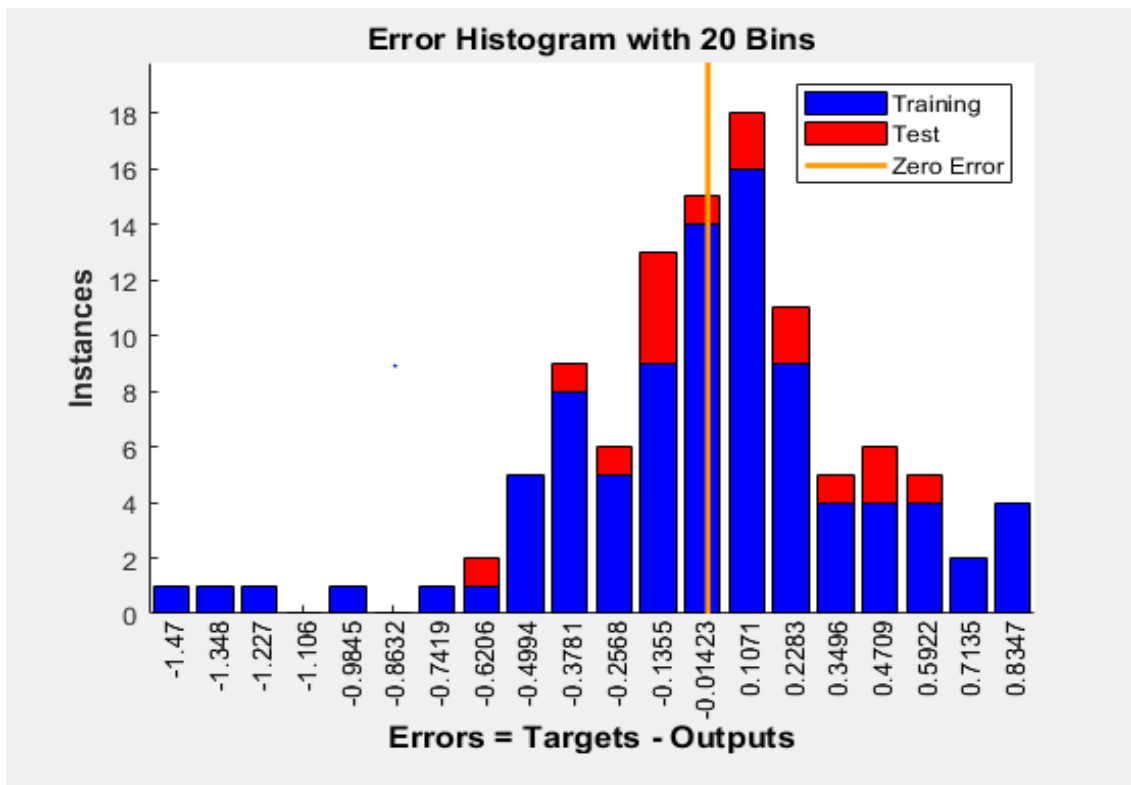
---

In the detailed analysis of the training state graph within the Artificial Neural Network (ANN) model, a comprehensive examination of the model's training dynamics and convergence behavior was conducted. The training state graph, generated through the plot train state function, provided visual insights into key training metrics and performance indicators throughout the training process. It depicted the progression of metrics such as training error, validation error, and performance gradient over successive iterations or epochs of training. By scrutinizing the training state graph, we could discern the model's learning trajectory, identifying phases of rapid learning, convergence, or potential issues such as overfitting or underfitting. Fluctuations or trends in training and validation errors offered valuable diagnostic information regarding the model's capacity to generalize to unseen data and its susceptibility to overfitting. Additionally, the performance gradient curve depicted the rate of change in performance metrics, aiding in the assessment of training stability and optimization progress. By iteratively analyzing the training state graph and adjusting training parameters such as learning rate or regularization strength, the ANN model's training process, enhancing its predictive accuracy and robustness for construction cost estimation. Overall, the detailed analysis of the training state graph played a pivotal role in understanding the training dynamics of the ANN model, guiding parameter tuning efforts, and improving its performance in real-world applications.

#### 4.3.2.3 Error Histogram: -

In the detailed analysis of the error histogram within the Artificial Neural Network (ANN) model, a comprehensive examination of the distribution and characteristics of prediction errors was conducted. The error histogram provided visual insights into the frequency and magnitude of errors between predicted and actual construction costs across the dataset. By scrutinizing the error histogram, we could assess the presence of systematic biases, heteroscedasticity (unequal variance of errors), or non-normality in the distribution of prediction errors. Deviations from a symmetric, bell-shaped distribution suggested potential issues such as model misspecification or outliers influencing the prediction accuracy. The error histogram served as a diagnostic tool to identify patterns or anomalies in prediction errors, guiding further analysis and model refinement efforts. By iteratively analyzing the error histogram and addressing underlying issues, researchers optimized the ANN model's performance and enhanced its reliability for accurate cost estimation in construction projects. Overall, the detailed analysis of the error histogram played a crucial role in validating the ANN model's predictive capabilities, identifying areas for improvement, and ensuring its suitability for real-world applications.

Figure 15 Error histogram



#### 4.3.2.4 Comparison with Actual Cost: -

The comparison between actual costs and predicted costs serves as critical evaluation metric in assessing the performance and accuracy of predictive models, particularly in domains such as construction project management. This comparative analysis enables the identification of any discrepancies or deviations between the predicted and actual costs, thereby facilitating informed decision-making and course correction where necessary. A favorable outcome entails minimal variance between the actual and predicted costs, signifying the model's proficiency in capturing and forecasting project expenses accurately. Consequently, a comprehensive examination of the factors contributing to the variance between actual and predicted costs is imperative, paving the way for model refinement and enhancement. Through iterative refinement and validation processes, predictive models can be iteratively optimized to deliver increasingly accurate cost forecasts, thereby empowering stakeholders with invaluable insights for effective project planning, budgeting, and management.

Table 18 Results of neural network model at testing phase

No.	Actual Cost (ETB)	Estimated Cost (ETB)	Absolute Error AE (Birr)	Absolute Percentage Error (%)
Project 1	40861726.80	40537706.99	-324019.81	-0.79%
Project 2	12430760.12	11739930.06	-690830.06	-5.56%
Project 3	21389929.56	20253434.71	-1136494.85	-5.31%
Project 4	60507026.97	61362655.46	855628.49	1.41%
Project 5	32482085.35	31863503.85	-618581.50	-1.90%
Project 6	8711089.01	8421728.92	-289360.09	-3.32%
Project 7	23534040.36	21592841.33	-1941199.03	-8.25%
Project 8	49998034.18	50141985.50	143951.32	0.29%
Project 9	63452535.38	64648041.11	1195505.73	1.88%
Project 10	53531748.70	54266558.06	734809.36	1.37%
Project 11	17148169.15	16570369.76	-577799.39	-3.37%
Project 12	14282372.52	13400239.20	-882133.32	-6.18%
Project 13	69321598.74	68697421.09	-624177.65	-0.90%
Project 14	108691060.25	107944901.53	-746158.72	-0.69%
Project 15	52124005.22	49727769.51	-2396235.71	-4.60%
Project 16	6640870.62	6388283.84	-252586.78	-3.80%
Project 17	26525774.27	26168420.63	-357353.64	-1.35%
Project 18	42221910.77	42257119.53	35208.76	0.08%
Project 19	57608606.39	57362753.65	-245852.74	-0.43%
Project 20	84447326.55	83007171.02	-1440155.53	-1.71%
Project 21	5484148.63	5100498.01	-383650.62	-7.00%
Project 22	11613276.63	11405120.21	-208156.42	-1.79%
Project 23	87743703.88	89141659.31	1397955.43	1.59%
Project 24	38650841.42	39482408.72	831567.30	2.15%
Project 25	31961140.53	32588912.51	627771.98	1.96%
Project 26	24418840.14	25078081.06	659240.92	2.70%

