# Predicting Overhead Cost of Building Construction in Pakistan Using Artificial Neural Networks

By


**Muhammad Waleed Anwar**

(00000172632)

2016-NUST-MS-CE&M-08


Thesis submitted in the partial fulfillment of

the requirements for the degree of

**Master's in Science**

In

**Construction Engineering & Management**


**Department of Construction Engineering & Management**

**School of Civil & Environmental Engineering (SCEE)**

**National University of Sciences & Technology (NUST)**

**Islamabad, Pakistan**


**(2020)**

This is to certify that the thesis titled

**Predicting Overhead Cost of Building Construction in Pakistan Using Artificial Neural Networks**

Submitted by

**Muhammad Waleed Anwar**

2016NUSTMSCE&M0800000172632

has been accepted towards the partial fulfillment of the requirements for the degree of

**Master of Science in Construction Engineering and Management**

Thesis Supervisor

**Dr. Muhammad Usman Hassan**

Assistant Professor,

Department of Construction Engineering and Management,

School of Civil and Environmental Engineering (SCEE),

National University of Sciences and Technology (NUST), Islamabad

# THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by Mr. Muhammad Waleed Anwar, Registration No. 00000172632, of CEM-08 SCEE has been vetted by undersigned, found complete in all respects as per NUST Statutes / Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ___________________________________

Name of Supervisor: Dr. Muhammad Usman Hassan

Date: _______________________________________

Signature (HOD): _____________________________

Date: _______________________________________

Signature (Dean/Principal): ____________________

Date: _______________________________________

# ACKNOWLEDGEMENTS

In the name of Almighty Allah, the most Merciful, the Beneficent. All praise is only for Allah who created us and always planned the best for us. I am grateful to the Almighty Allah for His countless blessings and mercy bestowed upon me through the difficulties of life and I seek His guidance and pray to Him for blessings and ease throughout this life and the life to come.

I am in debt of gratitude to my research supervisor Dr. Muhammad Usman Hassan for his guidance, motivation and constant encouragement throughout this journey. I appreciate the valuable time and personal support accorded by him. I am also indebted to my respected GEC members for providing their sincere input helping me craft my art.

I am incredibly grateful to all the volunteers and respondents for their valuable contribution to this research study. And at the end, I would like to pay my earnest and honest gratitude to my friends and family especially my parents, for their unconditional support, encouragement and patience.

# ABSTRACT

The purpose of this research is to provide an efficient and effective mean of estimating overhead costs of construction in Pakistan as a percentage of the total costs. In order to achieve this, applications of artificial neural networks were implied. This research presents an in depth knowledge over the main factors affecting the overhead costs and then using those factors to develop a neural network model which would help predict overhead cost of the building projects of Pakistan. The main data collection was done through telephonic interviews with the contractor personnel and asking them about their views on the shortlisted factors in light of overhead costs. The main data set was then used to design, test and validate the neural network model. The model that presented best results was than selected as a predicting tool for the estimation of overhead costs.

## TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

*Chapter 1*

# INTRODUCTION

Construction industry has become the most important industry in the past few years. It contributes a lot in stabilizing economy of a country by bringing in jobs for labor, material suppliers, technical personal including engineers, supervisors and quantity surveyors. The progress of a country is measured in many parameters and construction is surely one of the important factors. Despite endless efforts trying to rectify the issues this industry is suffering today, we are far from solution. Construction industry has become more dynamic over the years and this shift has rendered commonly used practices, tools and techniques obsolete. New research has to be incorporated in this industry to tackle new challenges that construction managers are facing. This research will shed some light on modern issues construction industry is suffering from and provide an unparalleled solution. This chapter includes subheadings explaining the rationale of this research, problem statement, objectives and the importance of this research.

## 1.1 Background

Construction industry plays a very important part in maintaining the economy of a country. With the boom of population in the last century and with ever increasing demand for infrastructure, its significance is justified. Construction industry has a notorious reputation when it comes to delivering projects staying under the time and budget constraints. According to (Sambasivan & Soon, 2007) construction industry faces delays globally which result in over budgeting of the project.

Construction industry has become a very competitive industry controlled solely by price (Chan, 2012). As whether a construction company wins a contract or not is mainly dependent upon the price of the bid hence the determining factor in construction companies is the value of bid (Plebankiewicz & Leśniak, 2013). Construction costs are mainly composed of direct and indirect costs. Both of these costs are responsible in ensuring success of a project so for this sole reason, controlling and managing these costs is of utter importance. Indirect costs are also known as overhead costs, although these costs are not large as compared to the direct costs yet they still need to be managed and controlled properly. Overhead costs take up a valuable chunk of the construction costs and are often deemed very important for ensuring success of a construction project.

Direct costs of construction can be attributed to the labor, material, supplies and equipment costs that must be incorporated to complete a project. Total cost of construction is mainly composed of direct costs. Contractors are provided with detailed drawings and specifications regarding the project to estimate direct costs of construction. With the advancement of communication technology, contractors can easily source the best prices from specialists and suppliers and with small tolerance level, contractors can estimate direct costs with minimum error (Chan, 2012).

Indirect costs often referred to as overhead costs are the ones which are not directly attributable to a specific project but help facilitate the process of construction. They mainly include project staffing costs, mechanical plants, project insurance and bonds, site accommodations and temporary works and facilities (Chan, 2012). Indirect costs can be divided into two subheadings.

- Project Overhead Costs
- Company’s Overhead Costs (Plebankiewicz & Leśniak, 2013)

Nature and the attributes of these costs have been explained extensively in the literature review.

## 1.2 Motive behind the research

In the public sector construction in Pakistan, bids are won mainly on the basis of lowest price where contractor qualification and experience is given secondary consideration. Contractors are provided with the contract documents including drawings and specifications to estimate the direct cost of construction which shall not vary that much from contractor to contractor. So the competing factor for winning a bid is the contractor fee and the overhead cost (Indirect cost) to control total estimated cost of construction. The accuracy of estimating costs of construction is a critical factor in the success of a company (G. H. Kim et al., 2004).

In Pakistan, the overhead cost of construction is incorporated in the bid as a mere percentage of the total direct cost and that percentage is not calculated, but assumed on the basis of past construction experiences of the contractor. This practice can prove fetal for the contractor as if the assumed cost is not adequate enough, the contractor will bear the loss and if it is assumed a lot high as compared to the actual overhead, contractor might not win the bid. So there should be a system of properly calculating and predicting the overhead cost of construction. Cost estimation being one of the important tasks, it is very difficult to accurately estimate overhead costs because of the lack of information in the early stages of construction (G. H. Kim et al., 2004).

The percentage of overhead cost estimation is considered a principle parameter in estimating the financial value of bid. Many researchers argue that the percentage of overhead to the direct cost of construction lies between 9 to 14%. In many cases, the cost of overhead had also reached 20% of the total direct cost as a result of delays that these projects had to go through due to improper planning (Gardezi et al., 2014).

Many contractors do not consider the actual cost of overhead in order to win the bid putting them in risk of losing credibility as they are unable to construct staying under the budget constraint. Improperly incorporating overhead costs has forced some contactors out of business because these costs constitute significant amount of total construction costs (Dagostino, 2002). A contractor should be well versed with proper estimation of overheads to avoid any damages that might occur (Nabil I. El-Sawalhi, 2015).

## 1.3 Problem Statement

Construction industry of Pakistan is carried out under exceptional circumstances owing to the unstable economic and political conditions. As a result of high inflation, major problems arise related to availability of materials, closures, availability of services and equipment. The volume of construction in Pakistan varies according to the political situations and donor's interest.

Considering all these variations, contractors need to find a better way to properly estimate the overhead costs to make a realistic bid amount that would not only increase their chance of winning but also reduce the amount of risk involved in making an unrealistic bid. Contractor should be aware of the importance of accurate overhead cost estimation to avoid any damages (Nabil I. El-Sawalhi, 2015). Accurate estimates are also very important to manage the project in construction phase. An inaccurate estimate would provide an inaccurate baseline to compare the actual progress of a project. This practice can prove fetal for a contractor in terms of his business as well as his/her reputation.

For the past 2 decades, many mathematical and analytical models have been developed to estimate the cost of construction projects to get a better understanding of these costs in order to properly manage and control them. Artificial Intelligence (ANNs) in particular is a new technique to generate models to predict and estimate these costs. (ElSawy et al., 2011) & (G. H. Kim et al., 2004) established that ANNs provide better results as compared to multiple regression analysis models (MRAs) and case based reasoning (CBR) in estimating construction costs. The main reason being the latter techniques require accurate inputs for estimating construction costs which is not possible in the early stages of construction because there are too many uncertainties involved. This study is carried out to help contractors in Pakistan to estimate the bid price with minimum tolerance. In order to achieve this objective, artificial intelligence is incorporated as it is expected to be the best option available today.

## 1.4 Research aim & Objectives

Artificial Intelligence makes use of Artificial Neural Networks (ANNs) which resemble the neural networks of human brain to provide basis for analysis of the raw data to useful information. A neural network is a computer based simulation that imitates the learning process of human beings (G. H. Kim et al., 2004). Neural networks have the ability to memorize and analyze large amount of data collected from experimental and numerical analysis (A. Patil et al., 2017). They have vast applications in sciences and engineering. They can provide a major breakthrough in the field of construction engineering and management because of its endless adaptability and usefulness (G. H. Kim et al., 2004). Artificial Intelligence helps in automated data collection and analysis to improve several aspects of construction engineering management i.e. reducing time, prediction, risk analysis, decision making and optimizing construction costs (A. Patil et al., 2017). In construction industry, neural networks have been developed to help contractors in many crucial construction decisions. The main aim of this research is to develop an ANN model which will help contractors to predict the overhead cost which would help them in making a more competitive bid that would eventually increase their chances of winning a contract. Objectives of this research are as under.

1. Identification of the factors effecting overhead costs of building construction in Pakistan.

2. Developing ANN models from shortlisted factors and train them using data from the past building projects in Pakistan.
3. Suggesting a best model to predict the overhead costs of building construction in Pakistan using Absolute difference percentage.

## 1.5 Importance of research

There is no such study conducted in Pakistan regarding the applications of ANNs in construction industry. Failure to correctly estimate the overhead cost of construction can lead to excessive cost overruns in construction projects. With the growing need for infrastructure and ever increasing population, construction has become one of the most important industries in the world. If overhead costs are managed and controlled properly, it can significantly increase a firm's chance of winning a bid and completing construction with minimum liabilities. This study is justified as it would facilitate the estimation of overhead costs of construction.

As there is a dearth of research on this area at national level, there is a need of spreading awareness among industry practitioners to control this enigma of excessive overhead costs because majority of the projects that have been completed recently have more overhead costs than planned. It will also be helpful for the contractors to remain competitive through execution of the project efficiently by controlling the excessive overhead costs.

This study would help contractors to make a more competitive (accurate) and promising bid that would eventually increase their chances of winning the contract. Predicting the overhead cost would provide a ball park estimate (tolerance level) which will be helpful for the contractors to know the limits of overhead costs (positive and negative). The designed models would provide an easy and efficient way of predicting the overhead costs of the future construction projects. This study would be applicable to construction process during planning and construction phase.

*Chapter 2*

# LITERATURE REVIEW

This Chapter presents an insight about the different costs associated with the construction industry with a detailed commentary about the indirect costs. This chapter will be presenting a detailed perspective of academics about the composition, characteristics and factors affecting overhead costs. It will also through light on the prospects of artificial intelligence completely explaining its usefulness, adaptability and working in the field of construction engineering management. This will be achieved by analyzing the work done by academics previously on the undersigned topic.

## 2.1 Introduction

Literature review is undoubtedly one of the most important stages in this research. It will shed light on all other phases of this research. Thorough literature about the undersigned subject provides the means of narrowing down the methodology so it would help in the decision making process. This chapter would provide answers to the following questions.

- What is the meaning of construction overhead costs?
- What items make up the overhead costs of construction?
- What are different types of overhead costs?
- What are the main factors affecting the overhead costs?
- What are artificial neural networks?
- How ANNs are revolutionizing the field of Construction?

## 2.2 Definitions

### 2.2.1 General Company Overheads:

Overhead costs of construction are commonly referred to as general condition costs and defined as the costs of facilitating the process of construction (Emerging Professionals

Companion, 2013). These costs include temporary facilities, utilities, supervisory staff salaries, engineering consultants, other professional services staff, small tools, safety and security equipment etc. Costs of subcontractors, insurance costs, bonds and permits are also incurred as general condition costs.

(Holland & Jr, 1999) defined overhead costs as the costs incurred to facilitate and simplify the project. General overhead costs are the fixed costs of items necessary to run a construction project paid by the contractor. (Dagostino, 2002), (Cilensek, 1991) argues that overhead costs are not a component of actual construction but should be paid by the contractor to support the work. Another definition says that the overhead costs are those costs that cannot be associated to a single project (Zack, 2001) & (Coombs & Palmer, 1995). (Eksteen & Rosenberg[2], 2002) said the overheads cannot be associated or recovered directly from construction sites.

Generally, overhead costs are all the costs other than the costs directly attributable to a project. Overhead costs have two categories, General Overhead costs and Job overhead costs (Peurifoy & Oberlender, 2002). Alternative terms typically used by the industry professionals include; expenses of centralized and support functions, basic fixed expenses incurred to run a firm, general office expenses and the expenses of head office and site management and site preliminaries.

El-Sawalhi, (2015) found out that the term “overhead” is often replaced by different other terms such as:

- Home office management cost
- Home office administrative cost
- Jobsite general condition cost
- Jobsite administrative cost

The survey conducted by Eksteen & Rosenberg[2], (2002) showed that the contractor respondents classify overhead costs as:

- **Administration and Management** including site managers, Home office, space rentals, services, computers, IT, Office equipment, Cleaning, Electricity, refreshments, printing and stationery.
- **Communication** including telephone bills, cell phones, faxes, postage.

- **Human Resources** including salaries, medical aid, occupational health and safety, industrial relations, protective gear for workers.
- **Transportation** including site visits, transportation of Labor and materials,
- **Financial** including auditing, legal fees, depreciation, subscriptions, group fees, donations and sponsorships.

### 2.2.2 Indirect Costs

Total costs of construction can be categorized into Direct and Indirect costs. Costs that can directly be attributable to the construction work items are direct costs. Costs that cannot be attributable to a construction work items are called indirect costs or Overhead costs (Commonwealth of Massachusetts, Division of Capital Asset Management, 2006).

The Overhead costs generally have two categories:

1. **Home Office Overhead** covering administrative costs and profit incurred by the contractor for business management.
2. **Job Site Overhead** covering site management costs and field overheads incurred by the contractor to facilitate ongoing construction.

Stolz, (2010) said that the construction cost estimates are composed of direct cost, indirect cost and profit. Indirect cost combined with profit is the overhead cost. Estimator should take precautions not to duplicate costs in the both cost heads. This would make an estimate unrealistic and reduce chances of winning the bid. Although there is an established difference between direct and indirect costs, but this difference can sometimes get hazy. To keep things simple, direct cost estimate is completed before and indirect cost estimate is done using the direct cost estimate and a preliminary construction schedule. Below are the categories of indirect costs an estimator must incorporate to make an estimate.

### 2.2.3 Site and Office Overhead Costs

Hinze (1999) defines general costs of company as the costs of running a business for example administration costs, cost of services and supplies and staff salaries. These costs continue to incur if the company is only doing one job. Generally the contractors use a percentage of the direct cost for these expenses (Clough, 1986). There are two types of overhead costs in construction company overheads and project overheads (Assaf et al., 2001). Lew, (1987) argued that company overheads are one of the major reasons why the contractors are not able to realize any profit or stay in business. This means that failure to estimate these costs properly would result in the financial collapse of a company so estimating these costs is of paramount importance. Overhead costs vary significantly from time to time ranging from 8 -- 15% of the total construction cost (Pulver, 1989).

