# Investigation of Factors Impacting Construction Cost Estimate to Develop Construction-Driven Artificial Neural Network (ANN)

by

Salem Al Saber

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2023 by the
Graduate Supervisory Committee:

Kenneth Sullivan, Chair
Richard Standage
Kristen Hurtado

ARIZONA STATE UNIVERSITY

August 2023

## ABSTRACT

Construction industry is the backbone of any country's economy. It is a primary source of foreign investments, creates new jobs, and maintains the economy flowing in various trades. Accurate cost estimation is a critical aspect for the construction industry, directly impacting project success and profitability. This master's thesis focuses on comprehensively identifying the key factors that influence cost estimation and provides valuable recommendations for constructing an optimized Artificial Neural Network (ANN) model. Through an extensive research methodology encompassing literature review, surveys, and interviews with industry professionals, this study uncovers significant factors that exert a substantial impact on cost estimation practices. The findings emphasize the importance of seamlessly integrating project delivery systems, meticulously considering project duration, and incorporating diverse perspectives from global regions. By incorporating these insights, stakeholders can make informed decisions, enhance project planning, and elevate overall project performance. This study successfully bridges the gap between theory and practice, presenting invaluable insights for stakeholders within the construction industry.

Keywords: cost estimation, construction industry, Artificial Neural Network, factors, project delivery systems, project duration, global perspectives, informed decision-making, project planning, project performance

## DEDICATION

I dedicate this thesis to my beloved family: my father, my mother, and my brothers Saif and Fathi. Your unwavering love, support, and encouragement have been the driving force behind my academic journey. Your sacrifices, understanding, and belief in my abilities have inspired me to reach new heights. This accomplishment is a testament to the values you have instilled in me and the constant support you have provided. Thank you for always being my pillars of strength and for believing in my dreams.

I would also like to express my heartfelt gratitude to my dear friends and colleagues who have been by my side throughout this research endeavor. To engineer Bashar, engineer Mohammed Nemrawi, and engineer Ghada; your guidance, assistance, and unwavering support have been invaluable. Your expertise, insightful discussions, and encouragement have enriched my work and contributed to its success. I am grateful for the collaborative spirit and camaraderie we have shared.

To all my friends who have been a part of my academic journey, your friendship, encouragement, and understanding have made this experience more meaningful and enjoyable. Your presence and shared experiences have provided a sense of belonging and support that I will cherish forever.

Thank you to each and every individual who has played a role in shaping my academic and personal growth. Your support and belief in me have made this achievement possible.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to all those who have contributed to the completion of this thesis.

First and foremost, I extend my heartfelt appreciation to my advisor, Dr. Sullivan, for his guidance, expertise, and unwavering support throughout this research journey. His invaluable insights, constructive feedback, and dedication to my academic development have been instrumental in shaping this thesis. I am truly grateful for the opportunity to work under his mentorship.

I am also deeply grateful to the faculty and staff at Arizona State University for providing a stimulating academic environment and resources that have nurtured my intellectual growth. I extend my thanks to the faculty members in the Ira Fulton School of Engineering for their knowledge, guidance, and commitment to excellence in education.

I would like to express my appreciation to my family for their unwavering support, love, and encouragement throughout my academic journey. I am especially grateful to my parents for instilling in me the values of perseverance, dedication, and the pursuit of knowledge.

Finally, I would like to express my deepest appreciation to all the individuals who have contributed to this thesis in any way, whether through their encouragement, feedback, or constructive criticism. Specifically, I would like to thank all the people who assisted and supported throughout the data collection process.

Thank you all for your contributions and for being a part of this important milestone in my academic and personal growth. Your involvement and support have been invaluable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

## LIST OF ABBREVIATION

ANN - Artificial Neural Network

BIM - Building Information Modeling

UAE - United Arab Emirates

MLP - Multi-Layer Perceptron

RNN - Recurrent Neural Network

PMI - Project Management Institute

AACE - Association for the Advancement of Cost Engineering

EVM - Earned Value Management

CBR – Case Base Resonance

MAE – Mean Absolute Error

MSE – Mean Squared Error

SGD – Stochastic Gradient Descent

MLP- Multilayer Perceptron

# CHAPTER 1

# INTRODUCTION

## Background

A construction project can be defined successful by being limited within the triple constraint or the golden triangle, which encompasses the interrelated factors of budget, time, and quality. Accurate cost estimation plays a crucial role in enabling stakeholders and decision-makers to conduct comprehensive feasibility studies, determine appropriate financial scales during the bidding stage, and effectively monitor cash flows throughout the construction phase (Elmousalami, 2020; Matel et al., 2019; Wang et al., 2022).

It is imperative to minimize errors in construction cost estimation as accurate information empowers stakeholders in construction engineering and management to make well-informed decisions aligned with the project's financial requirements (Wang et al., 2022). Moreover, accurate cost estimation plays a pivotal role in determining whether to proceed or to cancel the project (Matel et al., 2019).

Each construction project possesses unique characteristics, and the estimation of resource utilization and specification choices significantly impacts its associated costs. Throughout the various stages of a building project's life cycle, designing teams encounter risk and uncertainty. Therefore, it is essential to duly recognize and account for all relevant factors in order to mitigate the potential for project delays and cost overruns (Alqahtani & Whyte, 2016).

Wang et al. (2022) outlines that there is a general consensus in Hong Kong project cost estimation models where insufficient consideration is given to incorporating the trend and value of market indices in construction cost estimation, such as stock market index, construction indices, and daily wages. This oversight may contribute to significant disparities between the actual final cost and the initial construction projects estimates, despite the utilization of advanced estimation approaches by stakeholders and contractors. Incorporating the trend and value of these market indices into construction cost estimation is crucial to address the issue and improve the accuracy of cost projections (Wang et al., 2022).

Furthermore, the accuracy of cost estimates can vary among estimators due to differences in their levels of experience (Matel et al., 2019). Acquiring the expertise needed to perform cost estimation is a time-intensive endeavor for professional engineers, often taking several years to develop the necessary skills. Moreover, this expertise is frequently not documented or formally validated, making it susceptible to subjectivity and variability (Elfaki et al., 2014). According to Juszczyk et al. (2018), a project failure maybe caused by both underestimation and overestimation. Underestimating the cost of a construction project often results in cost overruns, leading to financial losses for stakeholders (Wang et al., 2022). Additionally, Elfaki et al. (2014) outlined that inaccurate cost estimation can result in various problems, including change orders, construction delays, and even bankruptcy.

In contrast, trained neural networks can leverage their knowledge and adaptability to produce accurate cost estimates within a shorter timeframe leading to more reliable decision-making in project management. This advantage is particularly valuable when limited project information is available (Matel et al., 2019). The adoption of advanced computational methods has become indispensable in addressing the complexities and challenges encountered in construction projects. These challenges serve as strong motivation for the application of intelligent techniques to effectively manage them. Intelligent techniques can be employed to address various challenges, including selecting the most suitable prime contractor, predicting project performance at different stages, and estimating the risk of cost overruns. In the field of civil engineering, Artificial Intelligence techniques are increasingly recognized as a valuable approach to tackle problem areas characterized by uncertainty and ambiguous definition (Elfaki et al., 2014).

**Problem Statement**

Cost overruns pose a significant challenge, particularly in the current landscape of stringent budget constraints. Such overruns can have severe consequences, including the potential cancellation of the project (Elmousalami, 2020).

During the tendering phase, project details and information are limited, whereas traditional estimation methods rely heavily on detailed information, resulting estimators to use their expertise and experience to make intuitive judgments. This approach can be time-consuming and costly (Matel et al., 2019). Furthermore, such qualitative approaches that rely on expert judgments may produce bias and yield inaccurate estimations. Many models have been proposed for construction cost estimation. However, various models primarily focus on project characteristics and overlook external economic factors (Wang et al., 2022).

This outlines that traditional estimation techniques using judgement and experience are no longer effective compared to Artificial Neural Network (ANN) models. Factors affecting cost estimation are the backbone of any ANN model, as they characterize the input variable that shapes the output variable. The chosen factors should have the biggest impact in affecting cost estimations, where the ANN model will get trained on these factors and will reduce their uncertainty.

The current body of knowledge focus on basic factors such as project locations, various elements of the structure, with minor considerations to the construction index, usage of advanced technology, linking usage of Building Information Modeling (BIM) and advanced technologies to mitigate delays and abortive works, social environment, and international market conditions. These neglected factors could have higher impact on cost estimation which is why ANN should not be trained only on structural elements factors such as number of stories, type of foundation, etc.

**Research Aim & Objectives**

The aim of this thesis is to identify and prioritize key factors that have a significant impact on cost estimation for building projects in the United Arab Emirates (UAE). These factors will have value as they represent the input variables for the next stage where an ANN model will be developed with the aim of providing an efficient and accurate cost estimation method. The research will involve an extensive analysis of existing literature, as well as conducting a survey and interviews to develop the survey results. By identifying the most influential and relevant factors, this thesis seeks to enhance the accuracy, efficiency, and interpretability of ANN models specifically tailored for the construction industry. The findings of this research will benefit various stakeholders involved in construction projects, such as owners, contractors, and consultants, by enabling them to determine parameters that affect cost estimation and can cause project cost overruns. The ultimate goal of this study is to prepare input parameters that will enable to provide valuable insights for future development of ANN models in the field of construction.

The Main objectives of this thesis can be defined under the two points below:

1. Identify new key factors affecting the accuracy of estimation of building project within the UAE and similar markets with focus on structural factors and new prospective factors such as safety requirement, type of contract, usage of advanced techniques, etc.
2. Conduct a comprehensive analysis for the determined key factors, followed by usage of tools for the development of data collection using various methods such as pilot testing and expert reviews that will be further discussed in chapter three. Expert reviews are conducted with high management professionals such as planning manager, contracts manager, technical manager, estimation manager and general manager to prepare for the development of the optimum Neural Network model in the next phase.