No.	Actual Cost (ETB)	Estimated Cost (ETB)	Absolute Error AE (Birr)	Absolute Percentage Error (%)
Project 27	14532331.09	14510797.89	-21533.19	-0.15%
Project 28	88792043.54	88260754.23	-531289.31	-0.60%
Project 29	8432040.49	7846313.09	-585727.40	-6.95%
Project 30	5641930.20	5243481.10	-398449.10	-7.06%
Project 31	31154647.43	31804520.30	649872.87	2.09%
Project 32	12988293.74	12204034.56	-784259.18	-6.04%
Project 33	12872737.33	12082521.04	-790216.29	-6.14%
Project 34	54208036.91	53769073.98	-438962.94	-0.81%
Project 35	15893199.19	15618683.94	-274515.26	-1.73%
Project 36	12650344.41	12491331.97	-159012.4427	-1.26%
Project 37	40007868.61	39145763.12	-862105.4941	-2.15%
Project 38	101957280.89	103501092.45	1543811.565	1.51%
Project 39	112775914.65	116529628.96	3753714.31	3.33%
Project 40	7630649.00	7489698.97	-140950.0346	-1.85%
Project 41	18367499.04	18300997.39	-66501.64259	-0.36%
Project 42	26609068.35	26467591.60	-141476.7521	-0.53%
Project 43	369550687.40	391727830.70	22177143.3	6.00%
Project 44	162826141.75	168113943.41	5287801.662	3.25%
Project 45	41826359.39	39865410.20	-1960949.195	-4.69%
Project 46	54912760.84	54368314.61	-544446.2294	-0.99%
Project 47	47230711.81	46766453.40	-464258.4085	-0.98%
Project 48	11872486.72	11784792.16	-87694.5594	-0.74%
Project 49	10282565.33	10083652.44	-198912.888	-1.93%
Project 50	5223894.16	4872702.59	-351191.569	-6.72%
Project 51	15930735.01	15835555.48	-95179.53105	-0.60%
Project 52	277309895.79	298696039.83	21386144.04	7.71%
Project 53	53940554.84	52864143.18	-1076411.658	-2.00%

No.	Actual Cost (ETB)	Estimated Cost (ETB)	Absolute Error AE (Birr)	Absolute Percentage Error (%)
Project 54	17306160.71	17406374.28	100213.5739	0.58%
Project 55	91217009.14	90985206.88	-231802.2556	-0.25%
Project 56	69352141.94	68675037.30	-677104.6426	-0.98%
Project 57	7167050.75	6800734.62	-366316.1267	-5.11%
Project 58	17736640.67	17619094.83	-117545.8413	-0.66%
Project 59	99514915.42	101836779.47	2321864.052	2.33%
Project 60	35678208.62	35098653.95	-579554.6746	-1.62%
Project 61	6608412.35	6416344.11	-192068.24	-2.91%
Project 62	16432680.12	16223766.44	-208913.6777	-1.27%
Project 63	205487827.96	211659396.50	6171568.545	3.00%
Project 64	105753773.82	110423330.19	4669556.371	4.42%
Project 65	22575567.04	22841458.85	265891.8141	1.18%
Project 66	456063149.60	484999938.67	28936789.07	6.34%
Project 67	22913317.66	23633807.92	720490.2587	3.14%
Project 68	52616121.57	51547957.48	-1068164.087	-2.03%
Project 69	202277080.65	208352895.96	6075815.305	3.00%
Project 70	122145608.76	125956050.19	3810441.429	3.12%
Project 71	7180899.49	7008724.90	-172174.5857	-2.40%
Project 72	109462873.05	110515037.16	1052164.113	0.96%
Project 73	15902596.57	15834565.97	-68030.60084	-0.43%
Project 74	148055283.35	156490372.50	8435089.154	5.70%
Project 75	67425693.35	67589396.16	163702.8054	0.24%
Project 76	46647616.60	45532899.50	-1114717.098	-2.39%
Project 77	151824732.18	155212879.68	3388147.496	2.23%
Project 78	55987406.25	59780119.76	3792713.505	6.77%
Project 79	10926574.45	10935408.51	8834.05558	0.08%
Project 80	83117571.40	82678331.64	-439239.7599	-0.53%

No.	Actual Cost (ETB)	Estimated Cost (ETB)	Absolute Error AE (Birr)	Absolute Percentage Error (%)
Project 81	264384973.97	275101090.91	10716116.94	4.05%
Project 82	14784204.84	14864978.53	80773.69297	0.55%
Project 83	15625135.83	15887138.67	262002.8423	1.68%
Project 84	229368185.00	237780337.74	8412152.744	3.67%
Project 85	19224472.30	19294144.04	69671.73514	0.36%
Project 86	38405669.71	37479671.79	-925997.921	-2.41%
Project 87	79653675.07	79712607.83	58932.7634	0.07%
Project 88	89890707.75	93224510.10	3333802.354	3.71%
Project 89	15230200.98	15026269.24	-203931.7462	-1.34%
Project 90	5743293.17	5537160.05	-206133.1192	-3.59%
Project 91	12104352.32	12144020.88	39668.55998	0.33%
Project 92	126904787.50	127567933.16	663145.658	0.52%
Project 93	148936564.80	155150509.14	6213944.34	4.17%
Project 94	60853936.20	61792720.58	938784.3796	1.54%
Project 95	15251049.73	15259423.78	8374.051162	0.05%
Project 96	15010243.68	14937493.09	-72750.58586	-0.48%
Project 97	65472907.16	65048607.44	-424299.723	-0.65%
Project 98	11450215.60	11265162.40	-185053.1972	-1.62%
Project 99	145950497.40	149907578.97	3957081.571	2.71%
Project 100	62757440.61	62286515.26	-470925.3465	-0.75%
Project 101	36100699.13	35424115.79	-676583.3371	-1.87%
Project 102	11293341.00	11170296.54	-123044.4581	-1.09%
Project 103	22515365.52	22866752.58	351387.0596	1.56%
Project 104	16235243.15	16222968.22	-12274.93367	-0.08%
Project 105	25986986.86	26742401.03	755414.1679	2.91%

---

## 5. CONCLUSION AND RECOMMENDATION

### 5.1 Conclusion

This study was undertaken with the primary objective of devising a novel technique for the early estimation of construction project costs within the Oromia Region. The overarching aim was to furnish stakeholders involved in such projects, including owners, contractors, and others, with comprehensive cost information during the nascent stages of project development, despite limited initial data availability. To achieve this, a meticulous series of steps and procedures were meticulously followed:

- ✚ Initial stages involved a multifaceted approach, encompassing questionnaire surveys, expert interviews, and exhaustive exploration of prior studies. Through these methods, a thorough identification of the Most Influential factors affecting building project costs was achieved. Fifteen key parameters emerged from this analysis, including but not limited to Types of building, area of typical floor, number of floors, Duration, Market Conditions, Risk Conditions, Safety conditions, Availability of Local Materials, Availability of Skilled Labor, Capacity of Contractors, Management Condition, Location, Floor Height, External Social Condition, and Complexity of the building.
- ✚ The study further involved the meticulous collection of historical data pertaining to building projects executed within the Oromia Region between 2009 and 2015. Data was sourced from various government ministries, contractors, and consultants. Following rigorous analysis, 105 projects were deemed suitable for inclusion, with this dataset being randomly divided into training, cross-validation, and testing sets.
- ✚ The subsequent phase focused on the development of an Artificial Neural Network (ANN) model, with MATLAB19b being selected as the preferred application due to its proven efficacy in previous research endeavors, coupled with its user-friendly interface. Through meticulous encoding and data entry into MS Excel spreadsheets, the training process for various models commenced.
- ✚ Among the plethora of models constructed, the single-layer feed-forward back-propagation (FFBP) model emerged as the optimal choice, characterized by one input layer with 15 neurons, one hidden layer comprising 5 neurons, and one output neuron. Employing the purelin transfer function and Momentums learning rate, which pertains to the Back propagation algorithm, this model exhibited exceptional accuracy, recording a performance rate of 94%.

- 
- ✚ Ensuring the validity and reliability of the model necessitated the implementation of numerous statistical performance measures, including Mean Absolute Error (MAE) and Mean Squared Error (MSE). The outcomes of these assessments were deemed acceptable and reliable, with the model showcasing an average percentage error of 8%.
  - ✚ Crucially, the culmination of this endeavor materialized in the creation of a user-friendly GUI, capable of delivering precise project cost estimates utilizing the identified fifteen parameters during the project's embryonic phases.
  - ✚ The study encounter challenges in inaccurately estimating costs in scenarios where certain projects exhibit disparities in factors like local material and labor availability. For instance, a project with identical design specifications (floor area, floor height and number of floor) but access to superior local materials, urban project location, low risk condition and high labor availability was incurring higher costs compared to a project operating in areas with limited local resources, Low skilled labor, remote project location and heightened risk conditions. These limitations underscore the necessity of considering a broader range of variables and contextual nuances in future analyses.

Moreover, the study's effectiveness is also influenced by the predefined ranges and constraints placed on input parameters, such as the predetermined area limits for typical floors. These constraints may inadvertently restrict the model's adaptability and accuracy, particularly in scenarios where projects deviate significantly from the predefined parameters. Despite these inherent limitations, it's imperative to acknowledge that this study represents a significant advancement in refining cost estimation practices within the construction sector of the Oromia Region. By providing stakeholders with a robust tool for informed decision-making, the study lays a foundational framework for future research endeavors aimed at addressing and mitigating these inherent challenges. Through continued refinement and integration of additional variables, future iterations of this model hold promise for further enhancing the accuracy and reliability of cost estimation in construction projects.