Franks, (1984) & Lew, (1987) theorized a technique to allocate overhead costs for a specific time period by the scaling the overheads with the total cost of a project for the same time period to get the percentage value of overhead for the future projects. The total overhead rate calculated is than added to the total direct cost to arrive at bid price. This method of calculating overhead cost lacks accuracy due to a number of reasons but is a common practice among the contractors. The reason this method is inaccurate because it is obtained from the measuring parameters that are only obtained from the estimated costs. This is also a common practice with contractors, when adding overhead to the total price rarely use calculated results but assume different factors like project complexity, workshops, camps, fees, taxes, automobile costs and salaries of technical personnel to evaluate overhead costs.

Al-shanti, (2003) categorized overhead costs in to two groups, site overheads and general overheads. Site overheads are related to specific projects that the company is working on and general overheads are related to company itself. McCaffer & Baldwin, (1991) defined site overheads are those that can be directly attributable to a contract but are not specifically associated with work items. These costs will incur regardless the work is progressing or not (Hinze 1999).

As suggested by other academics, S. S. Patil & Bhangale, (2014) also illustrate two types of overhead costs, company overheads and project overheads. Company overheads can also be called general or administrative costs including the costs incurred by the contractor in maintain business and supporting the production process not directly related to a specific project. Company overheads differ from time to time but are generally in between 8 to 15 % of the total construction costs. Project overheads are also called site overheads which is specific to a project but not directly attributable to a single specific work item. Project overheads are the costs that a contractor incurs to manage a project on site (Apanavičien & Daugeliene, 2011).

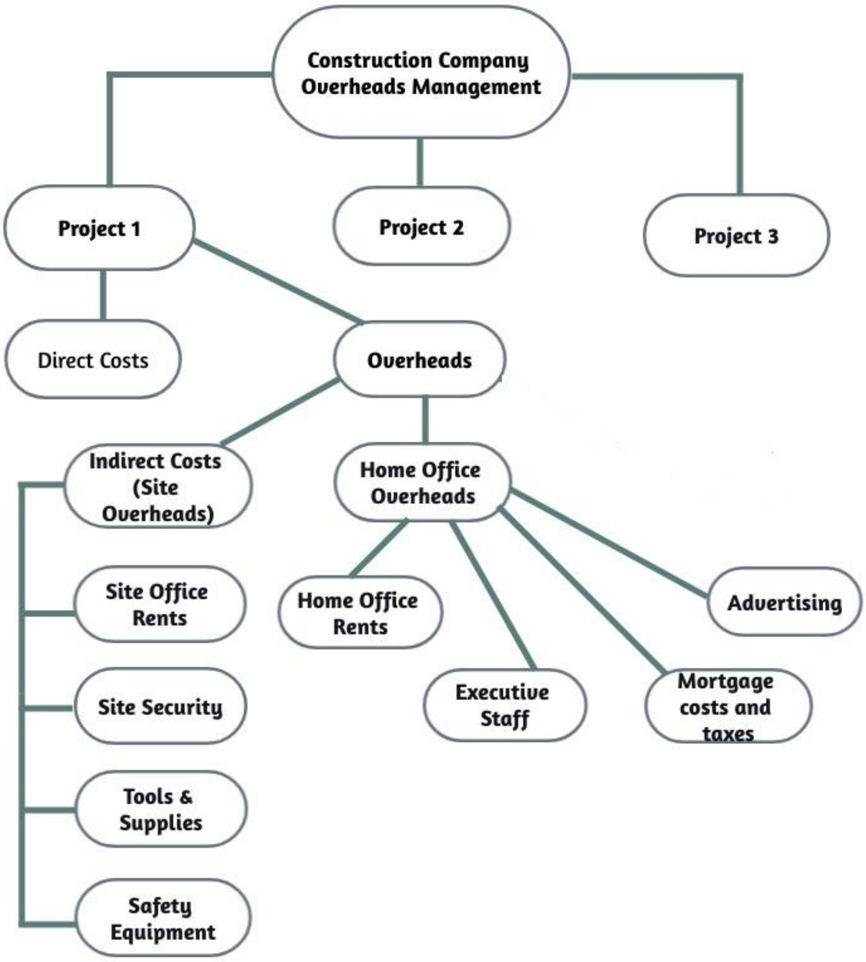


Figure 1: Cost Breakdown

### 2.2.4 Home Office Overhead

Home office overhead includes all expenses that cannot be associated with a single project like utilities, home office rentals and clericals (Neil, 1982). As they cannot be related to any single project, so these costs are divided among all the projects a company is currently undertaking by some basis (Holland & Jr, 1999). Construction overhead management is explained through Figure 1.

Zack, (2001) generally described home office overheads as the expenses that contractor has to pay for the benefit of all the running projects. These are actual costs and cannot be solely attributed to a single project. The contractor has a fare freedom in deciding the percentage of this cost but a He / She should always apply the same system for other contracts as well.

Shelton & Brugh, (2002) explained that home office overheads are required for the proper functioning of the company as a whole. These costs include administrative costs, advertisements, office rentals, taxes, furniture, utilities, owner's salary and office equipment stationeries etc. According to Taam & Singh, (2003) home office overheads contains training and recruiting employees, trade licenses, taxes, travelling, insurance, advertisements, data processing, cost of bidding, submittals and computing etc. (Nabil Ibrahim El-Sawalhi & Shehatto, 2014) concluded that overhead costs are a significant item of expense and generally run from 5 to 15 % of the total cost. General Overhead (home office overhead) generally include business expenses incurred by home office to support the construction process (Clough et al., 2000), (Nabil I. El-Sawalhi, 2015). To sum it up, home office overheads represent the costs to run activities in contractor's office to support ongoing projects in the field, (Irwin, 2005).

Lowe et al., (2003) listed out the items that would make up the home office overheads which are listed in table 1.

Table 1: Home Office Overhead Items (Lowe et al, 2003)

| | |
|---|---|
| Rents | Advertising |
| Furnishing | Non project related bonds and insurance costs |
| Office Equipment | Clerical staff |
| Mortgage costs | Estimators and Schedulers |
| Real estate taxes | Executive staff |
| Accounting and Data processing | Travel costs and automobile maintenance |
| Fees and registrations | Depreciation of assets and construction equipment |
| Utilities | Stationery |
| Interest | Legal services |
| Marketing | Hiring and training costs of employees |

### 2.2.5 Field Overheads

Field overheads are defined as the cost incurred to provide general plant and site services for example accommodations and insurances etc. (Nabil I. El-Sawalhi, 2015). Field overheads are the costs that are incurred in the field (Project) (Ruf, 2007).

Field overhead costs can be associated with a specific project excluding labor, materials or production equipment. Dagostino, 2002 stated that field overheads cannot be associated with a specific work item in a project but still required to construct a project. Field overheads also known as job overheads must be distributed on the project because it cannot be allocated to specific work items (Neil, 1982).

According to Lowe et al., (2003), a significant component of site overheads is labor overhead which is most commonly excluded in the overhead estimate. It includes sick leaves, vacations, unemployment, contribution for social security, excise and payroll taxes, retirement and medical insurance benefits and any other benefits made available to all employees by the contractor.

(Lowe et al., 2003) stated that the overheads attributed to specific contracts also include costs of indirect labor, supplies, inspection, quality control, depreciation, maintenance and support costs such as processing pay rolls etc. Ruf, (2007) pointed out the possible items on which field overheads are composed of which are listed in table 2.

Table 2: Possible field overhead items (Ruf, 2007)

| | |
|---|---|
| Trash Removal (Office) | Lodging (Home office personnel) |
| Water for field and office | Cell phones |
| Safety supplies | Office security |
| Portable toilets | Field office Expenses |
| Utilities | Telephones |
| Yard rents | Postage and shipping |
| Tools and supplies for field office | Office trailer rents |
| Field office rent | Risk insurance |
| Engineers Office rent | |
| Insurances required for contract | |

## 2.3 Cost Estimate

A cost estimate is at best an approximation of the expected cost of the project (Ahuja et al., 1994). The project cost estimate is the predicted cost of executing the work (Ritz, 1994). The society of cost estimating and analysis (SCEA) defined cost estimate as the art of approximating the probable worth or cost of activity based on the information available at time.

Advancement of cost engineering (ACE) International defines cost estimate as the approximation of cost that provides basis for business planning, project management, cost schedule and budget preparation. Dysert, (2006) defined cost estimate as a predictive process used to quantify the price of the resources required by the scope of an investment option. A cost estimate is crucial to construction contract tendering which provides basis for establishing cost of elements for the tender price of construction works (Akintoye, 2000). Cost estimation is the evaluation of factors that affect the budget of construction like labor and material (Smith & Mason, 1997).

Cost estimation is construction industry is done by three models depending on the level of accuracy required in the estimate (G.-H. Kim et al., 2004).

**Multiple Regression Model (MRA)** have been in use for estimation of costs since 1970 as it has the advantage of incorporating mathematical and statistical models and tests to check how well a curve matches a given data.

**Case Based Reasoning (CBR)** is based on experience and memory (Chen & Burrell, 2001). This technique provides estimation of costs by adapting solutions that were used to solve problems in the past projects.

**Artificial Neural Networks (ANNs)** is a computer program that imitates the learning process of a human. They have wide applications in different industries including construction industry.

## 2.4 Factors Affecting Overhead Costs

In order to develop a neural network to predict overhead cost of construction, it is necessary to find the factors which affect these costs. For this reason, an intensive review of literature is carried out. The literature on overhead costs shows that proper assessment of these costs is of utter importance and matters a great deal in ensuring success of a project. This concern has been highlighted in considerable amount of research for the identification, quantification and assessment of these costs. The following factors are extracted from intense review of the scientific studies carried out in between 1999-2018.

### 2.4.1 Explanation of the Factors

**Project Complexity:** Project complexity considers the variation and scope of work. (Akintoye, 2000) considered project complexity as the most influencing factor for overhead cost variation. Gross floor area, number of stories and basement levels, shape of site and soil conditions are a part of this factor.

**Project Location:** Travelling between located sites causes the transportation, security and other impacts which surges the overhead costs. Enshassi, Rashid Abdul Aziz, et al., (2008) concludes that distance of site from the company's seat is an important factor.

**Project size:** By increasing the size of the project, the requirement of project resources increases. Hesami & Lavasani, (2014) mentioned Size and complication affects the Organization structure which varies the overhead cost of the project.

**Payment Terms**: Delays in meeting the targeted schedule causes minute effect on the performance of the project but increases the overhead cost to meet the schedule. Delays in the payment is amongst the effective factor for the surge in overhead cost (Enshassi, Aziz, et al., 2008),(Assaf et al., 1999).

**Need for Work:** Assaf et al., (1999) enlightens that the amount and need of work is an important perimeter for determining the overhead cost of the construction project. Companies are mostly keen to induct the most experiences staff, otherwise to maintain the capital of the organization is obstacle for the company.

**The Client's Strictness in Supervision:** With the increasing demand of strictness on some critical projects best quality of the project is supposed to be subjected. As mentioned by (Y. Kim & Ballard, 2002) exposed that amongst many other reasons consultation and bureaucracy and two main causes for cost overhead overrun.

**Type of Contract:** Nature of the contract affects the overhead cost of the construction project by variations in term of liabilities and claims (Tak et al., 2002). By changing the scope of the project the cost of the contract changes and change in cost depends on the type of contract. This indirectly changes the overhead cost of the project.

**Number of competitors:** When there is more competition the lesser will be the profit margin and to remain in the competition the contractors have to minimize their overhead cost. (Assaf et al., 1999) considered the market competition to be the third most effecting factor for the variation of overhead cost of the construction project.

**Contractor's cash availability:** Contractor with low financial conditions effect the overhead cost of the construction project because of the lower finance in hand, contractor will not be able to make economical decision like buying quality product or procuring is lot. (Enshassi, Aziz, et al., 2008) has given the special influence to cash availability of the contractor and placed it at number two for the variation in overhead cost of the project.

**Percentage of Subcontracted work:** Assigning tasks to other subcontractors saves the main body from excessive spread over. But with the increase in number of parties the coordination requirement increase which slightly fluctuates the overhead cost. (Eksteen & Rosenberg2, 2002) ranked the above mentioned factors to lowest in term of its influence on the overhead cost of the project.

**Regional Economic Condition:** In the case of recession, companies have to lose their profit margin to remain in the competition. For most of the projects direct cost can't me minimized by a fair margin and to cop up with the situation companies have to reduce their overhead (Chan, 2012). These conditions directly affect the salaries, rents and quality.

**Inflation and insurance charges:** Puncreobutr & Khamkhong, (2017) had given too much emphasis on the interest rate and inflation as to be alarming factor for the increase in overhead cost of the project. Inflation tremendously affects the longer duration projects. Enshassi, Aziz, et al., (2008) had given inflation and interest rate the top priority in the overhead cost variation. Interest rate, inflation trends in the city of execution and inflation trends of the world etc.

**Stakeholder's profit:** Stakeholders profit is another factor to be incorporated in the overhead head. Chao,(2010) stated that if it is the case that the contractor in working without any profit margin than the contractor have to reduce its overhead to maintain the profit margin of the stakeholders.

**Tendering Method:** Tendering and the size of the project incur the definite effect on the overhead cost of the project Chan,(2012). Mostly the tenders are awarded by bidding process. Rates coated in bidding determine the overhead cost of the project.

**Method of performing a project:** To estimate the overhead cost an important aspect to be considered is method of performing the project. Chan,(2012) stated that ability and binding of the contractor can be determined by the way the contractor make sure to get his job done. Performing method effect procurement, financial statement and project delivery are common factors.

**Companies' classification:** PEC assigned 16 categories to construction companies, 8 comes under construction category and other 8 under operator's category. Classification is dependent on the experience and financial position of the organization. Top category firm requires large space and experienced staff to perform their work, so they would have more overhead (Enshassi, Aziz, et al., 2008)

**Type of project:** Another influencing factor on the overhead cost is the type and the nature of project. Massive projects like dam, canals, major highways would require more staff and for longer duration. Escalation is another factor in these types for project for increasing the overhead cost. (El-Sawy et al., 2011) established the type of project as among the third most influential factor for determining the overhead cost.

**Duration of project:** Project delay is another factor for determining the overhead cost of the project. Amongst the overhead cost activities 45% are those activities that are effect by the changing the duration of the activity (Chan, 2012).

**Experience of performing similar projects:** Contractor past experience would help to achieve high quality of work and take his decisions correctly and avoiding to make any false decision (Elazouni & Metwally, 2007) .Increase in cost and project delays can occur due to deficit of contractor experience (Kaming et al., 1997).

**Volume of work:** Specially talking about large organization with large staff circle, when there is no business and lesser market overhead remains the same but the earning starts deceasing (Chan, 2012). That would increase the overall overhead of the organization.

**Scope of Work:** Work scope effect the quantum of work which causes effect on the overhead cost of the construction project. By changing the work scope during the construction phase of the project affects the efficiencies of undergoing activities, as a result of which performance and overhead cost varies (Leśniak & Juszczyk, 2018).

**Project Management method:** Different organizations have different organization structure like weak, balance and strong matrix. By indulging more staff in organization structure the overhead cost increases. By adopting the organization structure that builds a team, beneficial results can be achieved (Jaya et al., 2009). It includes special construction techniques, bond requirement, Building information modeling and contractor's design input.