**Significance of the Study**

The thesis analyzes cost factors by weighting and rating of the major structural works in building projects and other economic factors such as high safety requirement, cost index, usage of BIM, stakeholder characteristics, etc.
The anticipated contributions of this thesis are expected to have relevance to both researchers and practitioners in the following ways:

1. For researchers, the findings will serve to emphasize the importance of exploring new parameters in cost estimation in order to assess the accuracy of Artificial Neural Network models across of building projects. By identifying and investigating these parameters, this research will contribute to expanding the knowledge and understanding of ANN model performance in the construction domain.
2. For practitioners, the findings will provide valuable insights to enhance the accuracy of cost estimation practices. By highlighting overlooked parameters that can lead to underestimation or overestimation of project costs, practitioners will be better equipped to conduct more precise estimate jobs. For instance, the use of BIM can offer a clearer deconstruction of the project compared to traditional 2D drawings, thereby reducing errors. However, it should be noted that the incorporation of BIM may come with additional costs related to the elaboration of the BIM model. It is the case that many estimators fail to precisely estimate the requirement to produce such models resulting in unpredicted costs, delays because of necessary qualified professionals and more serious implications.

Overall, this research aims to bridge the gap between research and practice by offering practical implications for practitioners while contributing to the advancement of knowledge in the field for researchers.

**Research Scope and Limitation**

This research primarily concentrates on gathering surveys from professionals who possess significant experience in building projects within the UAE. The scope of the study specifically focuses on the structural package and other economic factors. Consequently, the results obtained may not be directly applicable to other civil projects such as roads and infrastructure. Additionally, the unique characteristics of the UAE market and the concentrated expertise of professionals in this market may limit the generalizability of the findings to other markets. To gain a better understanding of different markets such as Europe, USA, and Asia, further evaluation and research would be required. It is important to recognize that each market has its own distinct features, regulations, and industry practices, which may influence the factors affecting cost estimation. Therefore, conducting specific studies in those regions would be beneficial for obtaining a comprehensive understanding of the factors relevant to cost estimation in different markets.

## Research Methodology

The aims of this research will be accomplished by executing the following procedures:

1. Conduct an extensive review of the literature pertaining to construction cost estimation, with a particular emphasis on the utilization of Artificial Neural Networks model.
2. Employ quantitative and qualitative survey techniques to identify the significant factors influencing the cost of building projects.
3. Integrate tools for development of data collection of survey results. It is based on exploratory interviews with various engineering positions to discuss data results and aim to reduce bias and properly highlight cost estimation key factors in the UAE market.

4. Deliver recommendations and conclusions regarding the parameters that should be considered in the subsequent phase of developing ANN models.

By implementing these steps, this study aims at enhancing the understanding of cost estimation in construction projects, specifically in the UAE market, and contributes to the development of ANN models for accurate cost estimation.

**Research Structure**

The thesis presented is structured into five chapters, which are outlined as follows:

Chapter 1: Introduction

The first chapter starts by a background explanation articulating a wide overview on traditional and computational estimation techniques advantages and inconvenient, followed by the problem statement, research aim and objectives, methodology, significance of the study and an outline of the thesis structure.

Chapter 2: Literature Review

In this chapter, an extensive examination of the existing literature related to cost estimating is presented. The review encompasses various aspects, including definitions, methods, processes, and techniques utilized in cost estimation. Furthermore, the chapter delves into the historical background of Artificial Neural Networks (ANNs), providing definitions, highlighting their advantages, discussing their structure, comparing them to conventional methods, exploring different types of ANNs, examining model training, and testing procedures, analyzing performance measures, evaluating sensitivity, and addressing potential challenges associated with ANNs. The chapter concludes by depicting applications review of ANN in the construction field in recent years.

Chapter 3: Research Methodology

The third chapter describes the methodology adopted for developing the thesis, including the process of acquiring data on key factors affecting cost estimating in building projects. It then briefs the tools for data development.

Chapter 4: Data Results

In this chapter, survey results statistical analysis is presented. Then, tools for data development used during study are discussed to confirm the data reliability.

Chapter 5: Conclusion and Recommendations

The final chapter summarizes the conclusions drawn from the study and provides recommendations for the next phase in which an ANN model will be build and cost estimation model at early stage will be the end result. It also outlines areas for further research and development.

# CHAPTER 2

# LITERATURE REVIEW

## Chapter Introduction

The objective of this chapter is to provide a review of the literature used in the thesis. The chapter communicates cost-related definitions, purpose, accuracy, and types of cost estimation. It also depicts the estimating process and its methods and describes the classification of construction costs. The chapter concludes by providing Artificial Neural Networks (ANNs) history, definition, structure, types, problems and challenges, and applications of ANN.

## Cost Engineering

Cost engineering can be defined as applying scientific principles and methods along with engineering expertise and judgment, cost engineering to address a variety of issues relating to estimation, cost control, business planning, and profitability analysis (Jinisha & Jothi, 2019).

## Cost Estimate

A cost estimate may be defined as the approximation of an operation, project, or program cost. The cost estimate is determined from a cost estimating process. The cost estimate may represent a single total value or may have many identifiable component values.

According to the Association for the Advancement of Cost Engineering International (AACE), Cost estimate provides the basis for project management, business planning, budget preparation, and cost and schedule control (AACE, 2020).

**Purpose and Accuracy of Cost Estimate**

Cost estimation accuracy is an essential factor for ensuring the success of any construction project. One of the major problems is cost overruns, especially with the current priority on tight budgets. A number of literature estimate that one in four projects can face overbudget issues. Many of the shelving or cancellation of projects are led by cost overruns (Feng & Li, 2014).

Hanna et al. (2004) indicated that the recommendation to eliminate the top three most common reasons for change orders, design errors, design changes, and additions, is to invest additional time and budget resources ahead of construction (Hanna et al., 2004).

Construction early-stage cost estimation, along with its accuracy, is one of the major challenges for decades. Initially, limited information is available, the estimation process becomes a very hard task for estimators and project engineers. An appropriate estimation can provide a clear vision for financial management in designing the budget for investors, hence making better decisions. It also provides project managers with the ability to manage their available resources and cash reserve funds during the entirety of project execution stages (Chandanshive & Kambekar, 2019). According to Kim et al. (2004), one of the main elements of decision making at the preliminary stage of construction is cost estimate. Hence, improved techniques will provide improved control of cost and time.

Factors such as construction inputs' dynamic nature of market prices and cost factors unique to individual construction projects, make accurate estimation of the cost of construction projects challenging. The early cost estimate demand increases the problem complexity because the final cost function is not possible at the early stages of planning and design. For example, public agencies are often challenged in the process of fine-tuning the accuracy of their early cost estimates, which may lead them to choose simplified traditional models over relatively more extensive, complex approaches (Karaca et al., 2020).

Accurate cost estimation enhances contracting and the success or failure of a project depends on cost estimation. This is influenced by major factors such as (Ibrahim & Elshwadfy, 2021):

- Scope of the project, community interest and macro-economy accuracy reflected in estimation.
- Quality of assumptions used and the accuracy of bill of quantities for preparing the estimate.
- Years and quality of experience of the estimating team
- Construction method/ techniques/technology used in estimation.

Vikas et al. (2011) outlines the purpose of cost estimate as the following:

1. Form the basis for planning and control by defining the scope of work and its associated estimated cost.

2. Determine what resources to commit to the project by providing much of the basic information (hours, resources, tasks, and durations) which is needed for preparing a schedule.
3. Projects can be easier to manage and control when resources are better matched to real needs.
4. Provides the financial input required to prepare a cash flow curve.
5. Client expects actual development costs to be in line with estimated costs.
6. Can be used to assess the impact of changes and support during re-planning.

**Types of Construction Cost Estimates**

According to Len Holm (2021) and Jinisha & Jothi (2019), estimate types often correlate with project design phases and are labeled as follows:

1. Conceptual design budget estimate: It is the rough approximation of estimating involving the prediction of the future costs of a project. It is prepared for information purposes only, and it precedes design drawings.
2. Schematic design budget estimate: It is an approximation based on well-defined cost data and established ground rules, prepared for allowing the owner to review the design before details.
3. Design development estimate: It is based on the detailed design where all drawings are ready, prepared to ensure the design is within financial resources and it assists in bids evaluation.
4. Construction document estimate: This is done by the contractor during the bid phase to price the contract.

According to Amade et al. (2015), the probability of the use of each of the estimate methods depends on the ease of its application, familiarity, and effectiveness along with a tolerable level of accuracy and reliability. They include, but are not limited to:

- Functional Unit: Also called unit-price. This method involves the usage of a single functional unit multiplied by the number of units used.
- Cube Method: Multiplying plinth area with the height of the building. The height of the building should be considered from floor level to the top of the roof level. It is more suitable for multi-storied buildings.
- Superficial Area: It is one of the most popular preliminary estimating methods. It is an approximate cost obtained by multiplying the area by the cost per square meter/feet.
- Superficial Parameter: It entails the analysis of cost, programmatic and technical data to identify cost drivers and develop cost models.
- Approximate Quantities: It is an approximate quantity method cost estimate, the total wall length of the structure is measured, and this length is multiplied by the rate per running meter which gives the cost of the building.
- Elemental Analysis: Since the 1950s, quantity surveyors have used this technique to base their predictions during the design stage. Total costs should be provided for each element and sub-elements as appropriate. Costs should be shown separately where required in the elemental definitions and for different forms of construction. The cost of the elements should total to the contract sum minus main contractor's profit, where identified; preliminaries; contingencies and, where appropriate,

contractor's design fees. The cost of each element and the items comprising it should correspond with the specification (BCIS, 2012).