---

## 5.2 Recommendations

The current study showed very promising results in predicting the cost of building projects, and this approach will continue to make impressive gains especially in civil engineering field. However, some recommendations should be presented for decision makers in the construction sector and future studies to support the findings of this study;

- ✚ All construction parties are encouraged to be more aware about cost estimation development and pay more attention for using this developed technique in estimation process.
- ✚ It is recommended that the construction firm, Government and engineering associations starts collecting and storing project cost data in order to create and establish a database for executed projects cost database that can be used as (training) data source for researchers to develop cost estimation process.
- ✚ For future studies, it is recommended to obtain more training data from newly projects and add them to the training data. This will improve the training process and produce more input choices.

Finally, Cost estimates in neural network models are related to the input costs that the model was built upon, so changing input will increase the estimate error or make the model unusable. Therefore, it is recommended to link the model with the inputs considered.

---

## 6 References

- Abd, A. M., Jasim, N. A., & Naseef, F. S. (2019). Predicting the final cost of Iraqi construction project using Artificial Neural Network (ANN). *Indian J. Scie. Technol.*, 12,, 18.
- Abdullah M. Alsugair, K. S.-G. (2023). Artificial Neural Network Model to Predict Final Construction. *Appl. Sci.* 2023, 13, 8078. <https://doi.org/10.3390/app13148078>, 1-19.
- Abdulmohsen S. Almohsen, N. M. (2023). Integrated artificial and deep neural networks with time series to predict the ratio of the low bid to owner estimate. *Engineering, Construction and Architectural Management Emerald Publishing Limited*.
- Abhijit N. Bhirud, V. R. (2017). pre-design stage construction cost prediction of building projects using artificial neural network.
- Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. A Textbook. Springer. <https://doi.org/10.1007/978-3-319-94463-0>.
- Alaloul, W. S. (2018). An Artificial neural networks (ANN) model for evaluating construction project performance based on coordination factors. *Cogent Engineering*, 5(1), 1-18.
- Alcineide, P., Gean, S., Luiz, M. F. M., Felipe, C. A., & Débora, de G. S. . (2021). Cost forecasting of public construction projects using multilayer perceptron artificial neural networks. *Ingenieria e Investigacion*, 41(3), 1–11.
- Alqahtani, A., & Whyte, A. (2013). Artificial neural networks incorporating cost significant items towards enhancing estimation for (life-cycle) costing of construction projects. *Australas J. Constr. Econ. Build.* 2013, 313, 51–64.
- Alqahtani, A., & Whyte, A. (2016). Estimation of life-cycle costs of buildings: regression vs artificial neural network. *Built Environment Project and Asset Management*, 6(1), 30–43.
- Al-saadi, A. M. (2017). Estimating the Optimum Duration of Road Projects Using Neural Network Model. *International Journal of Engineering and Technology (IJET) Vol 9 No 5*, 3458.
- Alshahethi, A. A. (2018). Estimating the final cost of construction project using ANN. *Engineering, Business*.
- Alshemosi, A. M. (2017). Cost estimation process for construction residential projects by using multifactor linear regression technique. *International Journal of Science and Research*. 6(6), 151-156.
- Ar. Ankita Gupta, A. P. (2022). Artificial Neural Network Analysis for Cost Estimation of Building Projects in India. *5th International Conference of Contemporary Affairs in Architecture and Urbanism (ICCAUA-2022) 11-13 May* , 193-205.
- Aswed, G. (2016). Productivity Estimating Model for Bricklayer In Construction Projects Using Artificial Neural Networks. *Al-qadisiya Journal of Engineering Sciences*, 9(2), 183-199.
- Badawy, M. A. (2022). Identifying the risk factors affecting the overall cost risk in residential projects at the early stage. *Ain Shams Engineering Journal*, Vol. 13 No. 2,.
- Barros, L. B. (2018). *development of a more accurate estimation technique for construction highway projects using ANN*. Brazil.
- Beltman, a. (2021). *Predicting construction costs in the program phase of the construction process: a machine learning approach*. University of Twente.
- Bhosale, A. & Konnur, B. A., (2019). Use of Artificial Neural Network in Construction Management. *International Research Journal of Engineering and Technology (IRJET)*, 06(05), 5905-5908.
- Brownlee, J. (2020). Deep Learning Performance. *Retrieved from Machine Learning Mastery*:.  
Chakraverty, S., Deepti, ., Sahoo, M., Nisha, ., & Mahato, R. . (2019). Concepts of Soft Computing Fuzzy and ANN with Programming.
- Chandanshive, V. &. (2019). Estimation of building construction cost using artificial neural networks. *Journal of Soft Computing in Civil Engineering*. 3(1), 91-107.
- Chandanshive, V. B., & Kambekar, A. R. (2019). Estimation of Building Construction Cost Using Artificial Neural Networks. *Journal of Soft Computing in Civil Engineering*, 3(1), 91–107.

- D.A. Clevert, T. U. (2016). Fast and Accurate Deep Network. *Proceedings of International Conference on Learning Representations*, 1-14, .
- Elevado, K. J. T., Galupino, J. G. & Gallardo, R. S.,. (2018). Artificial Neural Network (ANN) modelling of concrete mixed with waste ceramic tiles and fly ash. *International Journal of GEOMATE*, 15(51),, 154-159.
- Elmou salami, H. H. (2020). Artificial Intelligence and Parametric Construction Cost Estimate Modeling. *Journal of Construction Engineering and Management*, 146(1), 1943-7862.0001678.
- Elsawy, I., Hosny, H., & Razek, M. (2011). A neural network model for construction projects site overhead cost estimating in Egypt. *Int. J. Comput. Sci. Issues* 2011, 8, , 273–283.
- Fachrurrazi, Saiful, H., & Mubarak, T. (2016). Neural network for the standard unit price of the building area. *Sustain. Civ. Eng. Struct. Constr. Mater.* 2016, 171, , 282–293.
- Fatima, A., Prasad, B. S. R. K. & Sekhar, S. T.,. (2017). A Comprehensive Study of Dispute Resolution using Artificial Neural Network in Build Operate and Transfer (BOT) Project. *International Journal of Applied Engineering Research*, 12(24),, 14984-14988.
- Ghadiminia, S. S. (2023). A construction cost estimation framework using DNN and validation unit. *Building Research & Information*, DOI:, 1-9.
- Glinskiy, V., Serga, L., Khvan, M. & Zaykov, K.,. (2016). Fuzzy Neural Networks in the Assessment of Environmental Safety. *Novosibirsk, Procedia CIRP*,, 615-619.
- Goa, H. (2016). Safety Evaluation on Building Construction Based on Hopfield Neural Network. *Jinan, 5th International Conference on Civil, Architectural and Hydraulic Engineering*.
- Golizadeh, H. S. (2016). Automated tool for predicting duration of construction activities in tropical countries. *KSCE Journal of Civil Engineering*, 20(1), 12-22.
- Golnaraghi, S., Zangenehmadar, Z., Moselhi, O. & Alkass, S., . (2019). Application of Artificial Neural Network(s) in Predicting Formwork Labour Productivity. *Advances in Civil Engineering*, 1-11.
- Günaydin, H. M., & Doğan, S. (2004). A neural network approach for early cost estimation of structural systems of buildings. *International Journal of Project Management*, 22(7), 595–602.
- Günaydin, H., & Dogan, S. (2022). A neural network approach for early cost estimation of structural systems of buildings. *Int. J. Proj. Manag.* 2022, 22,, 595- 602.
- Haykin, S. (1999). *Neural networks: a comprehensive foundation*. Prentice Hall. Prentice Hall.
- Heravi, G. &. (2015). Applying Artificial Neural Networks for Measuring and Predicting Construction-Labor Productivity. *Journal of Construction Engineering and Management*, 141(10), 04015032.
- Jiang, Q. (2019). Estimation of construction project building cost by back-propagation neural network. *J. Eng. Des. Technol.* , 18,601–609. [CrossRef].
- Jiang, Z. &. (2016). The selection of firms based on productivity: Different roles of entry and overhead cost. *Economic Modelling*, Volume 54,, 537-544.
- Jinisha, R., & Jothi, R. (2019). Construction Prediction of Cost Overruns. *International Journal of Innovative Research in Science, Engineering and Technology-IJIRSET*, 8(3)-2695–2699.
- Jose, C. M. (2017). Prediction of Cost of Quality Using Artificial Neural Network In Construction Projects. *International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 03, Issue 04;*, 54.
- Juszczuk, M. (2019). *Modelling Construction Site Cost Index Based on Neural Network Ensembles*. Cracow University of Technology.
- Juszczuk, M., Leśniak, A., & Zima, K. (2018). ANN Based Approach for Estimation of Construction Costs of Sports Fields. *Complexity*.
- KabirYaqubi, M. (2019). The Automated Cost Estimation in Construction. *International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7* .
- Karim, A., & Bubshait, A. (2013). Karim, A., Cost estimation using artificial neural networks in construction projects: a case study of Saudi Arabia. [9] *Karim, A., & Bubshait, A.*