## 2.5 Artificial Neural Network

Artificial Intelligence makes use of Artificial Neural Networks (ANNs) which resemble the neural networks of human brain to provide basis for analysis of the raw data to useful information. A neural network is a computer based simulation that imitates the learning process of human beings (G. H. Kim et al., 2004). Neural networks have the ability to memorize and analyze large amount of data collected from experimental and numerical analysis (A. Patil et al., 2017). They have vast applications in sciences and engineering. They can provide a major breakthrough in the field of construction engineering and management because of its endless adaptability and usefulness (G. H. Kim et al., 2004). Artificial Intelligence helps in automated data collection and analysis to improve several aspects of construction engineering management i.e. reducing time, prediction, risk analysis, decision making and optimizing construction costs (A. Patil et al., 2017). In construction industry, neural networks have been developed to help contractors in many crucial construction decisions. In this research we will be going to develop an ANN model which will help contractors to predict the overhead cost which is an important prerequisite in estimating overhead cost of construction of the ongoing project.

Artificial Neural Networks (ANNs) are an analogy based technique best suited for cost prediction (Elhag & Boussabaine, 1998). ANNs learn by examples (historic data) to develop solutions for upcoming projects. They do not require any logical reasoning and set of mathematical rules to establish relation between a desired output and its influencing variables (factors). An ANN resembles a linear or multiple regression models having no observable linking variables. It means one cannot observe how a neural network actually works. It feeds effecting variables as input and presents an output. Once its parameters are specified, it can be presented as a statistical tool which can be used for forecasting, estimation and prediction purposes etc.

### 2.5.1 Structure of an Artificial Neural Network

ANNs consist of a set of neurons also known as nodes. Each neuron has small finite storage capacity necessary to operate with the local data supplied through the linkages called axons (McClelland et al., 1986). Nodes are connected numerically through axons which make way for the transfer of data from neurons. A basic neural network has an input layer, one or more middle layers also known as hidden layers and an output layer. Middle layer is called hidden because it is not known how the hidden layer interprets data to useful information as these networks do not work on mathematical reasoning defining the relationship between independent and dependent variables. Like any model, specification is the core task when using ANNs. It is very important to define the topology of a network the propagation rule, the activation function and the learning rate should also be specified to meet the specified information for which a neural network is designed in the first place. Figure 2 represents a neural network having 4 input neurons and a single output neuron.

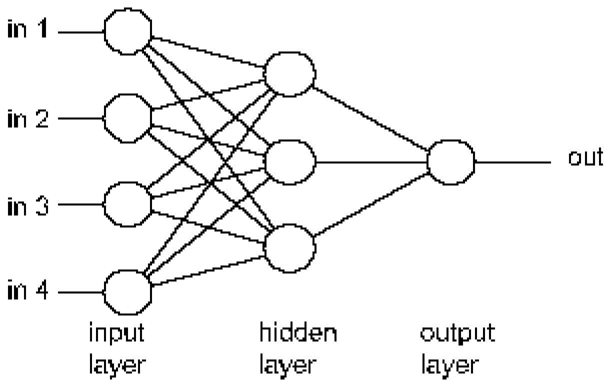


Figure 2: Basic Neural Network

### 2.5.2 Network Topology

The network topology of ANN defines how the constituent parts of a network are interrelated or arranged. It defines how many hidden layers a neural network should have, how many neurons it must have in the input, hidden and output layers etc. topology of a network varies depending on the information that a network is designed to produce.

## 2.6 Previous Work

Bastian (1994) developed a methodology to determine the number of hidden nodes for a three layered feed forward neural network for function approximation.

Elhag & Boussabaine, (1998) developed a cost assessment model using ANNs. For this, basic factors for building costs were identified and relevant data pertaining to those factors was gathered. Two ANN models were created where 13 factors were used as input nodes for model 1 and 4 factors were used for model 2 which were used to predict the lowest bid price for primary and secondary school construction.

Siqueira, (1999) used ANNs to develop a cost estimation method to generate conceptual cost estimates for the low rise prefabricated steel buildings. A computer software than automates those cost estimates by making use of neural networks to make direct cost estimates. Data was gathered of 75 such buildings over a period of 3 months. Once the ANN models were developed, they were tested against the results from the regression models by comparing with the project parameters that were not used to train the ANNs. The results produced from ANNs outperformed the ones obtained from the regression models.

Fang & Froese, (1999) designed a neural network structure to predict the cost of high strength concrete and formworks for high rise buildings. The structural elements used in the network consisted of shear walls, slabs, beams and columns. Influencing factors such as number of stories, height of walls, grid size were considered as inputs for the model to assess their effects on the cost of high performance concrete structures (HPC). Two strategies of cost estimation were used for developing ANN models, Hierarchical and hybrid strategies. From the training to validation, both strategies showed promising results but it was concluded that hybrid strategies did not produce accurate results as compared to the hierarchical but were easier to train as compared to the hierarchical strategies.

Assaf et al., (2001) inquired about the overhead cost measurement practices in Saudi Arabia with the help of a questionnaire to gather data about 61 building projects. He found out that the average percentage OH cost is slightly higher than mentioned in the literature. He also investigated about the factors that would have an impact on the overhead costs. He also argued that the unstable market is the main reason why contractors do not decide on an optimum level of overhead costs.

Al-shanti, (2003) investigated the practices for construction costs assessments in public sector building construction in Palestine to facilitate the local estimating practices.

G. H. Kim et al., (2004) compared the performance of three different cost estimation techniques: Neural Networks (NNs), Cost Based Reasoning (CBR) and Multiple Regression Analysis (MRA) on 530 past projects. He discussed that the Neural Network Model outperformed MRA and CBR estimating models because of its tendency to learn through trial and error process done using back propagation. In Long term use CBR performed well as compared to NNs. The reason for it is ANNs require continuous training from time to time with new data inputs (recently finished projects) to keep the results accurate and within the required tolerance level.

Luu & Kim, (2009) made use of the artificial neural networks to work out the total cost of apartment buildings in Vietnam. In order to identify the underlying factors (input variables), more than 90 responses were obtained through questionnaire survey. Training on the network was done using the 14 data sets acquired from already completed projects. C++ and MATLAB tools were used to apply the ANN to realistic projects. Results showed that ANN model is fairly competent in predicting the TCCs and further reinforced the reliability of using ANNs to develop cost models.

El-Sawy et al., (2011) developed a parametric cost predicting model for overhead cost on construction in Egypt using ANNs. To find out the factors affecting overhead cost, a questionnaire was circulated among the industry practitioners. More than 50 real life cases were taken in the period of 2002-2009 to train, test and validate the model. The ANN model topology consisted of 1 layer of 10 input nodes, 1 layer of 13 hidden nodes and 1 output layer of 1 node. Data entered to the input nodes using a sigmoidal function to normalize data values. Training was done through back propagation technique. The output node presented the overhead cost in percentage of the total cost. After proper training, model was tested for the data set value it was not familiar with and results were compiled which were under the acceptable tolerance.

Arafa & Alqedra, (2011) modeled a system that would estimate the cost of construction at early stages using ANNs. To perform this, a database was made comprising of more than 70 projects collected from Gaza Strip. 7 parameters were extracted that would provide input for the model. Those parameters would influence the cost of the structure skeleton for the building projects extracted from the drawings and specifications of the projects at early stages. The model had 7 input nodes in the input layer, 7 in the hidden layer and one node in the output layer. The results showed that ANN models were reasonably successful in predicting the cost of building projects with the basic information about the projects at early stages and do not require detailed information. The sensitivity analysis showed that number of stories; gross floor area and type of foundation were the most influencing factors that would contribute to increased construction cost at early stages of building construction.

Aibinu et al., (2011) developed and trained a feed forward ANN model with the database of 100 already completed projects with the help of nine input factors that affect the cost of these projects. The output variable was not the actual cost of the structure but the estimated accuracy of the model. The result showed that almost 74 % of all the projects whose cost was predicted did not differ 8.2% of the actual cost. This meant that the predicting power of ANN model was quite accurate when compared with the actual cost of projects. That model can be used as decision making tool for forecasting costs at pretender stages. It can be used to find percentage error in estimating the cost of the new projects. That difference can be kept as a cost contingency and should be included in a bid price. That model can also be used to predict the cost of new projects.

Nabil Ibrahim El-Sawalhi & Shehatto, (2014) developed a model to evaluate the cost of building construction projects with a high degree of accuracy and without the need for detailed information or drawings by using Artificial Neural Network (ANN), through developing a model that is able to help parties involved in construction projects (owner, contractors and others) in obtaining the total cost information at the early stages of project with limited available information. ANN is new approach that is used in cost estimation, which is able to learn from experience and examples and deal with nonlinear problems. It can perform tasks involving incomplete data sets, fuzzy or incomplete information and for highly complex problems. In order to build this model, quantitative and qualitative techniques were utilized to identify the significant parameters for the building project costs including skeleton and finishing phases. A database of 169 building projects was collected from the construction industry in Gaza Strip. The ANN model considered eleven significant parameters as independent input variables affected on one dependent output variable "project cost." Neuro solution software was used to train the models. The results of the trained models indicated that neural network reasonably succeeded in estimating the cost of building projects without the need for more detailed drawings. The average error of test dataset for the adapted model was largely acceptable (less than 6%). The performed sensitivity analysis showed that the area of typical floor and number of floors are the most influential parameters in building cost.

El-Sawah & Moselhi, (2014) presented a study on the use of artificial neural networks (ANNs) in preliminary cost estimating. The choice and the design of the ANN model significantly affect the results obtained from the model and, hence, the accuracy of the estimated cost. The study considered Back Propagation Neural Network.

Models were developed for order of magnitude cost estimating of low-rise structural steel buildings and short-span timber bridges. The study was conducted on actual data for 70 low-rise structural steel buildings and their respective cost was estimated using the developed regression and ANN models. These models were also applied to estimate the cost of a timber bridge extracted from the literature. The results showed that the mean absolute percentage error for the neural network models ranges from432q 16.83% to 19.35% whereas was equal to 23.72% for the regression model. Moreover, the linear regression model was more sensitive to the change of the number of the training data and that the PNN network was the most stable network among all the other estimating models as the maximum difference in MAPE percentage was only 2.46%. Whereas, the maximum difference in MAPE was 19.47%, 17.91%, and 61.45% for BPNN, GRNN and regression models respectively.

Lyne and Maximino (2014) developed an artificial neural network (ANN) model which could estimate the overall cost of building projects in the Philippines. Data which was thirty building projects were collected and randomly divided into three sets: 60% for training, 20% for validating the performance and 20% as a completely independent test of network generalization. Six input parameters, namely: number of story‘s, number of basements, floor area, volume of concrete, area of formworks‘ and weight of reinforcing steel. These inputs were entered into the ANN architecture and simulated in MATLAB. The feed forward & back propagation technique was used to generate the best model for the overall structural cost. The best ANN architecture consists of six input variables, seven nodes in the hidden layer and one output node. The resulting ANN model also reasonably estimated the overall structural cost of building projects with favorable training and testing phase outcomes.

Nabil I. El-Sawalhi, (2015) stated that the majority of contractors in Gaza Strip are aware towards overheads concept and they have good knowledge about the components of overheads. Accordingly, company owners or senior managers usually estimate overhead costs during the pricing of tenders. Around one third of contracting companies in Gaza Strip do not depend on historical data during pricing process. The OH cost is calculated based on detailed calculation for all items required by contractual conditions. No specific amount or percentage could be applicable to be added. Furthermore, during the bidding stage the overheads costs is equally distributed within each item proportionally to the total contract value. High competition in Gaza construction industry may force the contractors to reduce the HOOH percentage. Most contractors believe that submission of overhead breakdown within their bids will give them opportunity to review the overheads accurately before submission. R.Janani., (2015) discussed about overheads, overhead percentage on contractual value, factor affecting the overhead costs, major issues faced by contractors, how overhead costs affects the income, Engineers/Contractors view on overheads, investigation and control of overhead costs, creating cost awareness among employees, lists out the major items which affects the overheads, interviews with professionals and data collection from the projects and hence creates awareness while bidding and plan the financial resources effectively.

## 2.7 Summary

This chapter provided a detailed view on the overhead costs of construction, their significance and impact on the overall cost. This chapter explained the importance of accurate cost estimate in ensuring success of a project and presented a discrete background of usefulness, adaptability and applications of artificial intelligence in the field of construction engineering and management.

*Chapter 3*

# RESEARCH METHODOLOGY

This chapter will discuss the methodology used to satisfy the objectives and research aim of this study. The methodology is extracted from the literature and refined by seeking help from the academics. The main headings include questionnaire design, Artificial Neural Networks and analysis of the factors of overhead costs etc.

## 3.1 Introduction

Good methodology is the key for conducting good research. For developing the methodology for this research, intense literature study was conducted that would enlighten many different ways of fulfilling research aim. Literature review helped in pin pointing the factors that influence the overhead costs of construction. 35 factors were obtained from the literature published in between 2000 and 2018.

Ranking of these factors was done using the content analysis in which factors were ranked on their frequency (number of times they appear in literature). Once the factors are sorted from higher to lower frequency, Pareto analysis was performed to shortlist those factors.

Data collection was done in two ways:

1. The factors shortlisted from Pareto analysis were used to design the questionnaire to get the idea about the construction industry of Pakistan as to how industry professionals in Pakistan rate these factors according to their severity on the overhead costs.
2. Shortlisted factors were used to collect data from contractor firms working in Pakistan regarding the building projects they have completed, which would provide basis for the designing, training and testing of the artificial neural model.

The methodology for collecting data and designing the model is described in detail below.

## 3.2 Problem Identification

Construction industry of Pakistan is carried out under exceptional circumstances owing to the unstable economic and political conditions. As a result of high inflation, major problems arise related to availability of materials, closures, availability of services and equipment. The volume of construction in Pakistan varies according to the political situations and donor's interest. Considering all these variations, contractors need to find a better way to properly estimate the overhead costs to make a realistic bid amount that would not only increase their chance of winning but also reduce the amount of risk involved in making an unrealistic bid. This research wishes to provide an adequate source of predicting the overhead cost of building construction to the contactors working in Pakistan. For this sake, it will make use of artificial intelligence.

## 3.3 Factors affecting overhead cost

The first step in this research was to find out the factors that affect overhead costs of construction. For this reason, more than 20 research papers published in between 2000-2018 related to overhead costs of construction were read. A total of 35 factors were extracted from the literature. Many researchers used different names for similar factors which were not replicated for this study as it would have created confusion and would increase the number of factors with no added advantage. The detailed overview of the research papers used to find the factors influencing overhead costs is done in chapter 2.

## 3.4 Content Analysis

A total of 35 factors were analyzed through content analysis. For that, the frequency of every factor was measured describing how many times a specific factor been used in the research papers. In addition to that, every factor was also given a severity rating from high to low describing whether the factor mentioned in the research paper was significant or not according to the author (Ahmed et al., 2019). In order to provide factors with their severity rating, subjective judgment technique was used.

Table 3: Likert Scale 1

| Severity Rating | High | Medium | Low |
|---|---|---|---|
| Numeric counterpart | 5 | 3 | 1 |

By using above parameters, Literature Score was calculated for each factor using the below mentioned formula.

$$LS = \frac{F \times SR}{TRP \times 5}$$

Where;

LS is literature score

F is Frequency

SR is Severity rating

TRP is Total number of research papers used in content analysis

5 is the highest numeric counterpart corresponding to “high” severity rating in the table (3.1)

The factors were sorted from higher to lower on the basis of literature score value. Pareto technique was used to find the factors which affect 80% of the output variable (Overhead Cost). This technique helped to reduce 35 factors to 14 factors which would affect the overall overhead cost by 80%. Figure 3 represents Pareto analysis applied to shortlist the factors.