Estimate methods are classified differently by various literature. Classified estimate method types are (Samphaongoen, 2010):

1. Conceptual cost estimate during the schematic design phase and the error percentage reaches to 20%
2. Semi-detailed cost estimate during the detailed design phase and the error percentage is between 5-10%.
3. Detailed cost estimate during the issue for construction phase and the error percentage decreases to less than 5%

The first estimate type guideline was developed in 1958 by AACE Estimating Methods Committee and proposed four types:

1. Order of Magnitude estimate
2. Preliminary estimate
3. Definitive estimate
4. Detailed estimate

All these types are based on four estimate characteristics: purpose, accuracy, information available for estimating, and methods. The main tenet in estimation is a level of scope definition increased which improves the accuracy of the estimate. While the specifics have evolved, the general concept of classification or phased estimates remains the same.

## Estimating Process

Estimation is one of the essential steps in the project management process. The process of estimating refers to the procedure of approximating or calculating the costs, resources, time, and other factors associated with a particular project (Elfaki et al., 2014). As per Project Management Institute (PMI), the estimating process involves "developing an approximation or estimate of the costs of the resources needed to complete project activities" (Project Management Institute, 2017, p. 246). The Project Management Body of Knowledge (PMBOK) also emphasizes the importance of updating estimates throughout the project life cycle to reflect changes in project scope, risk, and other factors.

Elfaki et al. (2014) states that the process of estimating costs is of utmost importance in the initial stages of any construction project. As a result, the field of construction management has invested significant research efforts in the realm of construction cost estimation.

## Classification of Construction Costs

RAD (2002) states that construction costs are generally classified into six categories: direct and indirect costs, overhead, contingency, allowance, and profit. Kraus (2007) describes tools for mutual understanding classification of construction costs, specifically overhead, contingency and allowance. It is generally agreed that contracting is a job for the lowest offer, that is why it is important that contractors should understand calculating the indirect cost, profit, and overhead. Figure 1 below simplifies the classification.

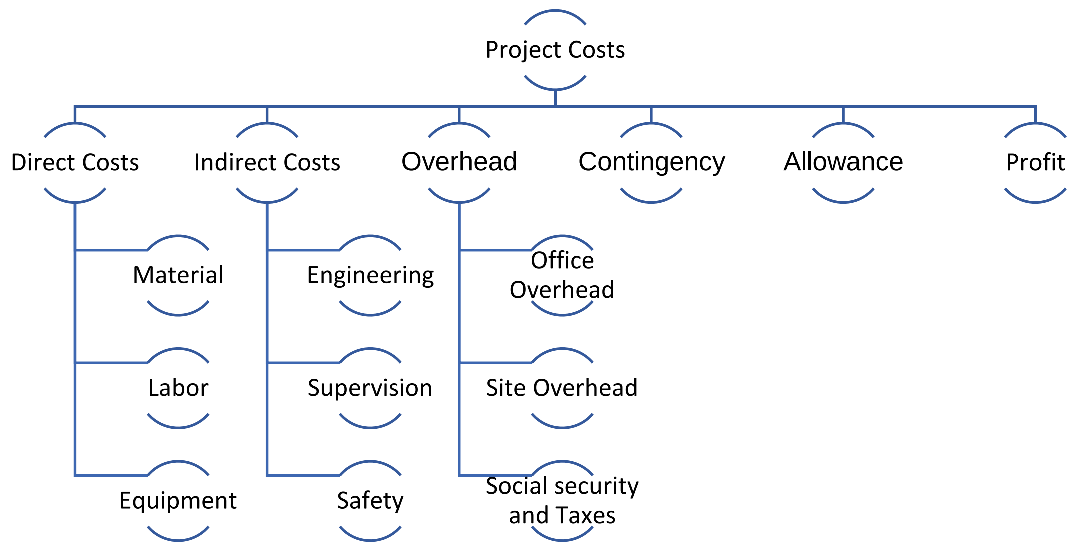



*Figure 1: Project Costs Classification*

***Direct Cost***

RAD (2002) clarifies the direct cost as the costs directly related to the achievement of the project (i.e., masons constructing a wall). The three main categories inside direct cost are labor, material, and equipment. This includes the salary, transportation, and the purchase or rental of equipment specifically for the project.

**Material Cost.**

According to Shehatto (2013), the material cost includes direct cost of the material items, transportation, sales taxes and freight costs, delivery, storage, sales and other taxes and losses.

#### Labor Cost.

Labor cost is the direct cost of the labor to the good achievement of a task or a project. It includes the wage rate, transportation of labor, their food and rest time. This cannot include supervision, or any other cost not associated directly with achieving the work (D. Grau, personal communication, March 28, 2022).

#### Equipment Cost.

It is the cost of rented or owned equipment, including cost of loading, transportation, unloading, erection, maintenance, fuel, dismantling and removal. Equipment used to transport labor or material is classified under the material or labor cost and not considered as equipment cost.

According to Shehatto (2013), there are two types of equipment, general use and specific use:

- General use equipment: Equipment that has shared utilization by many trades on the construction site and it is not associated with any particular task. i.e., Crane
- Specific use equipment: Equipment specific for a construction operations task and is removed from the jobsite soon after the task is completed. i.e., Marble Floor Polisher

### *Indirect Cost*

Indirect cost in construction are costs that lead to the successful achievement of project but are not directly related to project. In other terms, indirect costs can include but are not limited to engineering, supervision, management cost, equipment for support

(Scaffolding), sick leave, vacation, training, and even retirement benefits for the employees. (RAD, 2002)

Kraus (2007) states that indirect cost can be considered as site overhead and is defined as per AACE International as: All non-direct costs necessary to properly perform the installation and may include field administration, direct supervision, capital tools, startup costs, contractor's fees, insurance, taxes, etc.

### *Overhead*

Overhead costs are the costs of running a business, they are not specific and vary from one company to other for what is considered. RAD (2002) states that overhead costs can include compensation of the company senior management and the cost of infrastructure necessary for supporting project activities. He also debates items such as the cost of preparing unsuccessful proposals, marketing and public relations, and ongoing innovative ventures of the organization are included.

Kraus (2007) defined overhead as per AACE International's Recommended Practice No. 10S-90, A cost related to performing a task or a project that cannot be assigned or attributed to any part of the project. He classifies overheads into 2 sub-categories, general overhead, and site overhead.

General overhead could be costs such as office, plant, equipment, staffing, and expenses that are essential to be maintained by a contractor for general business operations. According to Kraus (2007), the difference between general overhead and site overhead is often misunderstood. Site overhead are indirect cost and is defined in above section.

Apanavičienė & Daugėlienė (2011) stated that contractors often fail to evaluate properly the actual overhead costs, which represent the largest part of indirect costs of construction. The authors provided figure 2 below to simplify the matter.

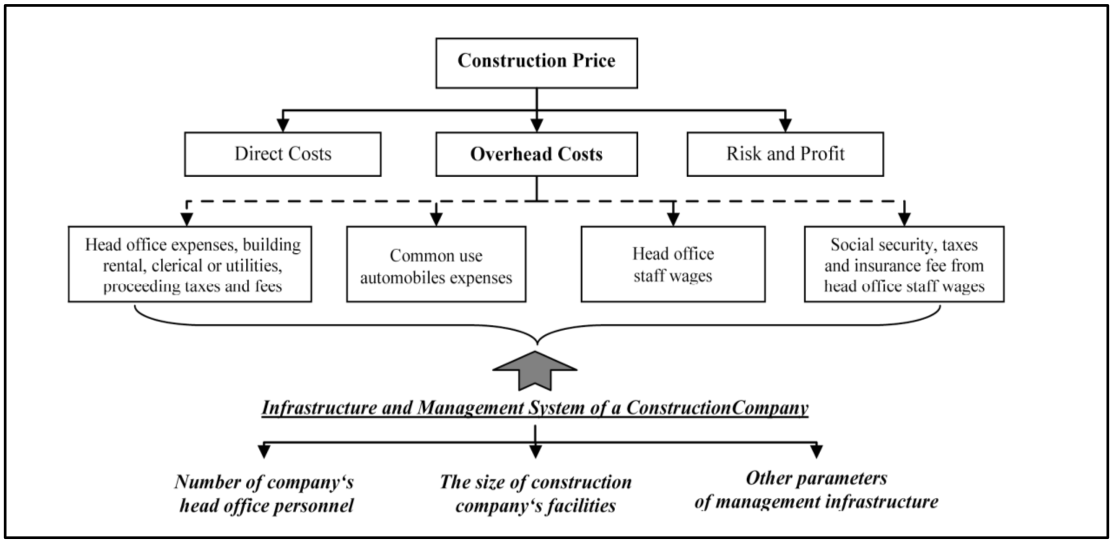



*Figure 2: The Structure of a Construction Company's Overhead Costs. Extracted From (Apanavičienė & Daugėlienė, 2011)*

### *Contingency*

Kraus (2007) defined Contingency as per AACE International as determined amount or percentage added to estimates to account for items, conditions, or events whose condition, occurrence, and/or impact are uncertain and likely to result in overall additional costs. The percentage depends generally on previous experience and volume of inaccuracies in the estimate caused by uncertainties in project details (RAD, 2002). It is generally included in most estimates and is expected to be expanded by the direction of the management (Kraus, 2007).

### *Allowance*

RAD (2002) states that allowance and contingency are akin to each other, only that allowance is defined as lump sum amount reserved for certain project task that are not calculated in a detailed estimate but is known to occur through the project. Contingency is an amount set for inaccuracy in estimate or change in price from tender stage to the procurement stage.

Kraus (2007) further explains allowance as per AACE International as “additional resources included in estimates to cover the cost of known but undefined requirements for an individual activity, work item, account, or subaccount.”