- (2013). *Cost estimation using arti International Journal of Engineering Business Management*, 5(1), 1-9.
- Kaviya, K. &. (2019). Prediction of compressive strength of high-performance concrete using Artificial Neural Network (ANN) models. *International Research Journal of Engineering and Technology*, 6,, 1378-1387.
- Khalaf, T. Z. (2020). Particle swarm optimization-based approach for estimation of costs and duration of construction projects. *Civil Engineering Journal*. 6(2),, 384-401.
- Kim, S., Choi, C., Shahandashti, M., & Ryu, K. (2022). Improving Accuracy in Predicting City-Level Construction Cost Indices by Combining Linear ARIMA and Nonlinear ANNs. *J. Manag. Eng.*, 38.
- Kulkarni, P. S. (2017). Artificial Neural Networks for Construction Management. *Journal of Soft Computing in Civil Engineering*, 1(2),, 70-88.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning . *Nature*, 521(7553), 436–444.
- Lee, D., Kim, S. & Kim, S.,. (2016). Development of Hybrid Model for Estimating Construction Waste for Multifamily Residential Buildings Using Artificial Neural Networks and Ant Colony Optimization. *Sustainability*, 8(9), , 870-883.
- Lee, J.G.; Lee, H.S.; Park, M.; Seo, J.O. . (2021). Early-stage cost estimation model for power generation project with limited historical data. *CrossRef*, 2599–2614.
- Len Holm, J. E. (2021). *Construction Cost Estimating*. *Routledge*. Retrieved from <https://doi.org/https://doi.org/10.1201/9781003023494>
- Lesniak, A. & Juszczuk, M.,. (2018). Prediction of site overhead costs with the use of artificial neural network-based model. *Archives of Civil and Mechanical Engineering*, 18(3),, 973-982.
- M. Juszczuk and A. Le´ sniak,. (2019). Modelling construction site cost index based on neural network ensembles. *Symmetry*, vol. 11, no. 3,, 411.
- M. Juszczuk, A. Le´ sniak, and K. Zima,. (2018). ANN based approach for estimation of construction costs of sports fields. [46] *M. Juszczuk, A. Le´ sniak, and K. Zima, “ANN based approComplexity*, vol. 2018, Article ID 7952434, 11.
- Magdum, S. K. (2017). Construction Cost Prediction Using Neural Networks. *Journal On Soft Computing*, October 2017, Volume: 08, Issue: 01, 1549.
- Marker, A. (2017). *The Ultimate Guide to Project Cost Estimating*. Retrieved from *Smartsheet*.
- Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H. (2019). An artificial neural network approach for cost estimation of engineering services. *International Journal of Construction Management*, 1–14.
- Mirahadi, F. & Zayed, T.,. (2016). Simulation-based construction productivity forecast using Neural-Network-Driven Fuzzy Reasoning. *Automation in Construction*, Volume 65,, 102-115.
- Mistry. Jignesh. Mukeshchandra, g. K. (2018). Implementation of Artificial Neural Network for Recognition of Factors Influencing Labor Production Rates for Concreting Activities. *A Review” IOSR Journal of Mechanical and Civil Engineering (IOSR- JM*, 50.
- Mohaghegh, S. (2000). *Virtual Intelligence And Its Applications In Petroleum Engineering*. Available at: <http://www.pe.wvu.edu/research/part1.htm>.
- Najafi, A., & Tiong, R. (2015). Construction productivity and cost estimation using artificial neural networks. 224-236.
- Nani, G., Mensah, I. & Kumi, T. A.,. (2017). Duration estimation model for bridge construction projects in Ghana. *Journal of Engineering, Design and Technology*, 15(6),, 754-777.
- Nicholas, J. M., & Steyn, H. . (2017). *Project Management for Engineering, Business and Technology*. *Routledge*.
- P.S.Kulkarni. (2017). Cost evaluation of construction using ANN. *volume 5/ issue 11/18.051154IJMTST*, 1-15.
- Pal, B. (2018). Cost Estimation Model (CEM) of Buildings by ANN (Artificial Neural Networks). *A Review” International Advanced Research Journal in Science, Engineering and Technology ISO 3297:2007 Certified Vol. 5, Issue 2*, 26.
- Pal, B. (2018). Cost Estimation Model (CEM) of Buildings by ANN (Artificial Neural Networks) . *International Advanced Research Journal in Science, Engineering and*

- Pham, T., Le-Hong, T., & Tran, X. (2021). Efficient estimation and optimization of building costs using machine learning. *Int. J. Constr. Manag.* 2021, 21,, 1–13.
- Polat, G. (2012). ANN approach to determine cost contingency in international construction project. *J. Appl. Manag. Investig*, 1, 195–201.
- Principe, J. e. (2010). NeuroSolution Help,. *NeuroDimension, Inc.*
- Raghd, A., Hanafy, H. & Tarek, A.,. (2018). Developing an ANN model to suggests the optimum typology for effectively managing construction projects. *International Journal of Advanced Research*, 6(7), 93-97.
- Ramos, D. (2017). Construction Cost Estimating. Retrieved from Smartsheet. Retrieved from <https://www.smartsheet.com/construction-cost-estimating>
- Saber, S. A. (2023). *Investigation of Factors Impacting Construction Cost Estimate to Develop ConstructionDriven Artificial Neural Network (ANN)*. ARIZONA STATE UNIVERSITY.
- Sharma, S. (2017). Activation Functions in Neural Networks. Retrieved from *Towards Data Science*.
- Sharmila, S., Lekha, G. & Kaushik, S.,. (2016). Cost and Time Effective Prediction of Soil Characteristics using ANN Model. *International Journal of Innovative Research in Science, Engineering and Technology*, 5(3),, 3829-3834.
- Shehatto, O. (2013). Cost Estimation for Building Construction Projects in Oromia Region Using Artificial Neural Network (ANN).
- ShehuWaziri, B. (2017). Artificial Neural Networks in Construction Engineering and Management. *International Journal of Architecture, Engineering and Construction Vol 6, No 1,*, 50-60.
- Shrestha, K. K. (2016). A Contingency Cost Estimation System for Road Maintenance Contracts. *Procedia Engineering, Volume 145,*, 128-135.
- Sitthikorn Sitthikankun, D. R. (2021). Construction Cost Estimation for Government Building Using Artificial Neural Network Technique. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 1-12.
- Squeira, I. (1999). *Neural Network-Based Cost Estimating*. Montreal, QC, Canada,: Department of Building, Civil and Environmental Engineering, Concordia University.
- Tijanić, K. C.-u. (2020). Cost estimation in road construction using artificial neural network. *Neural Computing and Applications*, 32(13), 9343–9355.
- Walia, A. S. (2017). Activation functions and it's types - Which is better? . Retrieved from *Towards Data Science*:.
- Wang, R. A. (2022). Assessing effects of economic factors on construction cost estimation using deep neural networks. *Automation in Construction*, 134.
- Waziri, B. S. (2017). Artificial Neural Networks in Construction Engineering and Management. *International Journal of Architecture, Engineering and Construction Vol 6, No 1,*, 50-60
- Wilmot, C., & Mei, B. (2005). Neural network modeling of highway construction costs. *J. Constr. Eng. Manag.* 2005, 131, , 765–771.
- Yadav, R., Vyas, V., Vyas, M. & Agrawal, S.,. (2016). Cost Estimation Model (Cem) for Residential Building using Artificial Neural Network. *International Journal of Engineering Research & Technology (IJERT)*, 5(1),, 430-432.
- Yuanyuan, T., & Libiao, B. (2022). Modeling for project portfolio benefit prediction via a GA-BP neural network. *Technol. Forecast. Soc. Chang.*, 183, 121939.
- Yun, D. P. (2023). Construction Cost Prediction Using Deep Learning with BIM. *Appl. Sci.* 2023, 13, 7207. <https://doi.org/10.3390/app13127207>, 1-14.
- Yun, S. (2022). Performance Analysis of Construction Cost Prediction Using Neural Network for Multioutput Regression. *Appl. Sci.* 2022, 12, 9592. <https://doi.org/10.3390/app12199592>, 1-17.