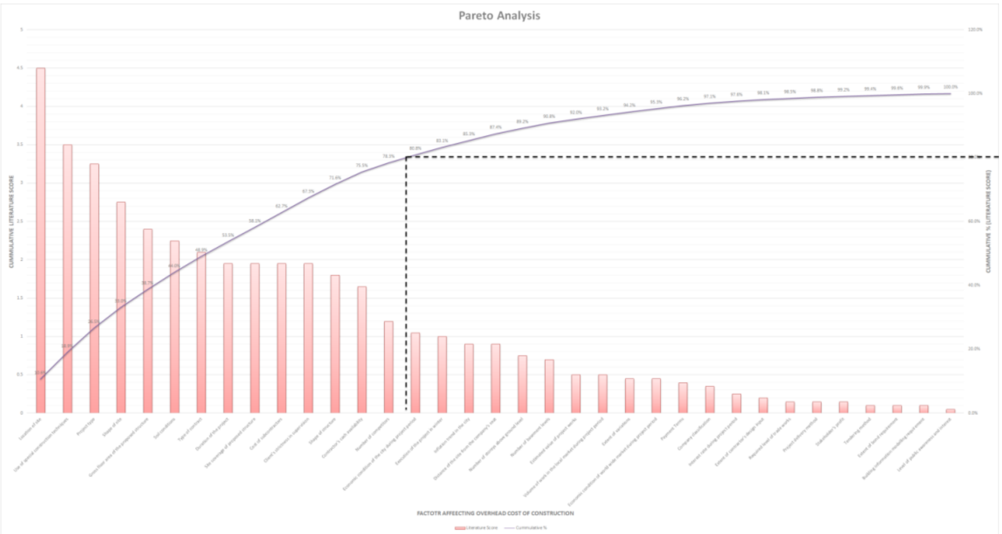


Figure 3: Pareto Analysis

## 3.5 Questionnaire Design

Content analysis was done in order to obtain and shortlist the factors affecting overhead costs. It included researches conducted all over the world. As this research will be conducted for the construction Industry of Pakistan so the perception of the industry professionals working in Pakistan regarding these factors was very important. El-Sawy et al., (2011) pointed out that the effective factors on the output can be different for different regions, in other words most effective factors on overhead costs in Pakistan might not be the same as the rest of the world. In order to cater this problem, a questionnaire survey was designed and circulated among the industry practitioners. The questionnaire survey consisted of 2 parts. The first part was to collect information about respondent e.g. amount of experience, nature of organization etc. the second part consisted of the 14 shortlisted factors obtained from the content analysis and the respondents were asked to rank those factors according to the likert scale which is represented in the Table 4.

Table 4: Likert Scale 2

| Likert scale | Never | Rarely | Sometimes | Often | Always |
|---|---|---|---|---|---|
| Numeric counterpart | 1 | 2 | 3 | 4 | 5 |

## 3.6 Analysis of the responses

After designing the questionnaire on Google forms, it was circulated among the construction industry practitioners in Pakistan. More than 72 responses were collected having quite a diverse demographic. Before analyzing responses, bogus responses had to be screened out. For this purpose, every response was checked for similarities in answers and patterned answers. 6 responses were found out to be bogus and further analysis was performed on the remaining 66 responses.

The main aim for gathering responses was to find out the field score for the shortlisted factors which would provide an insight of the perspective of industry practitioners regarding these factors. Reliability test was performed using SPSS (Statistical Package for Social Sciences) software and Cronbach’s Alpha value was computed. Tavakol & Dennick, (2011) illustrated that the Cronbach’s Alpha value in the range of 0.7 to 0.95 shows that the data is reliable for analysis. That value came out to be 0.74 greater than minimum tolerable value of 0.7 suggesting that the data set is indeed valid and reliable and can be further analyzed to compute results.

Respondent demographics showed that about 30% of the total respondents belonged to consultant organization, about 35% belonged to client organizations and about 24% belonged to the contractor organization. Almost 64% of the population had working experience in between 0 to 5 years, 25% had experience in between 6 to 10 years, 7% had experience from 11 to 15 years and the remaining respondents had experience above 16 years.

As respondents had diverse demographic meaning they belong to different nature of organizations e.g. client, contractor, consultant and other, anova test was performed to check whether there is a significant difference between the means of these groups. Anova test will tell whether there is a difference in opinion regarding the factors between the personnel belonging to different groups i.e. client, consultant and contractor etc. This can be accomplished by comparing the means of two or more groups. Test results showed that the difference in the means was significant between any two groups but anova doesn't specify which two groups. In order to find out which two groups had mean value significantly different than the other, two tailed anova test or T-test was performed. The results of the T-test showed that consultant and contractor were the two groups which do not belong to the same population and these two groups had the mean value significantly different than the other.

### 3.6.1 Relative Importance Index:

In order to find out the field score of the questionnaire, a widely known technique Relative Importance Index (RII) was used. This technique helps to rank the factors by comparing their importance with one another. Microsoft excel was used to calculate the RII (Field score) of the factors.

$$RII \ or \ FS \ = \frac{TW}{TR \times 5}$$

Where,

FS is the field score

TW is the total weightage

TR is the total number of responses used

5 represent maximum numeric value on Likert scale

### 3.6.2 Combined Score

In order to shortlist the final factors for the designing of neural network, combined score was calculated for the 14 factors. Literature score and field score was combined in different ratios (fractions) to get combined score for the factors (Ahmed et al., 2019).

50-50 combined score $(NLS \times 0.5) + (FS \times 0.5)$

40-60 combined score $(NLS \times 0.4) + (FS \times 0.6)$

30-70 combined score $(NLS \times 0.3) + (FS \times 0.7)$

Where,

NLS is normalized literature score

FS is field score

Factors were sorted out from highest to lowest combined score from each ratio. Top 7 factors (50%) were shortlisted which were same for every ratio so "40-60" was used to get final 7 factors on which further designing will be done. Having too many neurons in in a neural network can originate a lack of forecasting power because of over parameterization and too little input neurons can cause the neural network to memorize rather than predicting results. The amount of input neurons should be greater than 3 to stop the network from memorizing results and fewer than 10 to stop over parameterization (G. H. Kim et al., 2004), (Awad, n.d. 2018). The shortlisted factors are listed in table 5.

Table 5: Shortlisted Factors

| Sr.no | Factors | Notation | Combined Score |
|---|---|---|---|
| 1 | Location of site | $X_1$ | 0.873 |
| 2 | Use of special construction techniques | $X_2$ | 0.747 |
| 3 | Project type | $X_3$ | 0.698 |
| 4 | Duration of the project | $X_4$ | 0.660 |
| 5 | Client’s strictness in supervision | $X_5$ | 0.634 |
| 6 | Soil conditions | $X_6$ | 0.625 |
| 7 | Shape of site | $X_7$ | 0.602 |

## 3.7 Designing Artificial Neural Network

Artificial neural networks are widely used for prediction purposes in the field of engineering and management. In this study, Neural Designer will be used to design the models for the prediction the overhead cost of construction. Neural Designer is chosen due to its versatility and ease of application. It has a great user rating among other neural network software in 2020 and results obtained are also accurate compared to other softwares as well. Designing of a network will involve designing the topology and specific parameters (learning rate and transfer functions) related to a neural network. Many models will be designed by changing topology of the networks and performance will be measured by comparing the results with one another. The designed models will take input from the shortlisted factors and would be trained, tested and validated using Neural Designer software.

## 3.8 Steps to use Artificial Neural Networks

Artificial neural networks imitate the neural networks if a human brain. They require data for their training and application. Many researches claimed that they have better prediction capabilities as compared to CBR and MRA models. Unlike MRA, ANNs do not use the mathematical relations between the input and output variables for predictions rather they have weighed connections between the input and output variables which change during their training which cannot be justified as there is no understanding as to how neural networks predict data and this is the main reason why the functioning layers of a neural network are called “Hidden Layers”. In order to successfully deploy a neural network model in the industry, it must be first validated by carefully analyzing its performance. The 7 steps drafted by (Beale et al., 2012) are listed below.

1. Data collection
2. Creating a network
3. Configuring a network
4. Initialize weighs and Bias
5. Training a network
6. Validating a network
7. Use network

### 3.8.1 Data Collection

The preliminary survey helped shortlisting the factors affecting overhead costs of construction in Pakistan. Primary data for this research will be collected on 50 already completed building projects on the basis of those shortlisted factors. Contractors and clients will be the main source of data collection. Data will be gathered through telephonic interviews in which the respondent will be asked specific questions shortlisted through preliminary survey. This data will provide means to design, train and validate the neural networks.

### 3.8.2 Data Encoding

Neural network takes numeric values for the factors as inputs. For this reason raw data must be first converted into numerical values with the help of transfer functions (Kshirsagar & Rathod, 2012). Transfer function is a mathematical function that converts raw data (non-designable) into numerical data (designable). There are many transfer functions used in the neural network designing depending on what type of input data available and what type of output required. Unfortunately some transfer functions have better performance than the others. El-Sawy et al., (2011) showed that sigmoidal transfer functions have better performance as compared to tangential functions which can be used for predicting construction costs.

### 3.8.3 Training neural network

Training will be done on the basis of the data set. Minimum 70% of the projects in a data set will be adequate enough to successfully train the network. A data file will be created containing 40 projects which will be fed to the network for training. The actual percentage of overhead cost will be compared with the predicted overhead cost percentage to find mean square error (MSE). The MSE value will be back propagated to the network which would help the network to adjust its weighs and values in order to improve prediction results. Back propagation is a learning algorithm used by the neural network to gradually reduce the error between model output and the targeted output.

$$MSE = \frac{1}{\boldsymbol{n}}\sum_{1}^{n}(O-P)^2$$

Where “n” is the number of projects in data sheet

$O_i$ represents actual output

$P_i$ represents predicted output

Back propagation will continue up till the point where the model does not change its prediction i.e. the RMS value remains the same in successive learning sessions.

### 3.8.4 Validating neural network

A neural network model has to be validated before it can be put to use in the industry. In order to achieve this it has to be tested for the data set that it has not seen before (during training). 5 projects from the main data set will be selected at random to test the capability of the selected neural network. Absolute difference will be calculated and this value will tell the performance of the network. It is represented from the equation below.

$$ADV = \left(\frac{RO - PO}{RO}\right) \times 100$$

Where, “RO” is the real outcome

PO is the predicted outcome

ADV is the absolute difference value

An absolute value of ±10 will show that there is 10 percent difference between the actual overhead value and the predicted value of the neural network. An absolute difference value of ±4 will be the accepted tolerance level for this research.

*Chapter 4*

# MODELING AND ANALYSIS

This chapter is centered towards modeling neural networks to perform the undersigned tasks and then comparing these models with one another on the basis of their attributes. This is achieved by collecting data pertaining to building projects that were constructed in Pakistan in between 2009 to 2019. The goal of the study is to produce an artificial intelligent model that would help the contractors accurately predict the overhead cost of construction in pretender stages. This will help them to be able to make a realistic bid for a project increasing their chances of winning.

## 4.1 Introduction

An artificial neural network is an analogy based process which best suits the cost forecasting domain. The primary advantage of ANNs is that they learn from the past examples (Projects) and generalize solutions for forthcoming projects. They do not require a prerequisite to establish a relationship between an output with the potential input factors which makes them perfect to predict outcomes at the start of a project where necessary information to estimate output is limited.

This chapter will comprehensively explain step by step process of designing and analyzing the neural network models for the sake of fulfilling the objectives of this research. This chapter will also provide commentary on the process of data collection and compiling results of the selected model.

## 4.2 Data Set

Artificial Neural Networks learn from past examples, for that reason data was collected which would not only help in supervised learning but also used for validation of the best ANN model. A data collection form was designed in which contractors were asked to fill in the values of shortlisted overhead factors along with the actual overhead cost percentage that was incurred in their projects.

A data collection form consisted of two parts. First part was to collect details about the contractor organization along with their contact information. The second part consisted of the questions that would be later used for the modeling. The second part of the form consisted of 8 questions all of which were necessary to answer. Out of those 8 questions, 7 questions represented values for the input factors for a project and last question represented the target output variable, the overhead cost incurred on the project.

### 4.2.1 Data Collection

Once data collection form was designed, it was circulated to the contractors working in building construction all over Pakistan. Interviews with the contractor focal personnel (Project Managers) were conducted on telephone. Project managers were asked to gather data about their past projects according to the requirements of the data collection form through follow up calls every 3 days before the final interviews. During interviews, I would explain all the questions comprehensively to the contractors. Responses from the contractors were saved on an excel file to produce a data base. An average interview took 10 minutes to complete.

### 4.2.2 Data Statistics

For 53 valid responses, the maximums, minimums, means and standard deviations for the 7 input factors have been tabulated in table 6.

Table 6: Data Statistics

| Input Variable | Notation | Minimum | Maximum | Mean | Standard Deviation |
|---|---|---|---|---|---|
| Location of site | X1 | 1.0 | 10.0 | 5.79 | 2.29 |
| Use of special construction techniques | X2 | 1.0 | 10.0 | 5.33 | 2.09 |
| Project type | X3 | 1.0 | 10.0 | 4.67 | 2.52 |
| Duration of the project | X4 | 0.5 | 5.0 | 2.08 | 1.10 |
| Client's strictness in supervision | X5 | 1.0 | 10.0 | 5.60 | 1.89 |
| Soil conditions | X6 | 1.0 | 10.0 | 5.27 | 2.53 |
| Shape of site | X7 | 1.0 | 10.0 | 6.02 | 2.46 |

### 4.2.3 Input-Output Correlations

It might be interesting to look for dependencies between single input and single output variable. Correlation coefficient between inputs and OH% has been shown in figure 4.

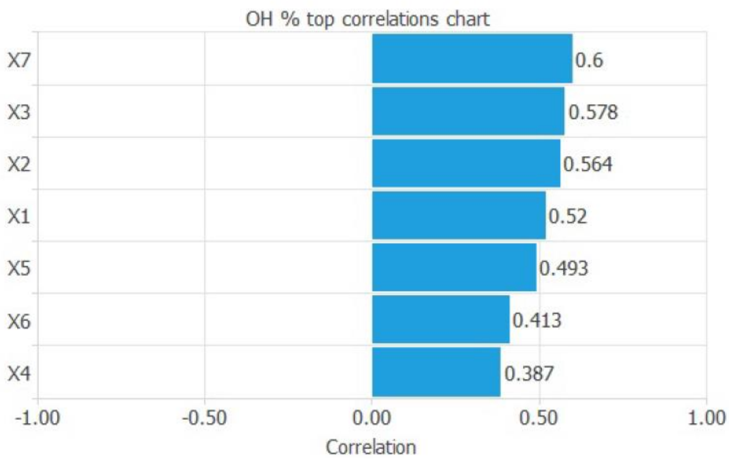


Figure 4: Correlations

Correlation coefficient close to 1 means a strong direct relation between input and output. Correlation coefficient close to '0' means that there is not a relationship between an input and an output variable. Consequently, a negative correlation between input and output would suggest an inverse relationship. In general, the output depends on many inputs simultaneously.

Figure 4 shows that the input factors X7, X3 and X2 have a strong direct relation while factors X1, X5, X6 and X4 have intermediate direct relationship with the overhead cost.