### *Profit*

Kraus (2007) explains profit as the amount included in a contractor’s tender price to compensate the contractor for undertaking the risks associated with the project and to provide a return on its investment. He further states that overhead and profit are frequently combined in a contract sense and referred to as “contractor’s overhead and profit”.

Shehatto (2013) defines profit as a percentage of the total contract price, or in some cases, as a percentage of each task in the project. The percentage of profit is set by the higher management or owners for each individual tender, depending on local market conditions, competition, risk, and the contractor`s need for new projects.

## Cost Estimation Techniques

Numerous literatures provide several techniques of cost estimates that are used in different industries and fields. Shehatto (2013) stated three main categories for cost estimation techniques:

1. Quantitative and Qualitative Technique
2. Preliminary and Detailed Techniques
3. Traditional and Artificial Intelligence Based Techniques

***Quantitative and Qualitative Technique***

Quantitative techniques rely on data and mathematical modeling, while qualitative techniques rely on expert judgment and intuition. The decision of usage of which technique to be used depends on the availability of data, the complexity of the project, and the level of accuracy required in the cost estimation (Juszczyk et al., 2018).

**Quantitative Technique.** refers to an approach that relies on numerical data and mathematical analysis to estimate the cost of a project (Ary et al., 2018). It depends on historical data, data collection and analysis using statistical and mathematical models to predict the cost of a project with a high degree of accuracy (Juszczyk et al., 2018). Examples of quantitative cost estimation techniques include parametric modeling, bottom-up estimating, three-point estimating, Earned Value Management (EVM) and Monte Carlo Simulation.

***Parametric Modeling.*** is a type of quantitative cost estimation technique used in construction that relies on analytical function and mathematical models to estimate project costs based on a set of parameters or factors. According to Chan and Kumaraswamy (1997), parametric models "estimate costs based on regression analysis, using selected cost drivers (parameters) to develop a mathematical formula for estimating the cost of a project or activity" (p. 72).

Construction projects with repetitive or standardized work may benefit from these models, as they can provide a more efficient and accurate estimation process. For a more comprehensive and accurate estimate of project costs, parametric models are often combined with expert judgment and historical data analysis (Lim & Mohamed, 2013).

***Monte Carlo Simulation.*** is a computational technique based on quantitative cost estimation. It models and analyzes uncertainties and risks associated with construction projects, allowing for more accurate cost estimation, schedule planning, and risk assessment. According to Kelton et al. (2010), Monte Carlo Simulation is defined as "a technique for conducting probabilistic simulations of deterministic models" (p. 3).

In construction, the project's key parameters for Monte Carlo Simulation mathematical model can include variables such as project duration, costs, resource availability, and other relevant factors. (Al-Sabah et al., 2016)

**Qualitative Technique.** entails an approach that prioritizes subjective interpretation, understanding, and exploration of data and experiences, including expert judgment, experience, interviews, observations, and intuition, to estimate the cost of a project (Juszczyk et al., 2018). It is often used when there is limited data available or when the project is new or innovative. Examples of qualitative cost estimation techniques include expert judgment, analogy-based estimating, and the Delphi method.

***Analogy-based technique.*** is a cost estimation method that relies on comparing history of similar past completed projects with the project being estimated. It is often used when there is limited information available about the new project, and the historical data of similar projects used as a basis for estimating the cost of the new project (Mendes et al., 2003; Khosrowshahi, 1994). Various literatures proposed different variations of analogy-based techniques, including the following:

1. Analogy-based parametric models (ABPMs). These models rely on identifying the key cost drivers of the project and using them to develop a mathematical formula that can be applied to estimate the cost of similar projects. (Ruwanpura & Bandara, 2009)
2. Analogy-based cost estimation (ABCE): ABCE is a technique that involves using a database of historical cost data to identify similar projects and estimate the cost of new projects based on their similarity. This technique involves a manual search process to identify relevant analogies that may require some level of expert judgment (Nguyen, Deeds, & Carpenter, 2014).
3. Analogy-based effort estimation (ABEE): ABEE is a technique used to estimate the effort required to complete a software development project based on the similarity of the project to past projects. This technique involves identifying and measuring the characteristics of past projects and using them to estimate the effort required for the new project. (Mendes et al., 2003)

***Delphi Method.*** is described by Elmousalami (2020) as a structured communication technique used to gather the opinions of experts regarding a specific case, with the aim of identifying all the parameters that impact the system. It involves a series of iterative rounds of collecting, ranking, and revising the parameters based on the feedback received from the experts. Experts provide anonymous feedback and revise their opinions based on the feedback from other participants to improve the quality of the survey. The Delphi rounds continue until a consensus is reached and no other opinions remain (Elmousalami, 2020).

Davé (2003) adds that "Traditional Delphi Method (TDM) is an anonymous, written, iterative survey method used to develop consensus among experts regarding a specific topic" (Davé, 2003, p. 135).

***Preliminary and Detailed Techniques***

A preliminary estimate is an approximation that relies on specific cost data and predefined guidelines. It allows the owner to review the design before proceeding with further details. The accuracy of this estimate is typically within a range of -30% to +50% (Jinisha & Jothi, 2019).

Detailed estimation techniques involve a series of methods and procedures that are employed to obtain an accurate cost estimate of a project based on analysis of detailed design specifications and drawings. These techniques entail deconstructing the project into its constituent elements, quantifying the required resources, considering factors such as labor, materials, equipment, and overhead costs, and incorporating current market rates and other pertinent variables to ensure accuracy in cost estimation (Jha & Bhandari, 2016).

### *Traditional and Artificial Intelligence Based Techniques*

The traditional percentage model is arbitrary and difficult to justify (Kwon & Kang, 2019). Günaydın & Doğan (2004) stated many types of traditional methods such as traditional detailed breakdown cost estimation; simplified breakdown cost estimation; cost estimation based on cost functions; activity-based cost estimation; cost index method; and expert systems. Other types can be found in different literatures such as analogue estimating (CII, 1996), and Three-Point Estimating (Fleming & Koppelman, 2016). Other methods are Monte Carlo simulations, and a third method is the regression model. These are advanced statistical tools using analytical predictions in forecasting the final cost of a project. These methods lack consideration of estimating risk costs (Kwon & Kang, 2019).

Alternative approaches were introduced recently using the concept of parametric models based on computerized techniques, such as fuzzy expert systems (also called fuzzy logic) and Artificial Neural Networks (ANN), which have been utilized in creating a model for estimating project cost contingency. These models are well-suited for handling non-linear data modeling, in contrast to linear methods like regression. While they prove valuable in assessing risks and their associated probabilities, they exhibit limited effectiveness when it comes to estimating cost contingency. (Kwon & Kang, 2019)

## Artificial Neural Networks (ANN)

Shehatto (2013) summarizes that Artificial Neural Networks (ANN) draw inspiration from the structure and functioning of neurons in the human brain. The brain demonstrates remarkable abilities to perform complex tasks with relative ease compared to computers. Consequently, researchers sought methods to integrate this intelligence into machines to enable them to handle such tasks effortlessly. ANNs emulate the human brain in two significant ways: the network acquires knowledge through a learning process, and the interconnected nodes (artificial neurons), known as synaptic weights, store this knowledge (Shehatto, 2013).

Elmousalami (2020) states that to achieve the desired level of accuracy in estimating construction projects, it is crucial to overcome significant challenges, especially during the early stages. These challenges primarily arise from the lack of preliminary information and increased uncertainties due to engineering design. To address the limited availability of detailed data, cost estimation techniques are employed to provide an approximate cost estimate that falls within an acceptable range of accuracy. (Elmousalami, 2020)

Shehatto (2013) states that not long ago, new artificial intelligence approaches, such as expert systems, Case-Based Reasoning (CBR), ANN, Fuzzy Logic (FL), Genetic Algorithms (GAs), and others, have found applicability in addressing cost estimation challenges.

***Historical Background of ANN***

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

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

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

### *Definition of ANN*

Artificial neural networks, commonly known as ANN, are computational models that draw inspiration from the functioning of the nervous system, specifically the behavior of neurons(Alcineide et al., 2021). ANN can be described as mathematical structures and their corresponding implementations, which encompass both hardware and software, and are designed to mimic and derive insights from the observed behavior of natural nervous systems (Juszczyk et al., 2018).

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

### *Advantages of ANN*

Artificial Neural Networks are among the innovative techniques that excel in handling incomplete data sets, fuzzy or incomplete information, and highly complex and ill-defined problems. ANNs have the capability to learn from examples and effectively address non-linear problems. A notable feature of ANN is their capacity to learn from experience and adapt to dynamic situations. They possess a natural ability to store experiential knowledge and make it readily accessible for application (Shehatto, 2013). Matel et al. (2019) stated that in the early design stage of construction projects, ANNs are widely used to estimate project costs and duration. ANN are beneficial because they can self-learn, which saves development time, and they can identify non-linear relationships between cost factors and project cost without extra effort. By using an ANN model, accurate predictions can be obtained, even when there is limited information available during the early stages of the design process.

Wang et al. (2022) noted that the regression analysis method is characterized by its simplicity and the generation of straightforward predictions. However, it has limitation that its reliance on a predetermined mathematical form, which restricts its suitability for datasets exhibiting high nonlinearity. Similarly, Support Vector Machine, Decision Tree, and Random Forest methods face the challenge of overfitting when applied to regression problems contrary to ANNs which have the ability to handle datasets that exhibit strong non-linear relationships between outcome variables and predictors, thereby providing more precise outcomes.