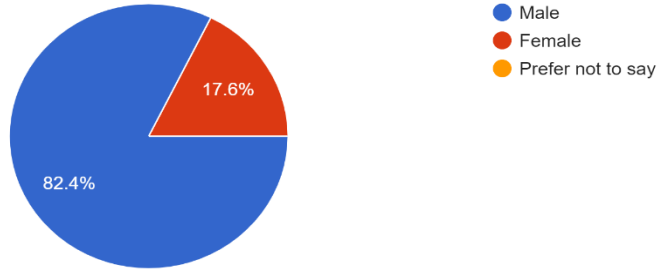
---

## APPENDECIS

### Appendix 1: Response of Sample Survey on Influential Factor on Construction Cost Estimation Section I: Demography

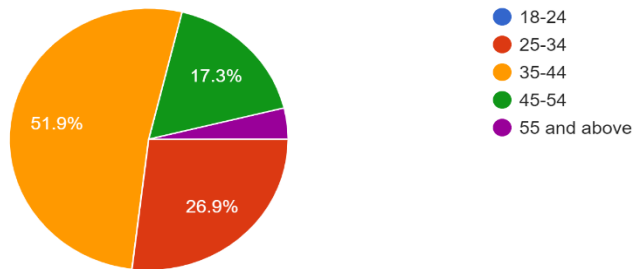
Gender

51 responses



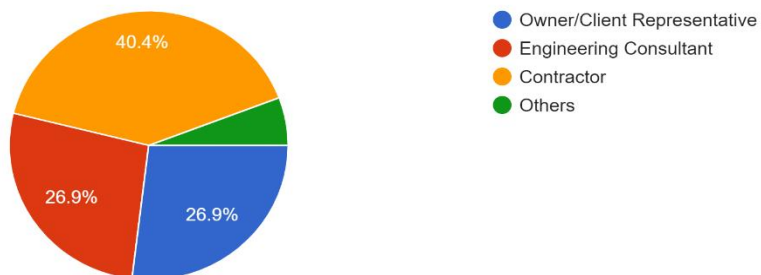
Age

52 responses



You are Representing?

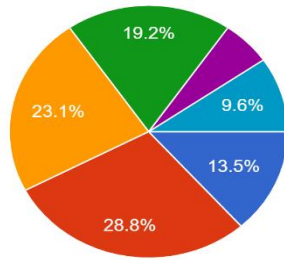
52 responses



---

Your current role regarding construction is?

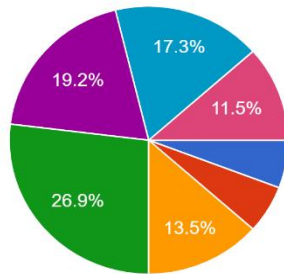
52 responses



- Owner/General Manager
- Construction manager
- Resident Engineer
- Site Engineer
- Quantity surveyor
- Others

Your years of industry experience is?

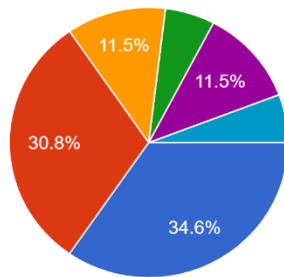
52 responses



- 0-2 years
- 3-5 years
- 6-8 years
- 9-11 years
- 12-15 years
- 15-20 years
- Over 20 years

How much do you rate the usage of advance technology in construction cost estimation? (such as ANN: Artificial Neural Network)

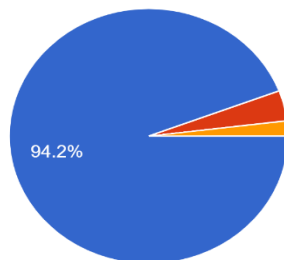
52 responses



- Not sure
- 0-10%
- 11%-25%
- 26%-50%
- 51%-75%
- 76%-100%

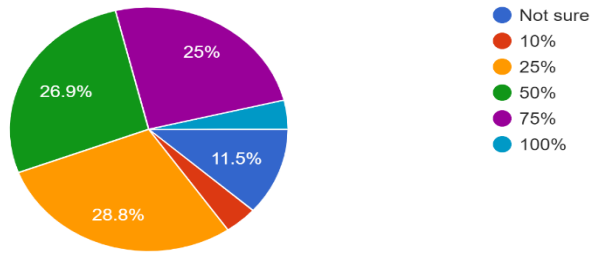
Do you think inaccurate estimation is one of the reasons for cost overrun in construction projects?

52 responses



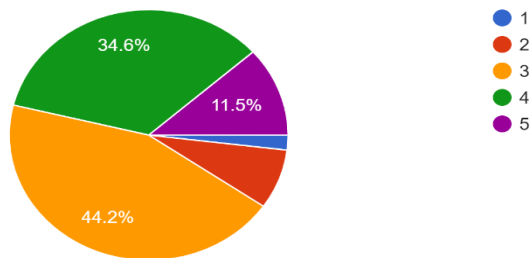
- Yes
  - No
  - Not Sure
-

How common do projects face cost overrun, or overbudget issues due to inaccurate estimation?  
52 responses

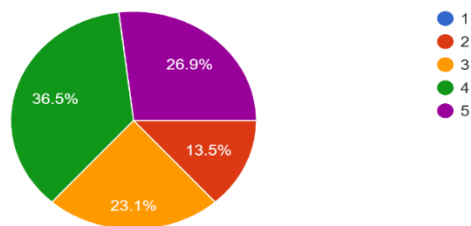


## Section II: Influential factors on construction cost estimation

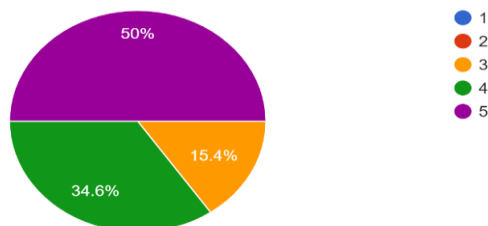
Location of the Project (Region, Zone, woreda) degree of effect on construction cost overrun 1=very low to none, 2= low, 3= moderate, 4=high, 5= very high  
52 responses



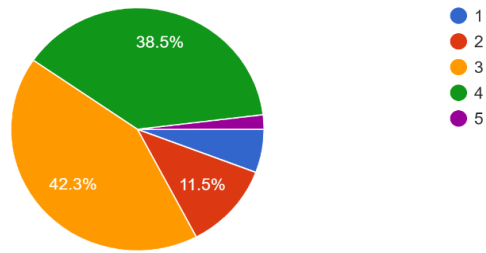
External Social Condition (War, Disease,) degree of effect on construction cost overrun 1=very low to none, 2= low, 3= moderate, 4=high, 5=very high  
52 responses



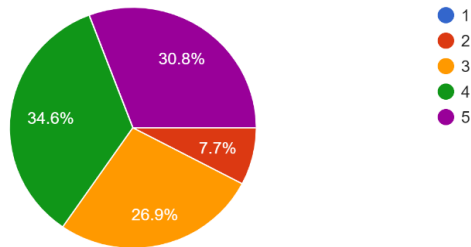
Market Condition (inflation, Local Currencies....) degree of effect on construction cost overrun, 1=very low to none, 2= low, 3= moderate, 4=high, 5= very high  
52 responses



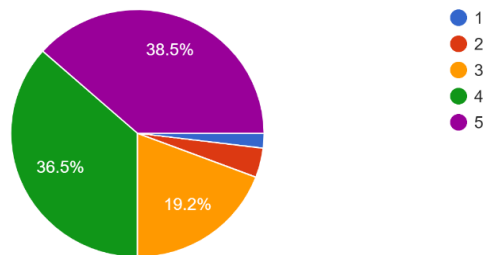
Safety Conditions(fire, electric, hazardous, security, emergency preparedness,...) degree of effect on construction cost overrun, 1=very low to none, 2= low, 3= moderate, 4=high, 5= very high  
52 responses



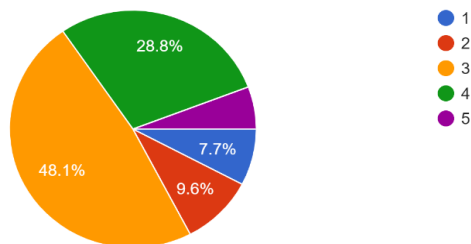
Availability of Skilled labor/Crew degree of effect on construction cost overrun, 1=very low to none, 2= low, 3= moderate, 4=high, 5= very high  
52 responses



Availability of Local Materials (Sand, Stone, Wood, Selected material,...) degree of effect on construction cost overrun, 1=very low to none, 2= low, 3= moderate, 4=high, 5= very high  
52 responses



Risk Condition (High, Medium, Low) degree of effect on construction cost overrun, 1=very low to none, 2= low, 3= moderate, 4=high, 5= very high  
52 responses