## 4.3 Neural Network Design Attributes

Artificial neural network design attributes should be fully understood before designing takes place. Design attributes depend on data set provided for modeling, the type of output required and the transfer functions.

### 4.3.1 Data Entry

Neural designer software takes in the numerical values for the inputs so for this reason the input data set has to be converted in numerical values. The data collection form for this study was already designed in such a way that it would ask the respondent to fill in numerical values against designated inputs for their projects. 7 input variables were used as linear continuous functions whose values were scaled between minimum and maximum values of the data set (1 and 10). Output variable (Overhead Percentage) was also a continuous function which has to be unscaled accordingly to get to actual output.

### 4.3.2 Data Outliers

Outliers are those instances (Projects) which do not provide any useful information to the neural network and rather negatively impact its generalization capabilities. The main aim of a neural network is to provide a generalization without any noise. Outliers should be screened out of the main data set as they would add noise which would affect the prediction capabilities of a neural network.

### 4.3.3 Data Subsets

Primary data set was composed of 56 real life projects. Out of those, 3 projects were screened out as outliers. The rest of the instances were divided in to 3 data subsets.

**Training or learning instances:**

41 instances were randomly selected for training the neural networks. Training instances are used to construct neural network models. Neural networks have to go through training before they can be deployed as forecasting tools. It has to be made sure that the training instances are composed of a wide range of projects with diversified values so that the neural network should not memorize the underlying pattern but provide a generalization for the instances that were not used to train it.

**Selection Instances:**

Selection instances are used to select a neural network model with best generalization capabilities. 7 selection instances were used for the modeling for this study. Selection instances are not used for training but they are used during training. A neural network computes error between the predicted target and real target of selection instances after each iteration. The model with minimum selection error is considered to have best generalization.

**Validation or Testing instances:**

Validating instances are used to test the prediction capabilities of a neural network. Once a neural network is selected after training on the basis of minimum selection error, it is put through validation or testing phase before it is deployed in the field as a forecasting tool. 5 instances were used for testing and validation the selected neural network.

### 4.3.4 Designing Neural Networks

Designing neural networks require a lot of work because there is no known methodology of modeling architecture of a neural network that would produce best results. Modeling is done on the data set by hit and trial method and the performance of a model is measured from errors.

#### 4.3.4.1 Architecture

The architecture of a neural network defines the complexity of a network. Neural designer suggests that for small data set and input variables, low complexity should be adequate enough to produce good generalization. Hegazy & Moselhi, (1995) stated that for cost predictions in civil engineering, a neural network with a single hidden (perceptron) layer having one half of the total number of input and output neuron is adequate. Overall simple architecture would produce good results as compared to complex neural network for our study. For this reason architecture chosen for this study is simple having at most 2 hidden layers. The classification of the neural networks designed for this study is categorized as follows.

**Shallow Neural Networks:**
Shallow neural networks are the ones having only 1 hidden or perceptron layer in its architecture with low complexity.

**Deep Neural Networks:**
Deep neural networks correspond to deep learning are the ones having more than one hidden or perceptron layers in its architecture which increase its complexity. They can have many hidden layers in their architecture but it surpasses the scope of this study. In order to stay within the scope, deep neural networks with only 2 hidden layers were be used.

#### 4.3.4.2 Structure of a Perceptron Layer

A perceptron or a hidden layer contains neurons that do not connect to the outside world but connect to the input and output neurons of a neural network (El-Sawy et al., 2011). Hidden layers are considered as the brain of a network. Figure 5 shows how a perceptron neuron connects to the input and output neurons. The scaled inputs are provided to the hidden layers through weighs which are then put through a combination function to produce a net-input value. The bias ‘b’ stores the value of the previous iteration. The combination function is represented by the eq.

$$Combination = Bias + \sum weighs * inputs$$

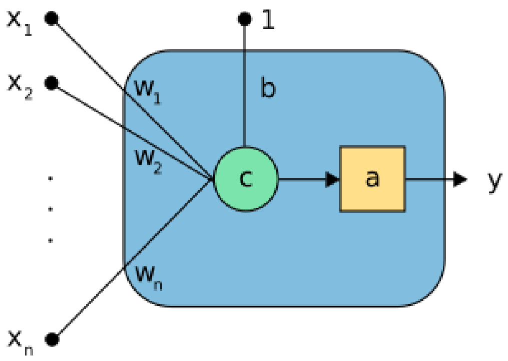


Figure 5: Perceptron Neuron

#### 4.3.4.3 Transfer Functions

Activation or transfer functions of the perceptron or hidden layers determine the function a neural network represents. Transfer functions have a direct impact on the output of generated by a neural network. The net-input (combination) is put through transfer function to produce an output.

$$Output = Transfer\ f(combination)$$

Different transfer functions produce different output. Approximation neural networks extensively make use of 3 different types of transfer functions which are used for this study as well.

**Linear Rectified:**

This function is also called a Relu function and is most commonly used function in neural networks. It is a modified form of a linear mathematical function which produces an output of 0 for negative input value and passes same positive input values as output.

$$Output = \begin{cases} 0 & if\ combination\ (input) < 0 \\ combination & if\ combination\ (input) \geq 0 \end{cases}$$

**Logistic Sigmoid:**

This is a non-linear function that can take in any value of a net-input and produce an output in between 0 and 1.

$$Output = \frac{1}{1 + e^{-combination}}$$

**Hyperbolic Tangent:**

This is a non-linear mathematical function that can take in any value of a net-input and produce an output in between -1 and 1.

$$Output = \tanh(combination)$$

#### 4.3.4.4 Nomenclature

90 neural network models were created for this study by altering the transfer functions, number of hidden layers and the number of neurons in the hidden layers. The number of input (7) and output neurons (1) remained the same for every model. Every model was given a name that would exhibit its characteristics.

- First letter of the name represents deep or shallow neural network.
- Numeric values in between represent the number of neurons in hidden layers.
- The last letters represent the transfer function.

Examples:

1. 'S-5-LS' means shallow neural network having 5 neurons in hidden layer having logistic sigmoid transfer function.
2. 'D-3-2-HT' means deep neural network having 3 neurons in first hidden layer and 2 neurons in second hidden layer with hyperbolic tangent transfer function.
3. 'S-13-LR' means shallow neural network having 13 neurons in hidden layer having linear rectified transfer function.

### 4.3.5 Errors

Training of a neural network is done through back propagating the error term measured between the actual target value and the predicted target value. The main aim of training a neural network is to reduce the error to an acceptable level. Mean squared error was used for training neural networks for this study. Two error terms are helpful in selecting the best model for a task.

1. **Training Error:** This is the final mean squared error between predicted targets of the training instances and their corresponding actual targets at the end of training.
2. **Selection Error:** This is the final mean squared error between the predicted targets of the selection instances and their corresponding actual targets at the ending of training.

### 4.3.6 Training a Neural Network

Once a neural network is designed and all the parameters are set, it has to go through training. Training instances are used for this purpose. The neural networks in this study were trained through supervised training. Supervised training is the one in which the actual target values are provided in the training data set for every instance and training progresses forward by estimating the mean squared error between the actual targets and the predicted targets by the neural network.

#### 4.3.6.1 Optimization Algorithm

Optimization algorithms are the methods which alter the parameters of a neural network during training. The main goal is to minimize the loss index in each iteration. Error generated from every iteration is back propagated through optimization according to which the weighs and biases of a network are altered so that the error term in the successive iterations will be reduced with minimum loss. Quasi-Newton optimization algorithm was used for training for the networks having non-linear transfer functions i.e. hyperbolic tangent and logistic sigmoid. Levenberg-Marquardt algorithm was used for the networks having linear transfer functions.

#### 4.3.6.2 Learning Rate

Learning rate is a positive value which determines how fast a network will adopt to a specific problem. Its value is in between 0 and 1. Keeping learning rate too high for a small data set would create a suboptimal generalization. So a learning rate value was set to 0.1 for this study.

#### 4.3.6.3 Stopping criteria

Training a neural network can take a lot of time and it can be very expensive. It depends on the type of optimization algorithm used, the learning rate and the size of training data set. Stopping criteria is a binary signal which allows early stopping of training session of a neural network. 3 stopping criteria were set for this study.

1. Reaching 50 iterations.
2. No difference in training error between the two successive iterations.
3. Training time of 3600 seconds (1 hour).

Training would stop if any one of the above stopping criteria is achieved as shown in figure 6.

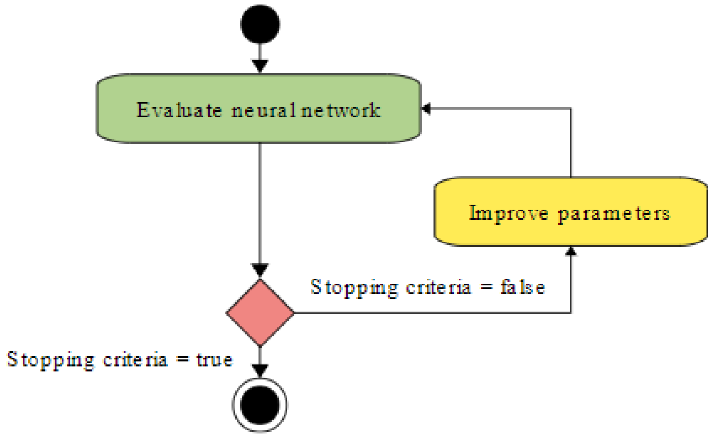


Figure 6: Training Flow chart

### 4.3.7 Generalization Fits

Generalization fits are helpful in selecting the best neural network model. Generalization of a network is figured out from the final training and selections error values.

**Over fitting:** If the training error is very small as compared to the selection error, the generalization is said to be over fit. This means that a neural network will perform exceptionally well in training but produce poor forecasts during testing.

**Under fitting:** If both the training and the selection errors are high, the generalization is said to be under fitting. In this case a network will not perform well in training as well as in testing.

**Good fit:** If the training and selection error of a neural network is minimum and the selection error is slightly larger than the training error. This is a good generalization because it provides a perfect balance between over fitting and under fitting. The network will perform well in training as well as in testing.

**Undefined fit:** if the training error is larger than the selection error, the generalization is said to be undefined. It will produce poor forecasts. It happens because of 2 main reasons.

1. Randomly selected training instances were hard to learn from.
2. Randomly selected selection instances were too easy to forecast.

## 4.4 Analysis

A total of 90 neural network models were designed according to the designed attributes mentioned in the previous section. 3 transfer functions were used which would make 30 models per transfer function. Out of those 30 neural models, 15 were shallow neural networks and 15 were deep neural networks. Their performance is explained as follows.

### 4.4.1 Shallow Neural Networks with Hyperbolic Tangent Transfer Function

Following table 7 shows the results of training of the shallow neural network models having hyperbolic tangent transfer function.

Table 7: S-HT

| Model ID | Mean Squared Error | | | Generalization |
|---|---|---|---|---|
| | Training | | Selection | |
| S1HT | 0.0854 | | 0.0865 | good fit |
| S2HT | 0.0825 | | 0.0908 | good fit |
| S3HT | 0.0367 | | 0.21 | over fit |
| S4HT | 0.0156 | | 0.103 | over fit |
| S5HT | 0.0337 | | 0.0798 | good fit |
| S6HT | 0.0049 | | 0.222 | over fit |
| S7HT | 0.0053 | | 0.17 | over fit |
| S8HT | 0.00569 | | 0.137 | over fit |
| **S9HT** | **0.00589** | | **0.0116** | **good fit** |
| S10HT | 0.007 | | 0.173 | over fit |
| S11HT | 0.0022 | | 0.219 | over fit |
| S12HT | 0.000311 | | 0.39 | over fit |
| S13HT | 0.000137 | | 0.245 | over fit |
| S14HT | 0.00248 | | 0.195 | over fit |
| S15HT | 0.0013 | | 0.603 | over fit |

**S9HT:**

Out of the above neural networks, S9HT neural network performed well as it was a good fit having smallest selection error at the end of training. The neural network has 9 neurons in its single perceptron layer. The architecture of this network is shown in the figure 7.

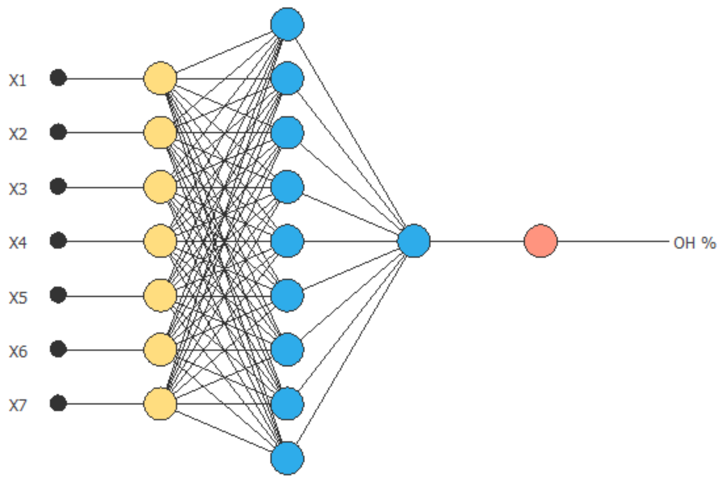

Figure 7: S9HT Architecture

From the figure, it is evident that this neural network has 7 input neurons, 7 scaling neurons (yellow), 9 hidden neurons (blue), one compiler neuron (blue), one unscaling neuron (red) and one output neuron (OH%).

### 4.4.2 Deep Neural Networks with Hyperbolic Tangent Transfer Function

Following table 8 shows the results of training of the deep neural network models having hyperbolic tangent transfer function.

Table 8: D-HT

| Model ID | Mean Squared Error (MSE) | | Generalization |
|---|---|---|---|
| | Training | Selection | |
| D11HT | 0.0837 | 0.0903 | good fit |
| D21HT | 0.0764 | 0.0924 | good fit |
| D22HT | 0.0727 | 0.0895 | good fit |
| D31HT | 0.068 | 0.0818 | good fit |
| D32HT | 0.0778 | 0.0829 | good fit |
| D33HT | 0.0404 | 0.313 | over fit |
| D41HT | 0.0368 | 0.153 | over fit |
| **D42HT** | **0.0265** | **0.0329** | **good fit** |
| D43HT | 0.0364 | 0.267 | over fit |
| D44HT | 0.0203 | 0.18 | over fit |
| D51HT | 0.0715 | 0.0875 | good fit |
| D52HT | 0.0468 | 0.0614 | good fit |
| D53HT | 0.0296 | 0.4 | over fit |
| D54HT | 0.0122 | 0.205 | over fit |
| D55HT | 0.00836 | 0.0463 | over fit |

**D42HT:**

Out of the above neural networks, D42HT neural network performed well as it was a good fit having smallest selection and training errors at the end of training. The neural network has 4 neurons in its first perceptron layer and 2 neurons in the second one. The architecture of this network is shown in the figure 8.

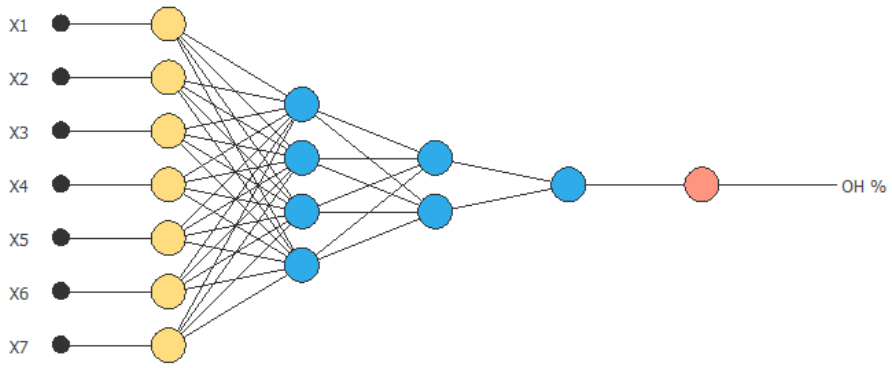


Figure 8: D42HT Architecture

### 4.4.3 Shallow Neural Networks with Logistic Sigmoid Transfer Function

Following table 9 shows the results of training of the Shallow neural network models having logistic sigmoid transfer function.