Juszczyk et al. (2018) summarized advantages of using ANNs in cost estimating problems, particularly in the construction, as follows:

1. ANN are well-suited for regression problems where understanding the relationships between the dependent variable and numerous independent variables is challenging.
2. ANN have the capability to acquire knowledge through automated training processes, eliminating the need for extensive manual investigation.
3. ANN can build and store knowledge based on the patterns observed in real-life training examples, allowing them to learn from experience.
4. ANN exhibit the ability to generalize knowledge, enabling them to make predictions for data that were not specifically included in the training process.

***Neural Networks versus Conventional Methods***

Matel et al. (2019) stated that ANNs work more accurately than Multiple Regression Analysis (MRA) and Case-Based Reasoning system (CBR) estimating models. According to Wang et al. (2022), Artificial Neural Networks (ANNs) possess the capability to effectively handle datasets that demonstrate complex non-linear relationships between outcome variables and predictors. As a result, ANNs offer more accurate outcomes in contrast to the limitations of the regression analysis method, which relies on predetermined mathematical forms that may not be suitable for highly nonlinear datasets. Additionally, other methods such as Support Vector Machine, Decision Tree, and Random Forest encounter challenges related to overfitting when applied to regression problems.

Elmousalami (2020) outlines that unlike conventional modeling methods, such as linear regression analysis, ANNs have the capability to approximate nonlinear functions with a desired level of precision. Shehatto (2013) shares differences between conventional computer method and ANNs by highlighting that conventional method tackles one task at a time, with no relation or experience between each task and use cognitive approach to provide solutions. Thus, problems should be clearly identified and articulated using concise and clear instructions that leave no room for ambiguity. ANNs on the other hand do not follow a sequential or deterministic approach. They lack complex central processors and instead consist of numerous simple processors that primarily calculate the weighted sum of their inputs from other processors. In ANNs, a vast number of interconnected processing elements (neurons) operate in parallel to address a specific problem, and they have the ability to learn through examples. Unlike traditional programming, ANNs cannot be explicitly programmed to perform specific tasks. (Shehatto, 2013)

***Neural Network Structure***

According to (Najafi & Tiong, 2015), the architecture of an artificial neural network (ANN) consists of individual neurons that comprise two main components: a summing junction responsible for aggregating inputs received from neighboring neurons, and an activation function that computes the output signal, which is then transmitted to other neurons. The activation function can take various forms, including signum, linear or semi linear, hyperbolic tangent, and sigmoid functions. Neurons are organized into different layers to form a network, namely the input layer, hidden layers, and output layer (Najafi & Tiong, 2015).

According to Elmousalami (2020), there are various types of activation functions that serve different purposes within artificial neural networks. These include the linear function, step function, ramp function, and tangent sigmoid function. The Selection of parameters in

ANN, such as the number of neurons, connections, transfer functions, and hidden layers, relies on the application of the ANN. Elmousalami (2020) defined various architectures of feedforward neural networks, such as multilayer perceptron networks (MLP), radial basis function networks, generalized regression neural networks, probabilistic neural networks, belief networks, Hamming networks, and stochastic networks. The choice of architecture depends on the specific problem at hand as each architecture is tailored to address different types of problems.

Shehatto (2013) says determining the appropriate number of hidden layers and neurons in a neural network is often considered a challenging aspect due to the lack of specific rules. It typically involves multiple trial and error iterations, requiring a significant amount of time and effort. Elmousalami (2020) confirms Shehatto by stating “No exact rule exists to determine the number of hidden layers and neurons in the hidden layer”. He then describes a neural network consisting of three layers, namely the input layer, hidden layer, and output layer, capable of addressing various prediction, approximation, and classification tasks. It is important to have a sufficient number of training cases compared to the size of the network to prevent overfitting and improve its ability to generalize. The learning process of artificial neural networks involves adjusting the weights and biases of the network to minimize the error observed during training (Elmousalami, 2020).

Elmousalami (2020) further explained, developing artificial neural networks can be summarized as follows:

1. Gather and preprocess historical cases.
2. Split the collected cases into a training set and a validation set.
3. Identify the relevant inputs and outputs for the network.

4. Determine the appropriate number of hidden layers.
5. Determine the optimal number of neurons in each hidden layer.
6. Choose the suitable transfer function for the network.
7. Initialize the weights of the network.
8. Select the learning algorithm to update the network's weights.
9. Train the model through multiple iterations to minimize the prediction error.
10. Evaluate the resulting error metrics such as mean absolute percentage error (MAPE).

Figure 3 below presents a schematic diagram of a hidden neuron, showing weight's summation part and transfer function part inside these neurons.

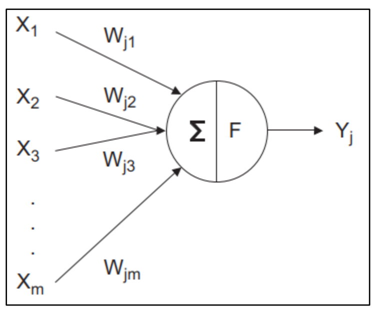



*Figure 3: Schematic Diagram of a Neuron. Extracted From (Najafi & Tiong, 2015)*

Figure 4 depicts the structure of ANNs as described in (Jinisha & Jothi, 2019). Each layer consists of several neurons except output layer where it represents the one neuron for output of training process.

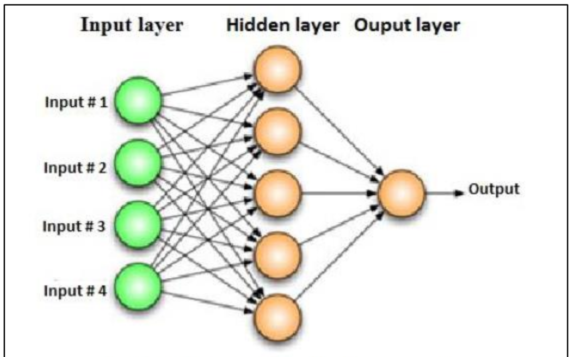



*Figure 4: The Structure of Artificial Neural Network. Extracted From (Jinisha & Jothi, 2019)*

### *Terminology Used In Artificial Neural Network*

The architecture and functionality of ANN are described using a particular group of terminology. For an understanding of how ANN function within themselves, it is fundamental to comprehend these terminologies.

**Neurons**. Also referred to as nodes, are fundamental processing units. These terms are often used interchangeably. Neurons receive inputs, carry out computations, and generate an output.

**Weight**. In the context of Artificial Neural Networks (ANNs), the term "weight" pertains to the parameters that gauge the strength of the connections linking neurons. These weights are responsible for determining how much influence each input has on the output produced by a neuron. (Chakraverty et al., 2019; Shehatto, 2013)

**Hidden Layer.** also referred to as “Intermediary layer”, is an essential part of the structure of any network. Günaydin & Doğan (2004) describes the hidden layer's role is to identify and retain relevant features and sub-features from the input patterns, enabling the network to make predictions about the output layer's values.

**Learning Algorithm**. learning can be defined as the method of assigning appropriate values to weights. It is generally associated with the training process; training refers to the process of attaining the desired output by adjusting the weights in the connections between layers of a network through multiple iterations. This allows the network to learn and improve its performance to accurately process information and produce the required results (Chakraverty et al., 2019). Haykin (2009) describes learning algorithms as a school process with two main categories of learning: supervised learning (learning with a teacher), and self-learning, divided into reinforced learning and unsupervised learning.

***Supervised Learning Method***. Najafi & Tiong (2015) defined the network is provided with input and output training pairs. The sample pairs presented to the network teach it to make predictions or produce desired outputs based on the given inputs. Haykin (2009) defines the teacher, equipped with knowledge of the environment, guides the neural network using input-output examples. The network adjusts its parameters based on the training data and error signal, aiming to emulate the teacher's knowledge stored in fixed synaptic weights as long-term memory. According to Shehatto (2013), supervised learning encompasses several learning algorithms, including the Back-propagation Learning rule, Gradient Descent Learning, and Delta Rule. Among these, the Back-propagation Learning rule is the most widely used method and offers different algorithms such as Levenberg-Marquardt and Momentum for implementing the backpropagation algorithm.

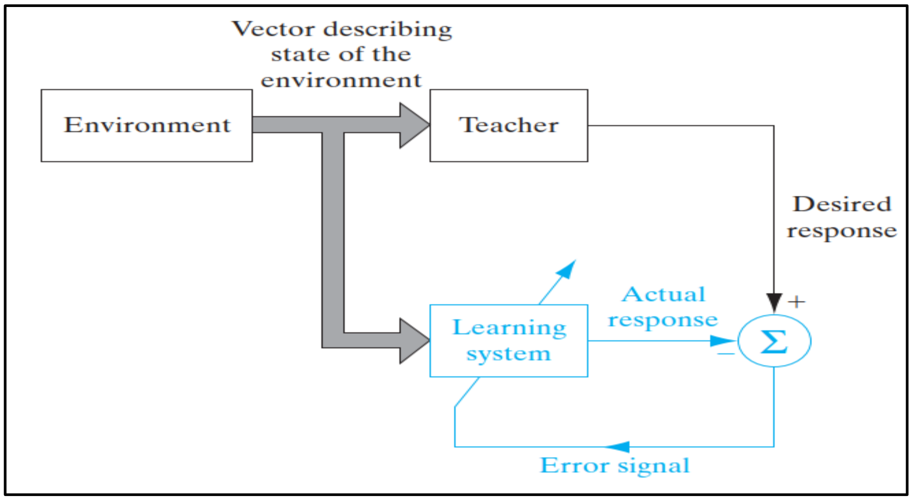



*Figure 5: Diagram of Learning With a Teacher. Extracted From (Haykin, 2009)*

***Unsupervised Learning Method.*** It is where the network does not have a predefined set of categories to learn from. Instead, it independently learns and develops its own representation of the input stimuli (Haykin, 2009). Chakraverty et al. (2019) describes unsupervised training in a neural network is performed without a teacher where the target output is unknown for training the input vectors. The network adjusts weights to group similar input vectors together. Implementing unsupervised training is complex and challenging.