## Appendix 2: Response of Sample Survey on Selected Influential Factor VS Project Cost

project	Project Types	Market condition	Duration	Floor Area	Availability of skilled labor	Availability of Local Material	Management condition	Complexity of Building	Location	Risk Condition	Floor Height	Number of floors	Safety Conditions	External social condition	Capacity of Contractor	Cost
1	7	1	3	5	2	3	4	2	3	3	1	1	3	3	1	40861726.80
2	8	1	3	4	2	3	4	1	3	2	1	1	3	3	1	12430760.12
3	5	1	3	8	2	3	3	1	3	2	3	1	3	3	1	21389929.56
4	8	1	3	9	2	3	4	2	3	2	1	2	3	3	1	60507026.97
5	5	1	3	8	2	2	2	1	2	2	3	1	2	2	1	32482085.35
6	1	3	3	1	4	3	3	1	2	1	1	1	2	3	5	8711089.01
7	7	2	4	2	5	5	2	2	3	4	1	3	5	3	4	23534040.36
8	6	2	4	3	2	2	1	3	1	3	1	2	1	2	3	49998034.18
9	8	1	3	9	2	2	2	2	2	2	1	2	2	2	1	63452535.38
10	7	1	3	5	2	2	2	2	2	2	1	1	2	2	1	53531748.70
11	8	1	3	4	2	2	2	1	2	2	1	1	2	2	1	17148169.15
12	8	1	3	4	4	4	4	2	4	2	1	1	4	1	1	14282372.52
13	6	1	4	5	3	4	4	2	4	3	3	1	4	1	1	69321598.74
14	1	1	3	6	5	3	5	3	4	2	1	4	3	3	1	108691060.25
15	5	1	4	7	3	2	3	2	3	4	5	1	1	1	2	52124005.22
16	8	1	1	3	3	5	1	1	1	4	1	1	4	3	5	6640870.62
17	5	1	3	8	4	4	4	1	4	2	3	1	4	1	1	26525774.27
18	7	1	3	5	4	4	4	2	4	2	1	1	4	1	1	42221910.77
19	8	1	3	9	4	4	4	2	4	2	1	2	4	1	1	57608606.39
20	4	1	3	9	4	3	4	3	4	2	1	5	4	1	1	84447326.55
21	2	1	3	4	2	3	4	2	2	3	1	1	3	1	3	5484148.63
22	8	1	2	4	5	2	3	1	4	3	1	1	3	2	4	11613276.63
23	7	1	4	8	5	4	2	3	4	1	1	2	5	3	1	87743703.88
24	5	1	2	8	3	4	1	1	1	1	6	1	2	4	3	38650841.42
25	1	1	2	4	2	3	3	1	2	3	1	1	3	3	5	31961140.53
26	2	1	2	3	1	2	2	2	1	3	1	1	2	1	5	24418840.14
27	3	2	1	1	3	4	3	1	3	2	1	1	3	3	4	14532331.09
28	4	1	3	5	5	4	4	3	2	3	1	2	1	2	3	88792043.54
29	2	1	3	4	2	3	3	2	2	2	1	3	3	2	2	8432040.49
30	2	1	3	4	2	3	2	1	2	2	1	1	2	2	3	5641930.20
31	5	1	2	8	3	3	3	2	3	2	6	1	2	2	4	31154647.43
32	5	3	3	8	4	4	4	3	1	3	2	1	3	2	3	12988293.74
33	3	3	3	3	4	4	4	2	3	3	1	5	3	2	4	12872737.33
34	5	1	4	7	2	3	3	3	2	3	6	1	4	2	3	54208036.91
35	6	2	2	3	4	3	3	2	3	2	1	1	3	3	3	15893199.19
36	7	2	1	2	1	2	3	1	1	4	1	1	3	3	2	12650344.41
37	8	3	3	5	3	2	4	3	3	3	1	2	2	2	4	40007868.61
38	6	1	4	7	2	3	4	3	3	2	2	2	3	3	3	101925280.89
39	3	2	4	5	1	4	3	2	3	2	1	5	3	2	3	112775914.65
40	8	1	2	3	3	1	3	1	2	2	1	1	2	2	5	7630649.00
41	8	1	2	3	4	5	4	2	3	2	2	1	1	2	5	18367499.04
42	8	1	2	4	2	3	3	2	2	3	3	1	3	3	4	26609068.35
43	6	1	4	8	5	4	3	3	3	2	1	5	4	3	1	369550687.40
44	3	2	4	7	1	5	3	2	3	2	1	5	3	2	3	162826141.75
45	3	2	4	8	1	5	1	1	3	2	2	1	1	2	2	41826359.39
46	5	1	3	6	3	2	4	3	4	2	4	1	2	4	3	54912760.84
47	4	1	3	5	3	4	4	3	3	2	1	2	5	3	3	47230711.81
48	5	1	2	3	2	3	3	1	2	3	3	1	3	2	5	11872486.72
49	4	1	1	1	4	3	3	1	3	2	2	1	2	2	4	10282565.33
50	3	2	1	2	2	1	3	1	2	3	1	1	2	3	5	5223894.16
51	4	1	2	4	3	2	3	2	3	3	1	1	3	3	4	15930735.01
52	2	1	4	7	3	2	4	3	3	2	1	4	3	1	3	277309895.79
53	2	1	3	5	2	3	3	3	2	3	1	2	1	2	3	53940554.84
54	1	2	2	3	1	2	2	2	1	4	1	1	4	3	4	17306160.71
55	2	1	4	8	1	3	2	3	1	3	1	3	2	2	3	91217009.14
56	3	1	3	6	5	4	5	3	3	3	1	2	3	4	4	69352141.94
57	4	1	1	2	4	3	4	1	4	2	1	1	3	3	4	7167050.75
58	5	3	2	3	2	3	3	2	2	3	3	1	2	3	3	17736640.67
59	6	1	4	7	3	2	4	3	3	2	2	2	2	2	2	99514915.42
60	3	1	3	5	1	2	2	3	1	3	3	1	1	2	2	35678208.62
61	4	1	1	1	4	3	5	1	3	2	1	1	4	2	3	6608412.35
62	2	1	2	4	2	3	3	1	2	3	1	1	2	3	3	16432680.12
63	1	1	4	9	2	3	3	3	2	2	1	4	3	3	1	205487827.96
64	1	2	3	6	5	4	4	3	3	2	1	2	3	2	3	105753773.82
65	5	2	2	3	3	2	4	2	4	2	2	1	2	2	3	22575567.04
66	4	1	4	8	3	4	4	3	3	2	2	5	2	2	1	456063149.60
67	4	1	1	2	4	3	5	1	3	2	1	1	1	2	2	22913317.66
68	2	1	2	4	1	3	2	1	1	4	1	1	2	1	3	52616121.57
69	2	1	4	7	3	2	4	3	3	2	1	4	2	3	1	202277080.65
70	5	1	3	5	2	3	3	3	2	3	1	3	2	3	1	122145608.76
71	4	1	1	1	2	2	3	1	2	3	1	1	3	1	2	7180899.49
72	7	1	4	7	5	4	4	3	3	2	1	3	3	1	1	109462873.05
73	2	2	2	3	4	3	5	1	4	1	1	1	4	1	2	15902596.57
74	8	2	3	6	2	3	3	3	2	3	1	1	3	2	1	148055283.35
75	3	1	3	5	3	4	4	2	3	2	1	2	3	2	1	67425693.35
76	2	1	3	5	1	3	2	2	1	4	1	2	5	2	1	46647616.60
77	1	3	4	8	1	3	2	3	1	4	1	4	2	3	1	151824732.18
78	5	1	2	4	3	4	4	1	3	2	2	6	1	2	3	55987406.25
79	5	1	1	2	4	3	5	1	4	1	1	1	3	3	1	10926574.45
80	4	1	3	6	5	4	4	2	3	2	1	3	3	4	1	83117571.40
81	3	2	4	9	2	5	3	3	2	3	1	5	5	2	1	264384973.97
82	3	1	2	3	2	3	3	1	2	3	1	1	4	2	2	14784204.84
83	5	1	1	2	2	4	3	1	2	3	4	1	2	2	3	15625135.83
84	4	1	4	7	4	3	5	3	3	2	1	4	3	1	1	229368185.00
85	2	2	2	4	3	2	4	1	4	1	1	1	3	3	1	19224472.30
86	4	2	3	5	5	4	4	2	3	2	1	2	4	3	1	38405669.71
87	4	1	3	5	2	3	3	2	2	3	1	3	2	2	1	79653675.07
88	5	1	3	6	2	2	3	2	2	3	3	1	2	2	2	89890707.75
89	6	2	2	3	4	3	5	1	3	3	1	1	1	3	2	15230290.98
90	7	1	1	1	3	4	3	1	4	1	1	1	5	3	3	5743293.17
91	8	2	2	4	3	3	3	1	3	2	1	1	2	1	3	12104352.32
92	6	1	4	8	4	3	5	3	3	3	1	3	2	2	1	126904787.50
93	5	1	4	9	2	2	3	3	2	3	3	1	3	2	1	148936564.80
94	3	1	3	5	2	1	3	2	2	3	2	2	3	3	1	60853936.20
95	2	2	1	2	3	4	4	1	3	2	1	1	2	3	1	15251049.73
96	1	2	2	3	4	3	5	1	3	2	1	1	2	4	3	15010243.68
97	5	2	4	7	2	3	3	3	2	3	4	1	3	2	3	65472907.16
98	3	1	2	3	2	2	3	1	2	3	1	1	3	2	5	11450215.60
99	4	1	4	7	1	3	2	2	1	4	1	4	4	3	1	145950497.40
100	2	1	3	6	5	4	4	3	4	1	1	2	4	3	1	62757440.61
101	1	2	3	5	3	3	4	2	3	2	1	2	3	4	1	36100699.13
102	5	2	1	1	4	3	3	1	3	2	6	1	3	2	3	11293341.00
103	3	1	2	4	3	2	4	1	4	1	1	1	4			