Table 9: S-LS

| Model ID | Mean Squared Error (MSE) | | Generalization |
|---|---|---|---|
| | Training | Selection | |
| S1LS | 0.0821 | 0.084 | good fit |
| S2LS | 0.0807 | 0.089 | good fit |
| S3LS | 0.054 | 0.09 | good fit |
| **S4LS** | **0.0498** | **0.0531** | **good fit** |
| S5LS | 0.055 | 0.0671 | good fit |
| S6LS | 0.0281 | 0.0788 | good fit |
| S7LS | 0.0139 | 0.18 | over fit |
| S8LS | 0.00923 | 0.0827 | over fit |
| S9LS | 0.0146 | 0.12 | over fit |
| S10LS | 0.0231 | 0.199 | over fit |
| S11LS | 0.0145 | 0.163 | over fit |
| S12LS | 0.005 | 0.24 | over fit |
| S13LS | 0.00662 | 0.19 | over fit |
| S14LS | 0.01 | 0.23 | over fit |
| S15LS | 0.000936 | 0.213 | over fit |

**S4LS:**

Out of the above neural networks, S4LS neural network performed well as it was a good fit having smallest selection error at the end of training. The neural network has 4 neurons in its single perceptron layer. The architecture of this network is shown in the figure 9.

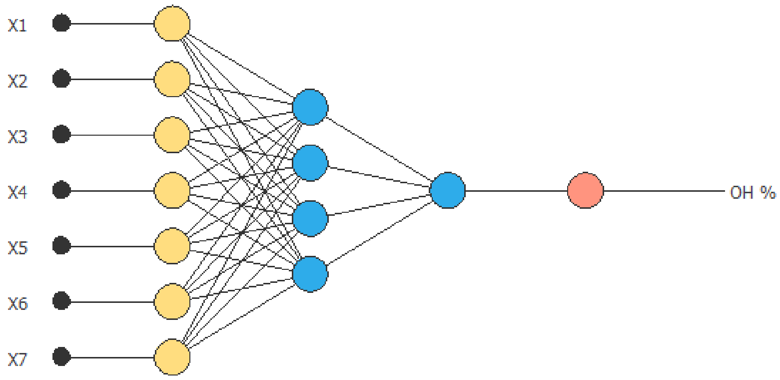


Figure 9: S4LS Architecture

### 4.4.4 Deep Neural Networks with Logistic Sigmoid Transfer Function

Following table 10 shows the results of training of the deep neural network models having logistic sigmoid transfer function.

Table 10: D-LS

| Model ID | Mean Squared Error (MSE) | | Generalization |
|---|---|---|---|
| | Training | Selection | |
| D11LS | 0.0829 | 0.0354 | undefined |
| D21LS | 0.0799 | 0.0807 | good fit |
| D22LS | 0.0519 | 0.0525 | good fit |
| D31LS | 0.0635 | 0.0447 | undefined |
| D32LS | 0.0457 | 0.0663 | good fit |
| D33LS | 0.0563 | 0.0878 | good fit |
| D41LS | 0.054 | 0.0565 | good fit |
| D42LS | 0.0347 | 0.142 | over fit |
| D43LS | 0.0549 | 0.0624 | good fit |
| **D44LS** | **0.0259** | **0.0281** | **good fit** |
| D51LS | 0.0309 | 0.0328 | good fit |
| D52LS | 0.0523 | 0.0303 | undefined |
| D53LS | 0.0226 | 0.106 | over fit |
| D54LS | 0.0168 | 0.0513 | over fit |
| D55LS | 0.0222 | 0.0719 | over fit |

**D44LS:**

Out of the above neural networks, D44LS neural network performed well as it was a good fit having smallest selection and training errors at the end of training. The neural network has 4 neurons in its first and second perceptron layers respectively. The architecture of this network is shown in the figure 10.

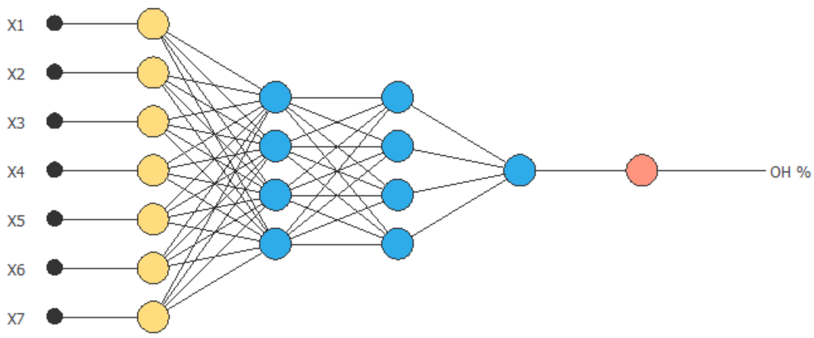


Figure 10: D44LS Architecture

### 4.4.5 Shallow Neural Networks with Linear Rectified Transfer Function

Following table 11 shows the results of training of the shallow neural network models having linear rectified transfer function.

Table 11: S-LR

| Model ID | Mean Squared Error | | Generalization |
|---|---|---|---|
| | Training | Selection | |
| S1LR | 0.407 | 0.549 | good fit |
| S2LR | 0.365 | 0.483 | good fit |
| S3LR | 0.286 | 0.369 | good fit |
| S4LR | 0.459 | 0.339 | undefined |
| S5LR | 0.617 | 0.718 | good fit |
| S6LR | 0.44 | 0.56 | good fit |
| S7LR | 0.356 | 0.314 | undefined |
| S8LR | 0.548 | 0.682 | good fit |
| S9LR | 0.58 | 0.711 | good fit |
| S10LR | 0.561 | 0.653 | good fit |
| S11LR | 0.473 | 0.342 | undefined |
| S12LR | 0.431 | 0.449 | good fit |
| S13LR | 0.366 | 0.426 | good fit |
| **S14LR** | **0.301** | **0.307** | **good fit** |
| S15LR | 0.467 | 0.328 | undefined |

**S14LR:**

Out of the above neural networks, S14LR neural network performed well as it was a good fit having smallest selection and training errors at the end of training. The neural network has 14 neurons in its single perceptron layer. The architecture of this network is shown in figure 11.

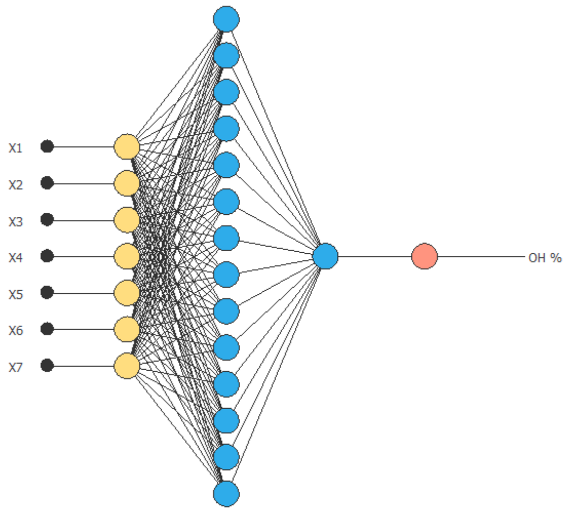


Figure 11: S14LR Architecture

### 4.4.6 Deep Neural Networks with Linear Rectified Transfer Function

Following table 12 shows the results of training of the deep neural network models having linear rectified transfer function.

Table 12: D-LR

| Model ID | Mean Squared Error (MSE) | | Generalization |
|---|---|---|---|
| | Training | Selection | |
| D11LR | 0.345 | 0.378 | good fit |
| D21LR | 0.302 | 0.352 | good fit |
| D22LR | 0.623 | 0.723 | good fit |
| D31LR | 0.624 | 0.598 | undefined |
| D32LR | 0.401 | 0.324 | undefined |
| D33LR | 0.478 | 0.58 | good fit |
| D41LR | 0.429 | 0.467 | good fit |
| D42LR | 0.37 | 0.415 | good fit |
| D43LR | 1.2 | 0.762 | undefined |
| **D44LR** | **0.288** | **0.315** | **good fit** |
| D51LR | 1 | 1.05 | good fit |
| D52LR | 0.406 | 0.52 | good fit |
| D53LR | 0.361 | 0.406 | good fit |
| D54LR | 0.399 | 0.332 | undefined |
| D55LR | 0.45 | 0.501 | good fit |

**D44LR:**

Out of the above neural networks, D44LR neural network performed well as it was a good fit having smallest selection and training errors at the end of training. The neural network has 4 neurons in its first and second perceptron layers respectively. The architecture of this network is shown in the figure 12.

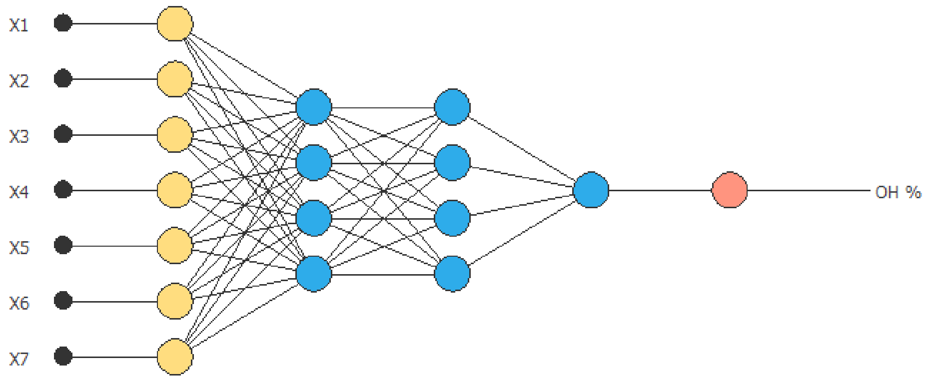


Figure 12: D44LR Architecture

## 4.5 Testing (Validation) of the Selected Neural Networks

90 designed artificial neural network models were categorized into 6 types having 15 models in each category. Best performing models from each category were selected on the basis of smallest selection and training errors. Validation was done on each selected neural network model with the help of test data set. Testing data set was composed of 5 randomly picked instances (projects) which were not used in training these networks. Input values for the instances were fed to the models to get the predicted overhead percentage value from the models. Predicted values were compared with the actual overhead percentage values to get absolute difference values and R-squared values. Best model to predict overhead cost of construction was selected on the basis of ADV and R-squared values.

**Absolute Difference Value:**

Absolute difference value (ADV) tells how much a predicted value is deviating from the actual value in terms of percentage tolerance. An absolute difference value of ±4 is selected for this study. For example, if the actual value of overhead is 10 than the predicted value of the network will only be considered correct of it is in between the tolerance of 9.6---10.4.

**R-Squared Value:**

R-squared value finds the correlation between the predicted and the actual overhead cost percentage. R-squared value closer to 1 means strong correlation. The neural model having R-squared value closest to 1 will be selected as the best model to predict overhead cost. The formula to compute R-squared is given as follows.

$$R^2 = 1 - \frac{\sum(RO - PO)^2}{\sum(RO - \overline{RO})^2}$$

Where,

RO is real output

PO is predicted output

$\overline{\text{RO}}$ is the mean of real outputs

### 4.5.1 Test results of S9HT

Table 13 shows validation results of the shallow neural network with hyperbolic tangent transfer function. 2 out of 5 projects in the testing data set had ADV greater than ±4 so the success percentage is 60.

Table 13: Validation of S9HT

| Project ID | ROH % | POH% | ADV | Status | Success % | R-Squared |
|---|---|---|---|---|---|---|
| P-1 | 7.9 | 7.738577 | 2.043325 | Pass | 60 % | 0.8359 |
| P-2 | 11.75 | 9.394989 | 20.04265 | Fail | | |
| P-3 | 10 | 10.16706 | -1.67056 | Pass | | |
| P-4 | 14.5 | 15.63448 | -7.82399 | Fail | | |
| P-5 | 8.5 | 8.469321 | 0.36093 | Pass | | |

In the following figure 13, trends of the real and predicted overhead percentages have been mapped against project IDs. Blue line represents real overhead % and red line represents predicted overhead % by the neural network.

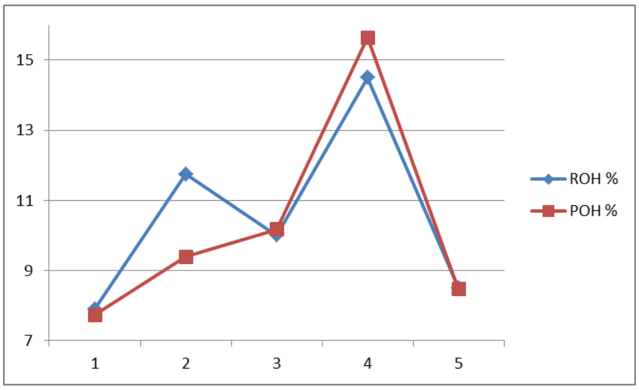


Figure 13: Trend Line (S9HT)

In the following figure 13(b), correlation between the real overhead % and the predicted overhead % has been graphed. R-squared value has also been computed.

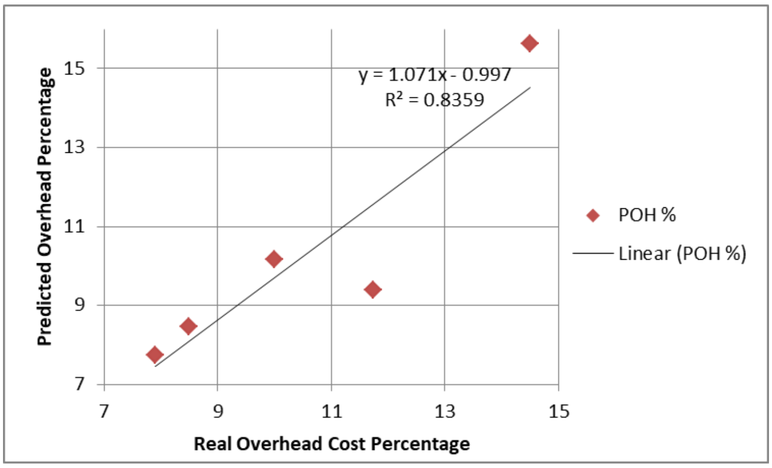


Figure 13(b)

### 4.5.2 Test results of D42HT

Table 14 shows validation results of the deep neural network with hyperbolic tangent transfer function. 2 out of 5 projects in the testing data set had ADV greater than ±4 so the success percentage is 60.

Table 14: Validation of D42HT

| Project ID | ROH % | POH% | ADV | Status | Success % | R-Squared |
|---|---|---|---|---|---|---|
| P-1 | 7.9 | 7.8070 | 1.1768 | Pass | 60 % | 0.9039 |
| P-2 | 11.75 | 11.778 | -0.2386 | Pass | | |
| P-3 | 10 | 11.7103 | -17.103 | Fail | | |
| P-4 | 14.5 | 14.4774 | 0.1558 | Pass | | |
| P-5 | 8.5 | 7.8419 | 7.7414 | fail | | |

In the following figure 14, trends of the real and predicted overhead percentages have been mapped against project IDs. Blue line represents real overhead % and red line represents predicted overhead % by the neural network.