***Reinforcement Learning.*** It is a machine learning approach where a network learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or punishments. Through a process of trial and error, the network aims to discover the optimal actions to maximize its cumulative reward (Sutton & Barto, 2018). Researchers like Mnih et al (2015) suggest reinforcement learning is often implemented using algorithms such as Q-learning, SARSA, or Deep Q-Networks (DQNs) (Mnih et al., 2015).

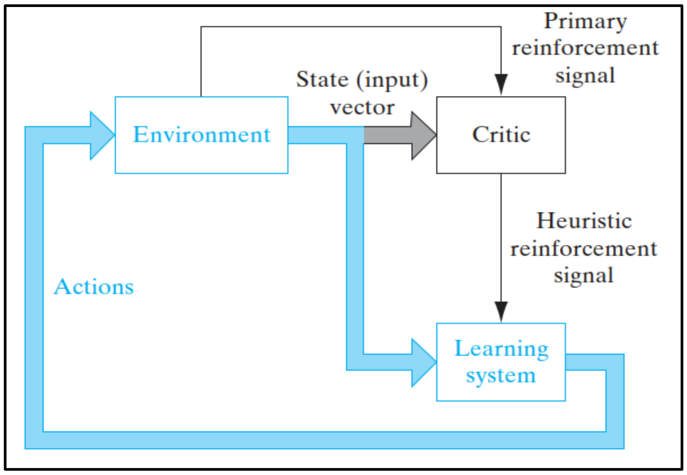



*Figure 6: Diagram of Reinforcement Learning. Extracted From (Haykin, 2009)*

**Activation Functions.** Sometimes referred to as transfer function. It refers to a mathematical function applied to the input of a neuron to determine its output. The activation function introduces non-linearity to the network, enabling it to learn and approximate complex relationships in the data (Aggarwal, 2018). Activation functions play a crucial role in capturing the non-linear relationships between input and output in neural networks. Different types of activation functions, including sigmoid, tanh, relu, and linear, are used to achieve this mapping of non-linearity (Wang et al., 2022). Najafi & Tiong (2015) suggest that the activation function in neural networks can take various forms, including signum, linear, semi-linear, hyperbolic tangent, and sigmoid functions.

| ANN | Number of neurons in the hidden layer | Activation function hidden layer | Activation function output layer | Training algorithm |
|---|---|---|---|---|
| MLP$_{e\text{-}l}$ 7-2-1 | 2 | Exponential | Linear | BFGS |
| MLP$_{e\text{-}ht}$ 7-2-1 | 2 | Exponential | Hyperbolic tangent | BFGS |
| MLP$_{e\text{-}l}$ 7-3-1 | 3 | Exponential | Linear | BFGS |
| MLP$_{s\text{-}l}$ 7-5-1 | 5 | Sigmoid | Linear | BFGS |
| MLP$_{ht\text{-}e}$ 7-5-1 | 5 | Hyperbolic tangent | Exponential | BFGS |

*Figure 7: Details of the Selected Anns for Further Training. Extracted From (Juszczyk et al., 2018)*

Figures 7, 8, 9 and 10 below presents four types of transfer functions commonly used: Unit step (threshold), sigmoid, piecewise linear, and Gaussian.

***The Unit Step (Threshold) Activation Function.*** Also referred to as the Heaviside step function, is a commonly employed activation function in neural networks. It categorizes input values into binary outputs by comparing them to a designated threshold. For inputs below the threshold, the function outputs 0, while for inputs equal to or above the threshold, it outputs 1 (Haykin, 2009).

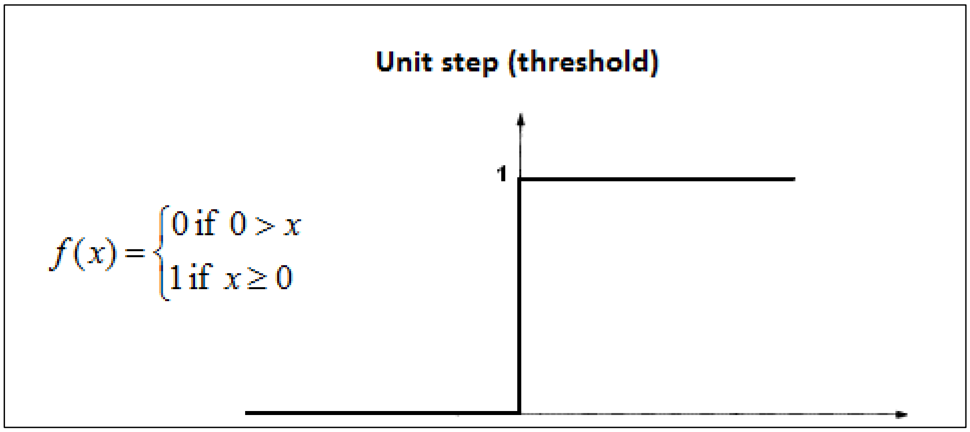



*Figure 7: The Diagram of Artificial Neural Network Function: Threshold. Extracted From (Artificial Neural Network, n.d.)*

***Sigmoid Activation Function.*** Also referred to as the logistic function, is widely used in neural networks. It transforms input values smoothly into a range from 0 to 1, enabling non-linear operations. It is commonly employed to introduce non-linearity in the network and is well-suited for binary classification tasks. Mathematically, it is defined as f(x) = 1 / (1 + exp(-x)), where exp represents the exponential function (Nguyen & Widrow, 1990).

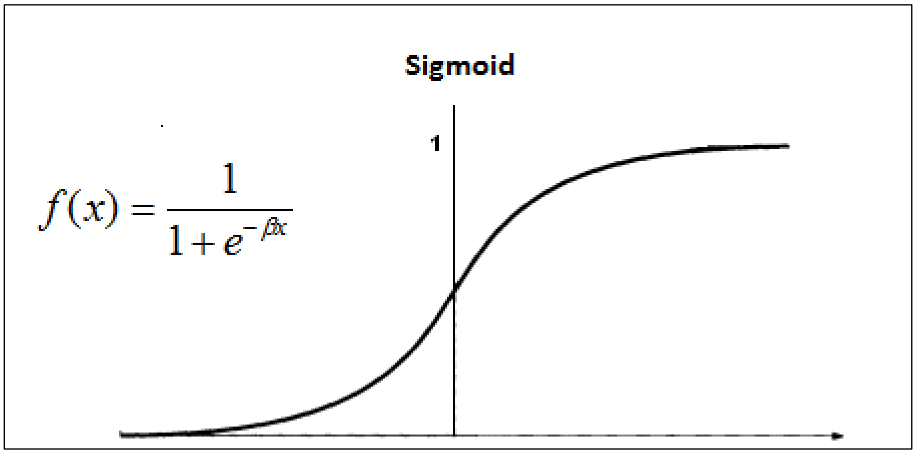



*Figure 8: The Diagram of Artificial Neural Network Function: Type Sigmoid. Extracted From (Artificial Neural Network, n.d.)*

***Piecewise Linear Functions.*** Bishop (2011) defines it as functions composed of multiple linear segments that are defined over distinct intervals or regions. These segments are connected at breakpoints, and the behavior of the function within each segment is determined by a linear equation. (Bishop, 2011)

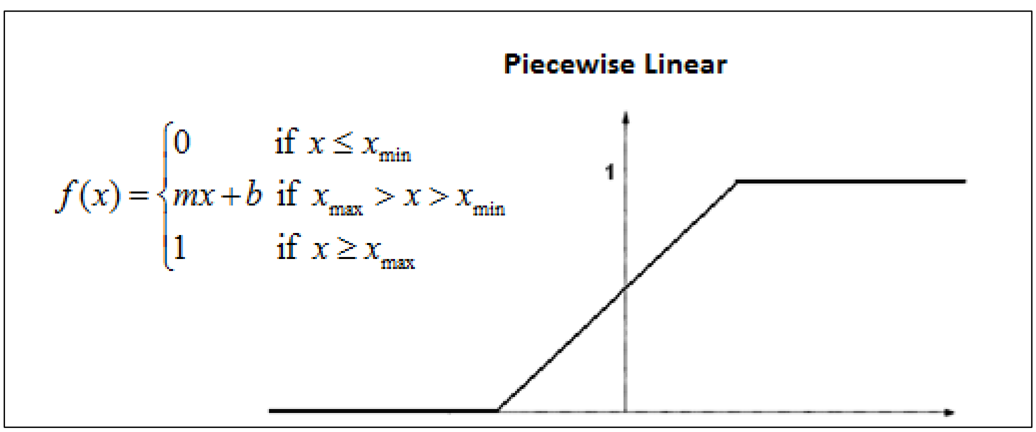



*Figure 9: The Diagram of Artificial Neural Network Function: Piecewise Linear. Extracted From (Artificial Neural Network, n.d.)*

***Gaussian Functions*****.** As defined by Bishop (2011), are mathematical functions that follow a bell-shaped curve with a symmetric distribution around the mean. The shape of the curve is determined by the standard deviation. These functions are commonly used in various fields, such as neural networks, for modeling continuous and smooth distributions of data. The mathematical formula for a Gaussian function is $f(x) = \exp(-(x - \mu)^2 / (2\sigma^2))$, where exp denotes the exponential function, x represents the input value, $\mu$ is the mean, and $\sigma$ is the standard deviation (Bishop, 2011).