## Appendix 3A: MATLAB code to define network architecture

```
Command Window
>> % my data has input features in the first columns and output/target values in the last column
% Split my data into input features and target values
X = burhan(:, 1:15); % Input features
Y = burhan(:,end); % Target values
% I want to normalize my input features
X = normalize(X);
% my neural network architecture
hiddenLayerSize = 15;
net = feedforwardnet(hiddenLayerSize);
% Train the neural network
net = train(net, X', Y');
% Display the network architecture
view(net);

fx >> |
```

## Appendix 3B: MATLAB code for Best Hidden Layer Size

```
Command Window
>> % Load your dataset and preprocess it as needed
% x2, Y2, x3, Y3 represent your training and testing data

% Define the range of values for the number of hidden layers and neurons
hiddenLayerSizes = [5, 10, 15, 20]; % Example: Test with 5, 10, 15, and 20 neurons
numHiddenLayers = [1, 2]; % Example: Test with 1 and 2 hidden layers

% Initialize variables to store results
bestAccuracy = 0;
bestHiddenLayers = 0;
bestNeurons = 0;

% Iterate over different configurations
for i = 1:length(numHiddenLayers)
    for j = 1:length(hiddenLayerSizes)
        % Create and train the neural network with current configuration
        net = feedforwardnet(repmat(hiddenLayerSizes(j), 1, numHiddenLayers(i)));
        net = train(net, x2, Y2);

        % Predict using the trained network
        Y_pred = net(X3);

        % Calculate accuracy (or any other performance metric)
        accuracy = calculateAccuracy(Y_pred, Y3);

        % Check if current configuration gives better accuracy
        if accuracy > bestAccuracy
            bestAccuracy = accuracy;
            bestHiddenLayers = numHiddenLayers(i);
            bestNeurons = hiddenLayerSizes(j);
        end
    end
end
fx end

Command Window
% Display the best configuration
fprintf('Best Number of Hidden Layers: %d\n', bestHiddenLayers);
fprintf('Best Number of Neurons per Hidden Layer: %d\n', bestNeurons);
fprintf('Best Accuracy: %.2f%%\n', bestAccuracy * 100);
Unrecognized function or variable 'Y2'.

>> % Load your dataset and preprocess it as needed
% x2, y2, x3, y3 represent your training and testing data

% Define the range of values for the number of hidden layers and neurons
hiddenLayerSizes = [5, 10, 15, 20]; % Example: Test with 5, 10, 15, and 20 neurons
numHiddenLayers = [1, 2]; % Example: Test with 1 and 2 hidden layers

% Initialize variables to store results
bestAccuracy = 0;
bestHiddenLayers = 0;
bestNeurons = 0;

% Iterate over different configurations
for i = 1:length(numHiddenLayers)
    for j = 1:length(hiddenLayerSizes)
        % Create and train the neural network with current configuration
        net = feedforwardnet(repmat(hiddenLayerSizes(j), 1, numHiddenLayers(i)));
        net = train(net, x2, y2);

        % Predict using the trained network
        Y_pred = net(x3);

        % Calculate accuracy (or any other performance metric)
        accuracy = calculateAccuracy(Y_pred, y3);

        % Check if current configuration gives better accuracy
        if accuracy > bestAccuracy
```

```

        % Check if current configuration gives better accuracy
        if accuracy > bestAccuracy
            bestAccuracy = accuracy;
            bestHiddenLayers = numHiddenLayers(i);
            bestNeurons = hiddenLayerSizes(j);
        end
    end
end

% Display the best configuration
fprintf('Best Number of Hidden Layers: %d\n', bestHiddenLayers);
fprintf('Best Number of Neurons per Hidden Layer: %d\n', bestNeurons);
fx fprintf('Best Accuracy: %.2f%%\n', bestAccuracy * 100);

```

## Appendix 3C MATLAB code for Activation Function

Command Window

```

% Define the activation functions to test
activationFunctions = {'tansig', 'logsig', 'purelin', 'relu'};

% Initialize variables to store the best performance and the corresponding activation function
bestPerformance = inf;
bestActivationFunction = '';

% Loop through each activation function
for i = 1:length(activationFunctions)
    % Create a feedforward neural network
    net = feedforwardnet(10); % you can adjust the number of neurons as needed

    % Set the activation function for each layer
    for j = 1:length(net.layers)
        net.layers(j).transferFcn = activationFunctions(i);
    end

    % Split data into training, validation, and testing sets
    net.divideParam.trainRatio = 70/100;
    net.divideParam.valRatio = 15/100;
    net.divideParam.testRatio = 15/100;

    % Train the network
    [net, tr] = train(net, x2, y2);

    % Evaluate the network performance on the validation set
    yPred = net(x2(:, tr.valInd));
    performance = perform(net, y2(:, tr.valInd), yPred);

    % Store the best performing activation function
    fx if performance < bestPerformance

```

Command Window

```

        % Store the best performing activation function
        if performance < bestPerformance
            bestPerformance = performance;
            bestActivationFunction = activationFunctions(i);
        end

        fprintf('Activation function: %s, Performance: %.4f\n', activationFunctions(i), performance);
    end

% Display the best performing activation function
fprintf('Best Activation Function: %s, Performance: %.4f\n', bestActivationFunction, bestPerformance);

% Split data into training, validation, and testing sets
[trainInd, valInd, testInd] = divideblock(size(x2, 2), 0.7, 0.15, 0.15);
xTrain = x2(:, trainInd); yTrain = y2(:, trainInd);
xVal = x2(:, valInd); yVal = y2(:, valInd);

% Train the network using custom training loop
net = configure(net, xTrain, yTrain);
net = init(net);

% Set training parameters
epochs = 100; % Number of epochs
learningRate = 0.01;

% Custom training loop
for epoch = 1:epochs
    % Forward pass
    yPred = net(xTrain);

    % Compute loss

```

## Appendix 3D MATLAB code for Loss Function

Command Window

```
>> % Define network architecture parameters
inputSize = size(x2, 1); % Number of features in input data
hiddenLayerSize = 5; % Number of neurons in the hidden layer
outputSize = 1; % Number of outputs

% Split data into training and validation sets
trainRatio = 0.7; % 70% for training
valRatio = 0.15; % 15% for validation
testRatio = 0.15; % 15% for testing
[x2_train, x2_val, x2_test, y2_train, y2_val, y2_test] = dividerand(x2, trainRatio, valRatio, testRatio);

% List of loss functions to try
lossFunctions = {'mse', 'mae', 'crossentropy'}; % Example list

% Initialize results storage
results = struct();

% Loop over each loss function
for i = 1:length(lossFunctions)
    lossFunction = lossFunctions{i};

    % Create and configure the ANN
    net = fitnet(hiddenLayerSize);

    % Set the division of data for training, validation, and testing
    % Compute loss
    loss = lossFunctions{i}(yTrain, yPred);

    % Backward pass (Gradient Descent)
    gradients = (yPred - yTrain) .* net.LW(2, 1)' .* (net.IW(1, 1)' * xTrain)';
    net.IW(1, 1) = net.IW(1, 1) - learningRate * gradients;
    net.LW(2, 1) = net.LW(2, 1) - learningRate * (yPred - yTrain) * net.LW(2, 1);
end

% Evaluate the network performance on the validation set
yPredVal = net(xVal);
performance = perform(net, yVal, yPredVal);

% Store the best performing loss function
if performance < bestPerformance
    bestPerformance = performance;
    bestLossFunction = func2str(lossFunctions{i});
end

fprintf('Loss function: %s, Performance: %.4f\n', func2str(lossFunctions{i}), performance);
end

% Display the best performing loss function
fprintf('Best Loss function: %s, Performance: %.4f\n', bestLossFunction, bestPerformance);

% Custom loss functions
function loss = mseLoss(yTrue, yPred)
    loss = mean((yTrue - yPred).^2, 'all');
end

function loss = maeLoss(yTrue, yPred)
    loss = mean(abs(yTrue - yPred), 'all');

    % Set the division of data for training, validation, and testing
    net.divideParam.trainRatio = trainRatio;
    net.divideParam.valRatio = valRatio;
    net.divideParam.testRatio = testRatio;

    % Set the loss function
    net.performFcn = lossFunction;

    % Train the network
    [net, tr] = train(net, x2_train, y2_train);

    % Test the network
    y2_pred = net(x2_test);

    % Calculate performance
    switch lossFunction
        case 'mse'
            performance = mse(net, y2_test, y2_pred);
        case 'mae'
            performance = mae(y2_test - y2_pred);
        case 'crossentropy'
            performance = crossentropy(net, y2_test, y2_pred);
    end

    % Store the results
    results.(lossFunction).net = net;
```