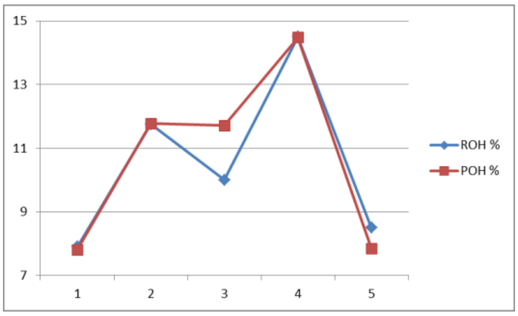


Figure 14: Trend Line (D42HT)

In the following figure 14(b), correlation between the real overhead % and the predicted overhead % has been graphed. R-squared value has also been computed.

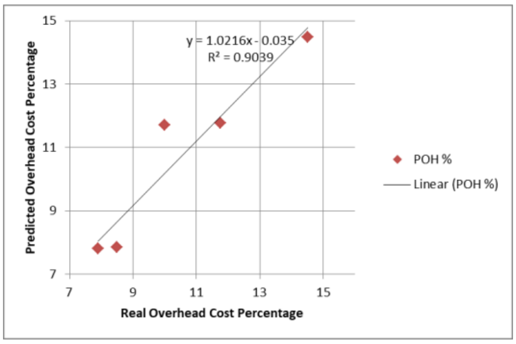


Figure 14(b)

### 4.5.3 Test results of S4LS

Table 15 shows validation results of the shallow neural network with logistic sigmoid transfer function. 1 out of 5 projects in the testing data set had ADV greater than ±4 so the success percentage is 80.

Table 15: Validation of S4LS

| Project ID | ROH % | POH% | ADV | Status | Success % | R-Squared |
|---|---|---|---|---|---|---|
| P-1 | 7.9 | 8.454018 | -7.01289 | Fail | 80 % | 0.9953 |
| P-2 | 11.75 | 11.50483 | 2.086586 | Pass | | |
| P-3 | 10 | 9.958851 | 0.411487 | Pass | | |
| P-4 | 14.5 | 13.99686 | 3.469958 | Pass | | |
| P-5 | 8.5 | 8.558362 | -0.68662 | Pass | | |

In the following figure 15, trends of the real and predicted overhead percentages have been mapped against project IDs. Blue line represents real overhead % and red line represents predicted overhead % by the neural network.

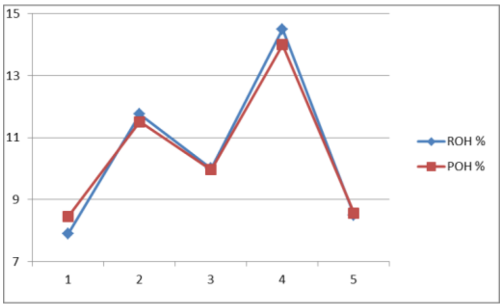


Figure 15: Trend Line (S4LS)

In the following figure 15(b), correlation between the real overhead % and the predicted overhead % has been graphed. R-squared value has also been computed.

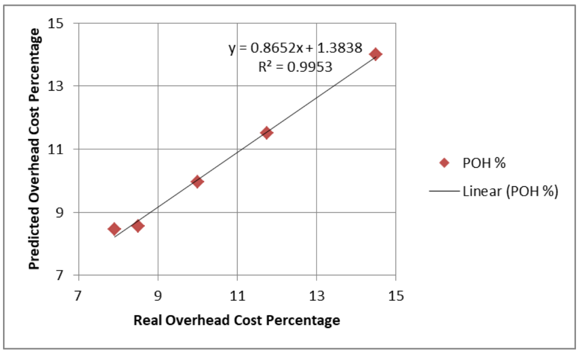


Figure 15(b)

### 4.5.4 Test results of D44LS

Table 16 shows validation results of the deep neural network with logistic sigmoid transfer function. 1 out of 5 projects in the testing data set had ADV greater than ±4 so the success percentage is 80.

Table 16: Validation of D44LS

| Project ID | ROH % | POH% | ADV | Status | Success % | R-Squared |
|---|---|---|---|---|---|---|
| P-1 | 7.9 | 8.111347 | -2.67528 | Pass | 80 % | 0.9858 |
| P-2 | 11.75 | 11.75489 | -0.0416 | Pass | | |
| P-3 | 10 | 9.911999 | 0.880009 | Pass | | |
| P-4 | 14.5 | 14.56179 | -0.42613 | Pass | | |
| P-5 | 8.5 | 7.799834 | 8.237249 | Fail | | |

In the following figure 16, trends of the real and predicted overhead percentages have been mapped against project IDs. Blue line represents real overhead % and red line represents predicted overhead % by the neural network.

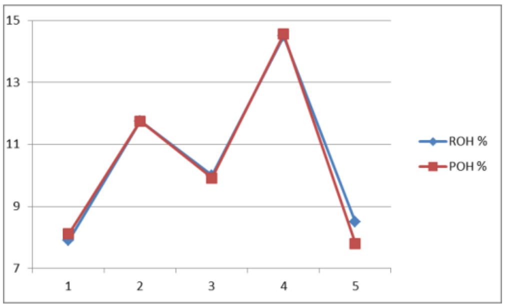


Figure 16: Trend Line (D44LS)

In the following figure 16(b), correlation between the real overhead % and the predicted overhead % has been graphed. R-squared value has also been computed.

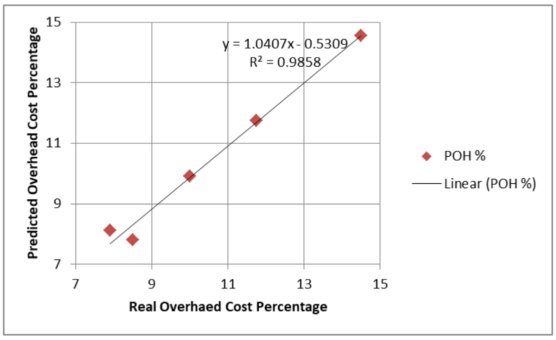


Figure 16(b)

### 4.5.5 Test results of S14LR

Table 17 shows validation results of the shallow neural network with linear rectified transfer function. 5 out of 5 projects in the testing data set had ADV greater than ±4 so the success percentage is 0.

Table 17: Validation of S14LR

| Project ID | ROH % | POH% | ADV | Status | Success % | R-Squared |
|---|---|---|---|---|---|---|
| P-1 | 7.9 | 10.94409 | -38.5328 | Fail | 0 % | 0.0892 |
| P-2 | 11.75 | 10.97959 | 6.556645 | Fail | | |
| P-3 | 10 | 11.34543 | -13.4543 | Fail | | |
| P-4 | 14.5 | 10.99738 | 24.15601 | Fail | | |
| P-5 | 8.5 | 11.25094 | -32.364 | Fail | | |

In the following figure 17, trends of the real and predicted overhead percentages have been mapped against project IDs. Blue line represents real overhead % and red line represents predicted overhead % by the neural network.

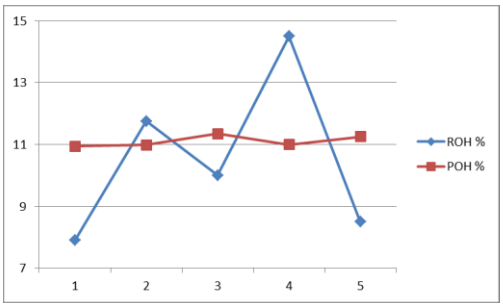


Figure 17: Trend Line (S14LR)

In the following figure 17(b), correlation between the real overhead % and the predicted overhead % has been graphed. R-squared value has also been computed.

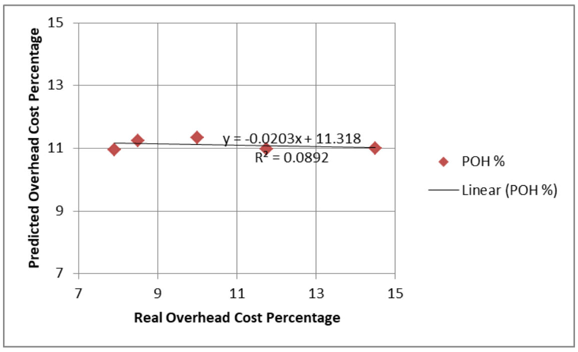


Figure 17(b)

### 4.5.6 Test results of D44LR

Table 18 shows validation results of the deep neural network with linear rectified transfer function. 5 out of 5 projects in the testing data set had ADV greater than ±4 so the success percentage is 0.

Table 18: Validation of D44LR

| Project ID | ROH % | POH% | ADV | Status | Success % | R-Squared |
|---|---|---|---|---|---|---|
| P-1 | 7.9 | 10.91932 | -38.2192 | Fail | 0 % | 0.5218 |
| P-2 | 11.75 | 11.08966 | 5.619906 | Fail | | |
| P-3 | 10 | 11.14902 | -11.4902 | Fail | | |
| P-4 | 14.5 | 11.11579 | 23.33936 | Fail | | |
| P-5 | 8.5 | 10.87585 | -27.9512 | Fail | | |

In the following figure 18, trends of the real and predicted overhead percentages have been mapped against project IDs. Blue line represents real overhead % and red line represents predicted overhead % by the neural network.

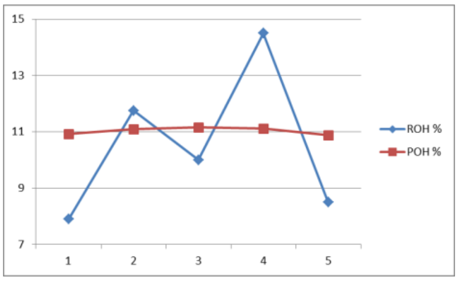


Figure 18: Trend Line (D44LR)

In the following figure 18(b), correlation between the real overhead % and the predicted overhead % has been graphed. R-squared value has also been computed.

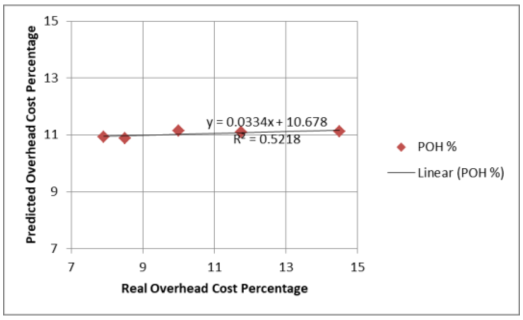


Figure 18(b)

### 4.5.7 Discussions

Out of the 90 designed artificial neural network models for this study. 6 networks were selected on which validation was done with the help of testing data set. Testing data set consisted of 5 projects. Input values of all those projects were fed to the networks to obtain the predicted results. Performance of each neural network was measured on the basis of absolute difference value and R-squared value. Summary of these results is compiled in the following table 19.

Table 19: Performance comparison of selected models

| Model ID | Transfer Function | Deep or Shallow Neural Network | # of neurons in first hidden layer | # of neurons in second hidden layer | Generalization | Success % | R-squared value | Rank |
|---|---|---|---|---|---|---|---|---|
| S-9-HT | Hyperbolic Tangent | Shallow | 9 | 0 | Good fit | 60% | 0.8359 | 4 |
| D-4-2-HT | Hyperbolic Tangent | Deep | 4 | 2 | Good fit | 60% | 0.9039 | 3 |
| **S-4-LS** | **Logistic Sigmoid** | **Shallow** | **4** | **0** | **Good fit** | **80 %** | **0.9953** | **1** |
| D-4-4-LS | Logistic Sigmoid | Deep | 4 | 4 | Good fit | 80% | 0.9858 | 2 |
| S-14-LR | Linear Rectified | Shallow | 14 | 0 | Good fit | 0% | 0.0892 | 6 |
| D-4-4-LR | Linear Rectified | Deep | 4 | 4 | Good fit | 0% | 0.5218 | 5 |

From the results presented in table 19**,** it is evident that (S4LS) shallow neural network with logistic sigmoid transfer function performed exceptionally well with 80% success rate and R-squared value **0.9953**. This study also advocates the works of (Hegazy & Moselhi, 1995) & (El-Sawy et al., 2011). A close runner up is D44LS which is a deep network with logistic sigmoid transfer function with the success rate of 80% and R-squared value of 0.9858. Overall performance of neural networks with non-linear transfer functions (hyperbolic tangent & logistic sigmoid) was satisfactory as it is evident that both deep and shallow neural networks with hyperbolic tangent transfer functions also produced a success rate of 60% and R-squared value above 0.8.

Linear transfer functions performed poorly in testing with 0% success rate. It is justified as the best selected models from linear categories had higher training and selection errors as compared to non-linear models despite the fact they had good generalization. S14LR is a shallow neural network with linear rectified transfer function having R-squared value of 0.0892 which shows a very small to no correlation between the predicted and real overhead cost percentage. On the other hand, D44LR (deep network with linear transfer function) has R-squared value 0.5218 which shows an intermediate relationship between the real and predicted overhead percentage.

## 4.6 Model Deployment

Main aim of this study is to deploy the best model in the construction industry of Pakistan which would help contractors to predict the overhead cost of construction beforehand. In order to make this happen, a complete mathematical expression of the selected neural network (S4LS) is presented below which would help contractors use the model on PYTHON or MATLAB based software as well completely rectifying the need to buy Neural Designer software.

### 4.6.1 Mathematical Expression

scaled_X1 = 2*(X1-1)/(10-1)-1;

scaled_X2 = 2*(X2-1)/(10-1)-1;

scaled_X3 = 2*(X3-1)/(10-1)-1;

scaled_X4 = 2*(X4-0.5)/(5-0.5)-1;

scaled_X5 = 2*(X5-1)/(10-1)-1;

scaled_X6 = 2*(X6-1)/(10-1)-1;

scaled_X7 = 2*(X7-1)/(10-1)-1;

y_1_1 = Logistic (-0.0978892+ (scaled_X1*0.852964)+

(scaled_X2*-2.72385)+ (scaled_X3*2.24602)+ (scaled_X4*-3.

19312)+ (scaled_X5*-0.240281)+ (scaled_X6*-0.697885)+

(scaled_X7*-4.26786));

y_1_2 = Logistic (-2.1016+ (scaled_X1*2.02781)+ (scaled_X2*-0.

591945)+ (scaled_X3*-4.90658)+ (scaled_X4*-0.886507)+

(scaled_X5*-0.474421)+ (scaled_X6*-1.02012)+

(scaled_X7*-0.382349));

y_1_3 = Logistic (-0.580874+ (scaled_X1*0.412553)+

(scaled_X2*0.675904)+ (scaled_X3*1.92791)+ (scaled_X4*0.

472742)+ (scaled_X5*-0.153561)+ (scaled_X6*1.1475)+

(scaled_X7*-1.36859));

y_1_4 = Logistic (-3.1292+ (scaled_X1*5.13009)+ (scaled_X2*0.

296859)+ (scaled_X3*2.69098)+ (scaled_X4*-1.24596)+

(scaled_X5*2.80428)+ (scaled_X6*2.83716)+

(scaled_X7*-1.84259));

scaled_OH% = (0.342933+ (y_1_1*-0.915124)+ (y_1_2*-0.400771)+

(y_1_3*-0.321696)+ (y_1_4*1.07158));

OH% = (0.5*(scaled_OH%+1.0)*(15-7)+7);

logistic(x){

return 1/(1+exp(-x))

}

*Chapter 5*

# CONCLUSIONS AND RECOMMENDATIONS

## 5.1 Introduction

This research developed and tested cost prediction models to assess the percentage of overhead cost for building construction industry of Pakistan, using the artificial neural networks. 90 feed forward networks consisting of an input layer containing 7 neurons, 7 scaling neurons and one output layer containing one neuron with a different transfer functions were developed. Study was composed of two successive steps. First step was to identify key factors that affect the percentage of overhead cost through literature and field surveys. Second step was to gather data about past completed building projects against shortlisted factors. This chapter presents conclusions from the results obtained and recommendations for future works.