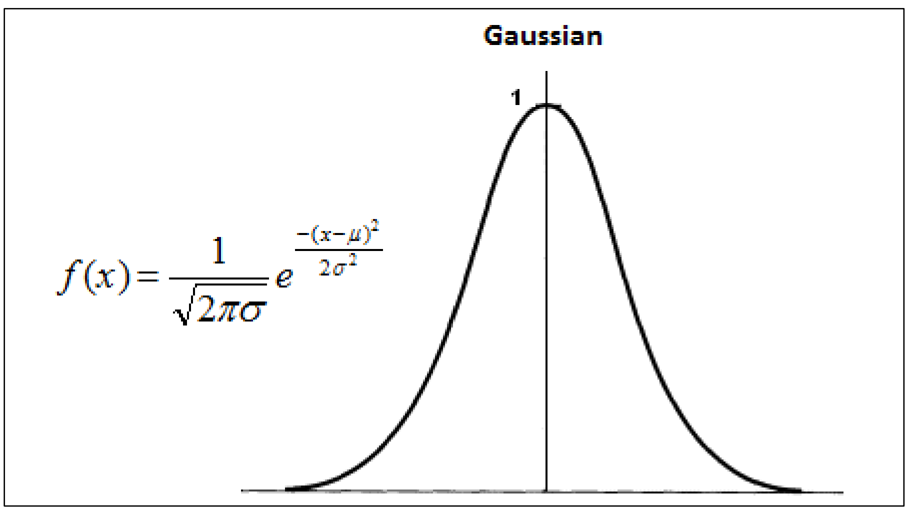



*Figure 10: The Diagram of Artificial Neural Network Function: Gaussian. Extracted From (Artificial Neural Network, n.d.)*

**Overfitting.** Occurs when a neural network achieves high performance on the training dataset but performs poorly on new, unseen data (Bishop, 2011).

**Generalization.** Refers to the network's ability to perform well on new, unseen data beyond the training dataset (Goodfellow et al., 2016).

**Regularization.** Encompasses methods used to prevent overfitting by imposing constraints on the weights or modifying the loss function (Chandanshive & Kambekar, 2019).

**Dropout.** It is a specific regularization technique that randomly deactivates some neurons during training to mitigate overfitting.

**Normalization.** It is the procedure of adjusting data to a standardized range in order to enhance the performance and convergence of an ANN. This entails rescaling the data so that it lies within a specific range, commonly 0 to 1 or -1 to 1. By normalizing the data, the dominance of features with larger scales is mitigated, promoting more balanced learning in the network. This ensures that the data is consistently formatted, enabling efficient training of the ANN (Geron, 2019).

**Denormalization.** normalization and denormalization are data preprocessing and postprocessing techniques in ANN that aim to improve the accuracy and interpretability of the network's outputs. Denormalization is the opposite of normalization, where the normalized data is transformed back to its original scale or range. It is commonly used on the output of an ANN to obtain meaningful predictions or results in the original data scale (Rumelhart et al., 1986).

### *Types of Artificial Neural Networks*

Haykin (2009) classified architecture of ANN into three distinct categories: Single Layer Feedforward Networks, Multilayer Feedforward Networks and Recurrent Networks. Elmousalami (2020) notes that the selection of the network architecture depends on the nature and requirements of the problem concerned.

**Single Layer Feedforward Networks.** Also known as a single-layer perceptron, is considered the simplest form and has only one layer of connection weights. The input units are connected to the output units, but there are no connections between input units or between output units as represented in figure 11. (Chakraverty et al., 2019; Haykin, 2009). They are limited in their ability to learn complex patterns and cannot capture non-linear relationships in the data.

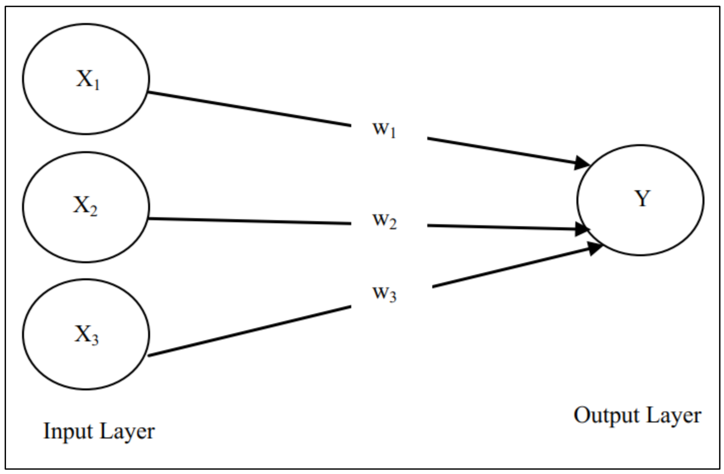



*Figure 11: Single Layer Feed Forward Network. Extracted From (Chakraverty et al., 2019)*

**Multilayer Feedforward Networks.** They consist of one or more hidden units/layers (intermediate layers) positioned between the input and output units. For every added hidden layer, the network capacity is increased in extracting higher-order statistics from its input (Haykin, 2009). They are employed for tackling complex problems that cannot be effectively solved by single layer networks due to their limited capabilities in training and performance (Chakraverty et al., 2019). There are various types of feedforward neural network architectures available, each tailored to specific problems including multilayer perceptron networks (MLP), radial basis function networks, generalized regression neural networks, probabilistic neural networks, belief networks, Hamming networks, and stochastic networks (Elmousalami, 2020). Figure12 shows the feed forward network.

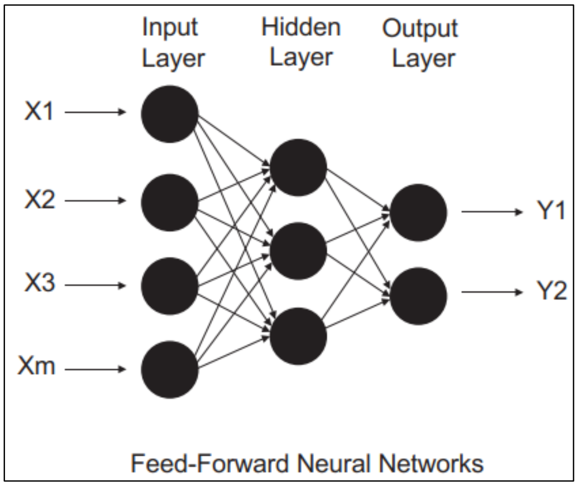



*Figure 12: Single layer feed forward network. Extracted from (Najafi & Tiong, 2015)*

**Multilayer Perceptron (MLP).** The MLP network requires predefining parameters such as input signals, hidden layers, and output neurons. There are no fixed rules for selecting the network structure, as it depends on the problem and is determined through computational tests. Training an MLP involves supervised learning with the error backpropagation algorithm. This algorithm consists of two phases: forward propagation and backward propagation. In forward propagation, input signals are processed through layers using synaptic weights and activation functions. The output layer is evaluated for errors compared to desired signals. Backpropagation updates the weights from the output layer to the input layer based on the obtained errors. This weight update process is known as backpropagation (Alcineide et al., 2021). Figure13 shows the feed forward network.

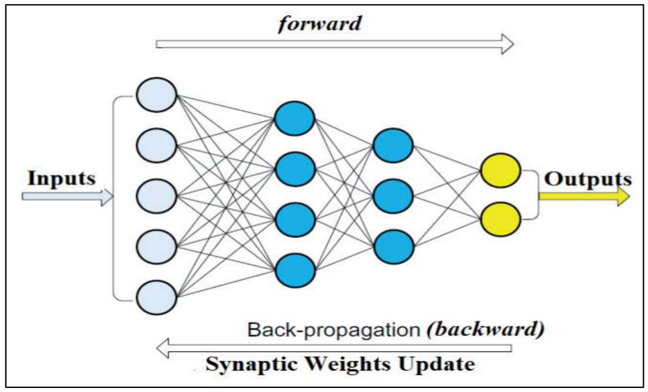



*Figure 13: Multilayer Perceptron (MLP) Showing Backpropagation Algorithm. Extracted From (Alcineide et al., 2021)*

**Recurrent Networks.** A key distinction between a recurrent neural network and a feedforward neural network is the presence of at least one feedback loop in the former (Haykin, 2009). By incorporating feedback connections, the preceding layers are able to receive data flow from the following layers (Najafi & Tiong, 2015). It is a specific type of neural network that incorporates feedback connections, enabling the flow of information from preceding layers to subsequent layers. Its purpose is to effectively handle sequential and time-dependent data (Goodfellow et al., 2016). He articulates that it is used to model relationships between sequences and other sequences rather than just fixed inputs.

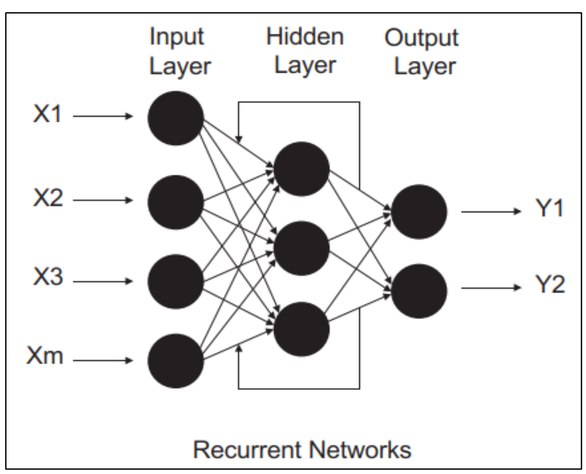



*Figure 14: Recurrent Networks Diagram. Extracted From (Najafi & Tiong, 2015)*

### Training of Neural Network

The training process is a crucial and important step in the development of ANN and the accuracy of its results. It is the process of attaining the desired output by adjusting the weights in the connections between layers of a network through multiple iterations. This allows the network to learn and improve its performance to accurately process information and produce the required results (Chakraverty et al., 2019). Shehatto (2013) articulates that the training process acquires weights meaningful information, whereas prior to training, they are initialized randomly and do not possess any specific meaning.