## Appendix 3E MATLAB code for Batch Size and Epochs

```
Command Window
>>
>> % Define network architecture parameters
inputSize = size(x2, 1); % Number of features in input data
hiddenLayerSize = 5; % Number of neurons in the hidden layer
outputSize = 1; % Number of outputs

% Split data into training and validation sets
trainRatio = 0.7; % 70% for training
valRatio = 0.15; % 15% for validation
testRatio = 0.15; % 15% for testing
[x2_train, x2_val, x2_test, y2_train, y2_val, y2_test] = dividerand(x2, trainRatio, valRatio, testRatio);

% List of batch sizes and epochs to try
batchSizes = [16, 32, 64, 128];
numEpochsList = [50, 100, 200, 300];

% Initialize results storage
results = struct();

% Loop over each combination of batch size and number of epochs
for i = 1:length(batchSizes)
    for j = 1:length(numEpochsList)

        % Test the network
        y2_pred = net(x2_test);

        % Calculate performance
        performance = mae(net, y2_test, y2_pred);

        % Store the results
        config = sprintf('batchSize_%d_epochs_%d', batchSize, numEpochs);
        results.(config).net = net;
        results.(config).performance = performance;
        results.(config).trainRecord = tr;
    end
end

% Find the best configuration based on performance
bestConfig = '';
bestPerformance = inf;

configNames = fieldnames(results);
for i = 1:length(configNames)
    config = configNames{i};

        net.IW{1, 1} = net.IW{1, 1} - learningRate * gradients;
        net.LW{2, 1} = net.LW{2, 1} - learningRate * (yPred - yTrain) * net.LW{2, 1};
    end

    % Evaluate the network performance on the validation set
    yPredVal = net(xVal);
    performance = perform(net, yVal, yPredVal);

    % Store the best performing loss function
    if performance < bestPerformance
        bestPerformance = performance;
        bestLossFunction = func2str(lossFunctions{i});
    end

    fprintf('Loss function: %s, Performance: %.4f\n', func2str(lossFunctions{i}), performance);
end

% Display the best performing loss function
fprintf('Best Loss function: %s, Performance: %.4f\n', bestLossFunction, bestPerformance);

% Custom loss functions
function loss = mseLoss(yTrue, yPred)
    loss = mean((yTrue - yPred).^2, 'all');
end

function loss = maeLoss(yTrue, yPred)
    loss = mean(abs(yTrue - yPred), 'all');
end
Activation function: tansig, Performance: 0.5728
Activation function: logsig, Performance: 1.2474
Activation function: purelin, Performance: 0.3026
```

## Appendix 3F MATLAB code for Learning Rate

Command Window

```
>>
>> % Define network architecture parameters
inputSize = size(x2, 1); % Number of features in input data
hiddenLayerSize = 5; % Number of neurons in the hidden layer
outputSize = 1; % Number of outputs

% Split data into training and validation sets
trainRatio = 0.7; % 70% for training
valRatio = 0.15; % 15% for validation
testRatio = 0.15; % 15% for testing
[x2_train, x2_val, x2_test, y2_train, y2_val, y2_test] = dividerand(x2, trainRatio, valRatio, testRatio);

% List of learning rates to try
learningrates = [0.001, 0.01, 0.1, 0.2, 0.3];

% Initialize results storage
results = struct();

% Loop over each learning rate
for i = 1:length(learningRates)
    learningRate = learningRates(i);

    learningRate = learningRates(i);

    % Create and configure the ANN
    net = fitnet(hiddenLayerSize);

    % Set the division of data for training, validation, and testing
    net.divideParam.trainRatio = trainRatio;
    net.divideParam.valRatio = valRatio;
    net.divideParam.testRatio = testRatio;

    % Set the training parameters
    net.trainParam.epochs = 100; % You can adjust this number
    net.trainParam.min_grad = 1e-6;
    net.trainParam.lr = learningRate; % Set learning rate

    % Train the network
    [net, tr] = train(net, x2_train, y2_train);

    % Test the network
    y2_pred = net(x2_test);

    % Calculate performance

% Find the best learning rate based on performance
bestConfig = '';
bestPerformance = inf;

configNames = fieldnames(results);
for i = 1:length(configNames)
    config = configNames{i};
    if results.(config).performance < bestPerformance
        bestPerformance = results.(config).performance;
        bestConfig = config;
    end
end

% Display the best learning rate and its performance
disp(['Best learning rate: ', bestConfig]);
disp(['Performance: ', num2str(bestPerformance)]);

% Use the best network for predictions
bestNet = results.(bestConfig).net;

% Make predictions with the best network
y2_best_pred = bestNet(x2);
```

## Appendix 3G MATLAB code for Regularization parameter

```
Command Window
>> % Define network architecture parameters
inputSize = size(x2, 1); % Number of features in input data
hiddenLayerSize = 5; % Number of neurons in the hidden layer
outputSize = 1; % Number of outputs

% Split data into training and validation sets
trainRatio = 0.7; % 70% for training
valRatio = 0.15; % 15% for validation
testRatio = 0.15; % 15% for testing
[x2_train, x2_val, x2_test, y2_train, y2_val, y2_test] = dividerand(x2, trainRatio, valRatio, testRatio);

% List of regularization parameters to try
regularizationParameters = [0, 0.001, 0.01, 0.1, 1, 10];

% Initialize results storage
results = struct();

% Loop over each regularization parameter
for i = 1:length(regularizationParameters)
    regParam = regularizationParameters(i);

    % Create and configure the ANN
    net = fitnet(hiddenLayerSize);

    % Set the division of data for training, validation, and testing
    net.divideParam.trainRatio = trainRatio;
    net.divideParam.valRatio = valRatio;
    net.divideParam.testRatio = testRatio;

    % Set the training parameters
    net.trainParam.epochs = 100; % You can adjust this number

    % Set the training parameters
    net.trainParam.epochs = 100; % You can adjust this number
    net.trainParam.min_grad = 1e-6;
    net.performParam.regularization = regParam; % Set regularization parameter

    % Train the network
    [net, tr] = train(net, x2_train, y2_train);

    % Test the network
    y2_pred = net(x2_test);

    % Calculate performance
    performance = mae(net, y2_test, y2_pred);

    % Store the results
    config = sprintf('regParam %f', regParam);
    results.(config).net = net;
    results.(config).performance = performance;
    results.(config).trainRecord = tr;
end

% Find the best regularization parameter based on performance
bestConfig = '';
bestPerformance = inf;

configNames = fieldnames(results);
for i = 1:length(configNames)
    config = configNames(i);
    if results.(config).performance < bestPerformance
        bestPerformance = results.(config).performance;
        bestConfig = config;
    end
end

performance = mse(net, y2_test, y2_pred);

% Store the results
config = sprintf('regParam %f', regParam);
results.(config).net = net;
results.(config).performance = performance;
results.(config).trainRecord = tr;
end

% Find the best regularization parameter based on performance
bestConfig = '';
bestPerformance = inf;

configNames = fieldnames(results);
for i = 1:length(configNames)
    config = configNames(i);
    if results.(config).performance < bestPerformance
        bestPerformance = results.(config).performance;
        bestConfig = config;
    end
end

% Display the best regularization parameter and its performance
disp(['Best regularization parameter: ', bestConfig]);
disp(['Performance: ', num2str(bestPerformance)]);

% Use the best network for predictions
bestNet = results.(bestConfig).net;

% Make predictions with the best network
y2_best_pred = bestNet(x2);
```

---

---

---