## 5.2 Summary

Construction industry plays a very important part in maintaining the economy of a country. As whether a construction company wins a contract or not is mainly dependent upon the price of the bid hence the determining factor in construction companies is the value of bid. Overhead cost of a project is a significant chunk of the total cost but it cannot be determined properly during pre-bid phase. In this study, researcher developed an approach to predict the overhead costs with the help of artificial neural networks. ANN is the basis of machine learning and it tries to imitate the learning process of human beings to provide solutions to the problems that are difficult to solve with preexisting techniques. The researcher explores the factors that affect the overhead cost of construction through literature and conducting surveys to find out how severely are these factors affecting the overhead cost. Main data collection was performed against the shortlisted factors getting information of completed projects to model the neural networks. Best performing model was tested for its capabilities through testing data set and methodology to deploy this model as a tool to estimate the overhead cost in construction industry is also explained.

## 5.3 Conclusions

The following conclusions were drawn from this study:

- Thirty five factors were obtained through literature which were ranked through Pareto analysis to get to top 14 factors would serve as basis for questionnaire survey.
- Questionnaire survey helped shortlist 7 top factors through RII.
- Questionnaire showed that Location of site, use of special construction techniques and type of project were top 3 factors affecting overhead cost in Pakistan having RII of 0.873, 0.747 and 0.698 respectively.
- Data set containing data about 53 past completed projects was collected from interviews with the contractors working all over Pakistan.
- Modeling and training of 90 ANN models was performed form the data set.
- Modeling and training data set was composed of 48 instances and validating data set was composed of the remaining 5 instances (projects).
- Best model had 0.0498 and 0.0531 mean squared error values for training and selection at the end of training.
- Best model has 4 neurons in single hidden layer with logistic sigmoid transfer function.
- Validation results for the best model produced R-squared of 0.998 which shows a strong correlation between predicted and real values.
- Selected model would predict overhead percentage within ±4 ADV 4 out of 5 times with a success % of 80.

## 5.4 Recommendations for Future Work

The current study showed very promising results in predicting the overhead cost of building projects, and this approach will continue to make impressive gains especially in civil engineering field. However, some recommendations are 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.

- The model should be augmented to take into consideration the other different types of Construction projects. For example: the infrastructure construction.
- The development of artificial neural network models requires the presence of structured database for the finished projects in the construction companies. Unfortunately most construction companies have no structured database system that can provide researchers with the required information. It is recommended that a standard database system for storing information regarding the finished projects should be developed and applied by the construction companies working in Pakistan.
- In order for this model to continue predicting accurately, it should be regularly (once a year) trained with new data of the newly finished projects.

To conclude, any rapid examination of cost data is very crucial and unworkable to achieve by manual calculations or estimations in this modern days, especially in the construction industry where decisions are taken in a very rushed and short period of time. That‘s why; computer based cost models are necessitated to enable accurate responds, ease the data analysis process and shorten the time required to accomplish the job.

# REFERENCES


R. J. (2015). a Systematic Study on Site Overhead Costs in Construction Industry. *International Journal of Research in Engineering and Technology*, *04*(10), 149–151. https://doi.org/10.15623/ijret.2015.0410026

Ahmed, M., Thaheem, M. J., & Maqsoom, A. (2019). Barriers and opportunities to greening the construction supply chain management: Cause-driven implementation strategies for developing countries. *Benchmarking*, *27*(3), 1211–1237. https://doi.org/10.1108/BIJ-04-2019-0192

Ahuja, H. N., Dozzi, S. P., & Abourizk, S. M. (1994). *Project management: techniques in planning and controlling construction projects*. John Wiley & Sons.

Aibinu, A. A., Dassanayake, D., & Thien, V. C. (2011). *Use of artificial intelligence to predict the accuracy of pre-tender building cost estimate*.

Akintoye, A. (2000). Analysis of factors influencing project cost estimating practice. *Construction Management and Economics*, *18*(1), 77–89. https://doi.org/10.1080/014461900370979

Al-shanti, Y. A. E. (2003). *A Cost Estimate System for Gaza Strip Construction Contractors*.

Apanavičien, R., & Daugeliene, A. (2011). New classification of construction companies: Overhead costs aspect. *Journal of Civil Engineering and Management*, *17*(4), 457–466. https://doi.org/10.3846/13923730.2011.625629

Arafa, M., & Alqedra, M. (2011). Early stage cost estimation of buildings construction projects using artificial neural networks. *Early Stage Cost Estimation of Buildings Construction Projects Using Artificial Neural Networks*, *4*(1).

Assaf, S. A., Bubshait, A. A., Atiyah, S., & Al-shahri, M. (1999). Project Overhead Costs in Saudi Arabia. *Cost Engineering*, *41*(4), 33–38.

Assaf, S. A., Bubshait, A. A., Atiyah, S., & Al-Shahri, M. (2001). The management of construction company overhead costs. *International Journal of Project Management*, *19*(5), 295–303. https://doi.org/10.1016/S0263-7863(00)00011-9

Awad, R. M. (n.d.). *By*. *November 2017*.

Beale, M. H., Hagan, M. T., & Demuth, H. B. (2012). Neural network toolbox user's guide R2012b. *The MathWorks, Inc*, *166*, 167.

Chan, C. T. W. (2012a). The principal factors affecting construction project overhead expenses: An exploratory factor analysis approach. *Construction Management and Economics*, *30*(10), 903–914. https://doi.org/10.1080/01446193.2012.717706

Chan, C. T. W. (2012b). The principal factors affecting construction project overhead expenses: An exploratory factor analysis approach. *Construction Management and Economics*, *30*(10), 903–914. https://doi.org/10.1080/01446193.2012.717706

Chao, L. C. (2010). Estimating project overheads rate in bidding: DSS approach using neural networks. *Construction Management and Economics*, *28*(3), 287–299. https://doi.org/10.1080/01446190903473782

Chen, D., & Burrell, P. (2001). Case-based reasoning system and artificial neural networks: A review. *Neural Computing & Applications*, *10*(3), 264–276.

Cilensek, R. (1991). Understanding contractor overhead. *Cost Engineering*, *33*(12), 21.

Clough, R. H., Sears, G. A., & Sears, S. K. (2000). *Construction project management*. John Wiley & Sons.

Eksteen, B., & Rosenberg[2], D. (2002). the Management of Overhead Costs in Construction Companies. *Association of Researchers in Construction Management*, *1*(September), 2–4. http://www.arcom.ac.uk/-docs/proceedings/ar2002-013-022_Eksteen_and_Rosenberg.pdf

El-Sawah, H., & Moselhi, O. (2014). Comparative study in the use of neural networks for order of magnitude cost estimating in construction. *Journal of Information Technology in Construction (ITcon)*, *19*(27), 462–473.

El-Sawalhi, Nabil I. (2015). An Overhead Costs Assessment for Construction Projects at Gaza Strip. *American Journal of Civil Engineering*, *3*(4), 95. https://doi.org/10.11648/j.ajce.20150304.11

El-Sawalhi, Nabil Ibrahim, & Shehatto, O. (2014). A Neural Network Model for Building Construction Projects Cost Estimating. *Journal of Construction Engineering and Project Management*, *4*(4), 9–16. https://doi.org/10.6106/jcepm.2014.4.4.009

El-Sawy, I., Hosny, H., & Razek, M. A. (2011). A Neural Network Model for Construction Projects Site Overhead Cost Estimating in Egypt. *International Journal of Computer Science Issues*, *8*(3), 273–283.

Elazouni, A. M., & Metwally, F. G. (2007). Expanding Finance-Based Scheduling to Devise Overall-Optimized Project Schedules. *Journal of Construction Engineering and Management*, *133*(1), 86–90. https://doi.org/10.1061/(ASCE)0733-9364(2007)133:1(86)

Elhag, T. M. S., & Boussabaine, A. H. (1998). An artificial neural system for cost estimation of construction projects. *Proceedings of the 14th ARCOM Annual Conference*.

ElSawy, I., Hosny, H., & Razek, M. A. (2011). *A Neural Network Model for Construction Projects Site Overhead Cost Estimating in Egypt*. *8*(3), 273–283. http://arxiv.org/abs/1106.1570

Enshassi, A., Aziz, A. R. A., & Karriri, A. El. (2008). Investigating the overhead costs in construction projects in Palestine. *Journal of Financial Management of Property and Construction*, *13*(1), 35–47. https://doi.org/10.1108/13664380810882066

Enshassi, A., Rashid Abdul Aziz, A., & Karriri, A. (2008). Investigating the overhead costs in construction projects in Palestine. *Journal of Financial Management of Property and Construction*, *13*(1), 35–47. https://doi.org/10.1108/13664380810882066

Fang, C. F., & Froese, T. (1999). Cost estimation of high performance concrete (HPC) high-rise commercial buildings by neural networks. *Durability of Building Material & Components*, *8*, 2476–2486.

Franks, J. (1984). *Building sub-contract management: purpose, tasks, skills, and responsibilities*. Construction Press.

Gardezi, S. S. S., Manarvi, I. A., & Gardezi, S. J. S. (2014). Time extension factors in construction industry of Pakistan. *Procedia Engineering*, *77*, 196–204. https://doi.org/10.1016/j.proeng.2014.07.022

Hegazy, T., & Moselhi, O. (1995). Elements of Cost Estimation: A survey in Canada and the unit. *Cost Engineering*, *37*(5), 27.

Hesami, S., & Lavasani, S. A. (2014). Identifying and Classifying Effective Factors Affecting Overhead Costs in Constructing Projects in Iran. *International Journal of Construction Engineering and Management*, *3*(1), 24–41. https://doi.org/10.5923/j.ijcem.20140301.03

Holland, N. L., & Jr, D. H. (1999). Indirect cost categorization and allocation by construction contractors. *Journal of Architectural Engineering*, *5*(2), 49–56.

Irwin, J. W. (2005). Current Issues to Watch For in Construction Claims. *Newsletter: Construction Update-March*.

Jaya, N. M., Pathirage, C. P., & Sutrisna, M. (2009). *Analysis of Critical Success Factors for Improving the Management of Project Overheads during the Construction Stage in the Construction Industry*. 1–12.

Kaming, P. F., Olomolaiye, P. O., Holt, G. D., & Harris, F. C. (1997). Factors influencing construction time and cost overruns on high-rise projects in Indonesia. *Construction Management and Economics*, *15*(1), 83–94. https://doi.org/10.1080/014461997373132

Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, *39*(10), 1235–1242.

Kim, G. H., An, S. H., & Kang, K. I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, *39*(10), 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013

Kim, Y., & Ballard, G. (2002). Case study-overhead costs analysis. *... IGLC-10, Gramado, Brazil,(August, 2002), Lci*, 1–13. http://www.leanconstruction.org/media/docs/IGLC02Kim-Ballard.pdf

Kshirsagar, P., & Rathod, N. (2012). Artificial neural network. *International Journal of Computer Applications*, 12–16.

Leśniak, A., & Juszczyk, 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. https://doi.org/10.1016/j.acme.2018.01.014

Lew, A. E. (1987). *Means interior estimating*. RS Means Company.

Lowe, S., Bielek, R., & Burnham, R. (2003). *Compensation for Contractors' Home Office Overhead* (Vol. 315). Transportation Research Board.

Luu, V. T., & Kim, S.-Y. (2009). Neural network model for construction cost prediction of apartment projects in Vietnam. *Korean Journal of Construction Engineering and Management*, *10*(3), 139–147.

McCaffer, R., & Baldwin, A. N. (1991). *Estimating and tendering for civil engineering works*. BSP Professional.

McClelland, J. L., Rumelhart, D. E., & Group, P. D. P. R. (1986). Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol. 1: Foundations. *Language*, *63*(4), 45–76.

Neil, J. M. (1982). *Construction cost estimating for project control*. Prentice Hall.

Patil, A., Patted, L., Tenagi, M., Jahagirdar, V., Patil, M., & Gautam, R. (2017). Artificial Intelligence as a Tool in Civil Engineering-A Review. *IOSR Journal of Computer Engineering* , *36–39*(Acbcda), 2278–8727. www.iosrjournals.org

Patil, S. S., & Bhangale, P. P. (2014). Overhead Cost in Construction Industry. *International Journal of Industrial Engineering & Technology (IJIET)*, *4*, 1–6.

Plebankiewicz, E., & Leśniak, A. (2013). Overhead costs and profit calculation by Polish contractors. *Technological and Economic Development of Economy*, *19*(1), 141–161. https://doi.org/10.3846/20294913.2012.763008

Puncreobutr, V., & Khamkhong, Y.-. (2017). Overheads Cost Management in Condominium Construction Project in Thailand. *SSRN Electronic Journal*, 1–10. https://doi.org/10.2139/ssrn.2970414

Ritz, G. J. (1994). *Total construction project management*.

Ruf, H. B., & Ruf, S. (2007). *DOCUMENTING AND PRESENTING COSTS IN UNDERGROUND CONSTRUCTION CLAIMS*.

Sambasivan, M., & Soon, Y. W. (2007). Causes and effects of delays in Malaysian construction industry. *International Journal of Project Management*, *25*(5), 517–526.

Shelton, F., & Brugh, M. (2002). Indirect costs of contracts. *Journal of Construction Accounting and Taxation*, *12*(4), 3–9.

Siqueira, I. (1999). *Neural network-based cost estimating*. Concordia University.

Smith, A. E., & Mason, A. K. (1997). Cost estimation predictive modeling: Regression versus neural network. *The Engineering Economist*, *42*(2), 137–161.

Taam, T. M. C., & Singh, A. (2003). Unabsorbed overhead and the Eichleay formula. *Journal of Professional Issues in Engineering Education and Practice*, *129*(4), 234–245.

Tak, C., Chan, W., & Pasquire, C. (2002). Estimation of Project Overheads: A Contractor's Perspective. *Association of Researchers in Construction Management*, *1*(September), 53–62. http://www.arcom.ac.uk/-docs/proceedings/ar2002-053-062_Chan_and_Pasquire.pdf

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53.

Zack, J. G. (2001). Calculation and recovery of home office overhead. *AACE International Transactions*, *2*, 1–6.

Dagostino, F. R. (2002). *Estimating in building construction. 6th edition*, PrenticeHall, Englewood Cliffs, N.J.'

Coombs, W. E., and Palmer, W. J. (1995). *Construction accounting and financial management. 5th edition*, McGraw-Hill, New York.

Peurifoy, R. L., and Oberlender, G. D. (2002). *Estimating construction costs. 5th edition*, McGraw-Hill, New York.

John M. Stolz, P.E. Overhead and Uncertainty in Cost Estimates: A Guide to Their Review

Hinze, J. (1999). *Construction Planning and Scheduling. Prentice Hall*, Columbus, Ohio.

Clough, R. (1986). Construction Contracting. 5th Ed, John Wiley & Sons, New York.

Emerging Professional‘s Companion (2013). Construction cost.

Commonwealth of Massachusetts, (2006). Consultants Estimating Maunal.

Pulver, HE. (1989). *Construction estimates and costs.* New York, NY,, McGraw-Hill.

Dysert, L. R. (2006). Is" Estimate Accuracy" an Oxymoron?. *AACE International Transactions*, ES11.