During the training process, the weights in the neural network are initialized with small random values. For each input, the network produces an output, which initially is random. The difference between this output and the desired output, known as the correct value, is computed using one of the performance measures. The training is repeated multiple iterations setting different weight configurations. The total error of the network is obtained by summing up the squared differences over all training examples. A perfect network would have an error of zero, and a smaller error indicates a better-performing network (Krogh, 2008). Training process can be described as the following steps (Jinisha & Jothi, 2019).

1. Initialize the network: Set up the network architecture, including the number of layers, neurons, and activation functions.
2. Define the loss function: Choose an appropriate loss function that quantifies the difference between the predicted outputs and the true outputs.
3. Initialize the weights: Assign initial random values to the weights and biases of the network.

4. Forward propagation: Pass the input data through the network to compute the predicted outputs.
5. Compute the loss: Compare the predicted outputs with the true outputs using the defined loss function to calculate the error.
6. Backpropagation: Propagate the error backward through the network, adjusting the weights and biases using optimization algorithms like gradient descent.
7. Update the weights: Use the computed gradients to update the weights and biases, aiming to minimize the loss.
8. Iterate the forward propagation, loss calculation, backpropagation, and weight updates for a certain number of epochs or until convergence.
9. Evaluate the model: Use a separate validation or test dataset to assess the performance of the trained model.
10. (Optionally) Fine-tuning and optimization: further refine the model by adjusting hyperparameters, exploring different architectures, or employing regularization techniques to improve performance and prevent overfitting.

There are various techniques used to train ANNs such as Backpropagation, Stochastic Gradient Descent (SGD), Adam Optimizer and Convolutional Neural Networks (CNN). The thesis will elucidate the first two techniques. Shehatto (2013) describes backpropagation algorithm as one of the most powerful and commonly used algorithms for training. According to (Jinisha & Jothi, 2019), Back-propagation learning algorithm was invented in 1969 for learning in multilayer network.

Tijanić et al. (2020) outlines that backpropagation, the squared error between the output values and the desired values is minimized using gradient descent. The error signals obtained from this process are utilized to compute the weight updates (Tijanić et al., 2020). The Backpropagation algorithm iteratively reduces the error between the model's output and the target output. By minimizing a parameter function, typically measured as the mean squared error over a training set, it learns the mapping from input to output (Günaydin & Doğan, 2004). Backpropagation is based on the error correction learning rule and consists of two phases: forward propagation and backpropagation (Alcineide et al., 2021). An epoch refers to a single pass through the entire training dataset during the learning process. It is a measure of the number of times the network has seen and processed all the training examples(Haykin, 2009). The network is exposed to a series of sample patterns in a repetitive manner until the error value is minimized (Vikas et al., 2011). The training process should be halted when the mean square error shows no further improvement over a specified number of epochs to prevent overtraining (Günaydin & Doğan, 2004).

In ANN, the Mean Squared Error (MSE) and Mean Absolute Error (MAE) are commonly used metrics for assessing the error. These metrics differ in their sensitivity to outliers and the way they quantify the error (Shehatto, 2013).

Both MSE and MAE provide a measure of the average error, but the choice between them depends on the specific application and the desired characteristics of the error measurement. The MSE is more sensitive to outliers and large errors, while the MAE is more robust to outliers and provides a more balanced representation of the error (Chollet, 2018).

Stochastic Gradient Descent (SGD): This method is a variant of gradient descent, but instead of performing computations on the whole dataset, which can be computationally expensive. SGD randomly selects a subset for each iteration.

### *Cross-Validation of Neural Network*

Cross-validation can be defined as a technique used in machine learning to evaluate the performance and generalization ability of a predictive model. Haykin (2009) outlines the data available is randomly divides into two major parts: training set and test set. The training set contains two subsets: estimation subset used to select the network model and a validation set (cross-validation) used to test and validate the model (Haykin, 2009).

Elmousalami (2020) articulates the validation set can range from 10 to 30 percent of the total sample to evaluate the model performance. The splitting of training dataset enables the implementation of K-fold cross-validation. The K-fold cross-validation data is utilized to determine the optimal hyperparameters for the model (Elmousalami, 2020). In the K-fold cross-validation, the dataset is divided into k equal-sized folds. The model is trained many times, with each fold serving as the validation set once while the remaining (k-1) folds are used as the training set. The performance of the model is then evaluated by averaging the results obtained from the k iterations (Goodfellow et al., 2016; Kohavi, 1995). Another variant of cross-validation is stratified k-fold cross-validation, which preserves the class distribution in each fold to ensure representative training and evaluation subsets (Kohavi, 1995). There are different variables of data validation such as Leave-one-out cross-validation (Kohavi, 1995) and Predictive sample reuse (Geisser, 1975).

Cross-validation serves the purpose of evaluating the model on a separate dataset from the one used for parameter estimation. It allows for assessing the performance of different candidate models using the training sample and selecting the best-performing model.

However, there is a risk of overfitting the validation subset with the chosen model's parameter values. To mitigate this risk, the selected model's generalization performance is measured on a distinct test set that is separate from the validation subset (Haykin, 2009).

Being mindful of potential biases and pitfalls in cross-validation is crucial. Varma & Simon (2006) highlight the biases that can occur when cross-validation is used inappropriately, such as data leakage and overfitting. To mitigate these biases, they suggest important precautions, including proper randomization of data prior to applying cross-validation and exercising caution when comparing models with differing complexities (Varma & Simon, 2006). Najafi & Tiong (2015) outline that the utilization of cross-validation ensures that the neural network is not overtrained, meaning it performs well on the training set but poorly on test data. By employing this technique, a stop criterion is selected to halt training as soon as the error on the cross-validation set begins to rise. Since there are no set rules for configuring neural networks, the practitioner should construct multiple networks and select the one with the lowest error (Najafi & Tiong, 2015).

***Testing of Neural Network***

During the testing phase, the accuracy of the model's predictions is assessed. The predicted results are compared to the actual results, and the percentage error in cost estimation is computed (Günaydin & Doğan, 2004). The author continues to describe that the impact of each network input on the network output can be examined through analysis, providing feedback on the most significant input parameters. This can be achieved through sensitivity analysis, a method that reveals the cause-and-effect relationship between inputs and outputs. By conducting sensitivity analysis, the network's size can be reduced, resulting in a simpler model and shorter training times. During sensitivity analysis, network learning is disabled, ensuring that the network weights remain unaffected. The approach involves modifying the inputs to the network and observing the resulting percentage change in the output (Günaydin & Doğan, 2004). Evaluation metrics such as accuracy, precision, recall, F1 score, or MSE are commonly used to quantify the performance of the network (Alpaydin & Bach, 2014).

The allocation of data for training and testing is dependent on data availability. The percentage of cases used for testing compared to training differs between the literature. Alcimede et al. (2021) state the percentage of training to testing can range from 70/30 (70% for training and 30% for testing), and up to 90/10 (90% for training and 10% for testing). Shiha et al. (2020) defined 80/20 (80% for training and 20% for testing). According to Shehatto (2013), a smaller subset of one-tenth of the training data can be allocated for cross-validation purposes. Jinisha & Jothi (2019) outlined in their study that they used a ratio of 70% for training, 15% for cross-validation and 15% for testing.

### *Performance Measures of ANN Model*

There are several performance measures used to evaluate the effectiveness of an Artificial Neural Network (ANN) model. Some commonly used performance measures in the context of ANN models include:

1. Mean Squared Error (MSE): MSE calculates the average of the squared differences between the predicted and actual values. The MSE is calculated by taking the average of the squared differences between the predicted and actual values. It gives

more weight to larger errors due to the squaring operation. The MSE can be expressed as: MSE = [(1/n) * Σ(y - ŷ)^2] (Chollet, 2018).

$$MSE = \frac{\sum_i^n (Y_{NN} - Y_A)^2}{n}$$

where n is the number of data points, y is the actual value, and ŷ is the predicted value. It provides a measure of the overall accuracy of the model's predictions.

2. Mean Absolute Error (MAE): it is calculated by taking the average of the absolute differences between the predicted and actual values. It treats all errors equally without giving more weight to larger errors. It can be expressed as:
MAE = [(1/n) * Σ|y - ŷ|] (Chollet, 2018). It provides a measure of the average magnitude of errors in the model's predictions.

$$MAE = \frac{\sum_i^n |Y_{NN} - Y_A|}{n}$$

3. Mean Absolute Percentage Error (MAPE): MAPE is calculated by taking the absolute difference between the actual and predicted values, dividing it by the actual value, and then multiplying by 100 to express it as a percentage. The absolute differences are then averaged across all data points to obtain the mean absolute percentage error (Hyndman & Koehler, 2006; Matel et al., 2019). Tijanić et al. (2020) defined MAPE by the formula below.

$$\text{MAPE} = \frac{1}{N}\sum \left| \frac{\text{Actual value} - \text{Estimated value}}{\text{Actual value}} \right|$$

4. Root Mean Squared Error (RMSE): it is a metric used to quantify the average magnitude of prediction errors in the same units as the target variable. It is obtained

by taking the square root of the MSE. RMSE provides a valuable assessment of how accurately a prediction model performs by measuring the average size of the errors (Hagan et al., 2014). Wang et al. (2022) defines the formula as shown below:

$$RMSE = \sqrt{\frac{1}{m} \times \sum_{i=1}^{m} (\hat{y}_i - y_i)^2}$$

5. R-squared (R^2): The coefficient of determination, is a statistical metric that quantifies the proportion of the total variance in the dependent variable explained by the independent variables in a regression model. R-squared ranges between 0 and 1, where a value of 1 indicates a perfect fit of the model to the data, and a value of 0 signifies no explanatory power. It is commonly used to evaluate the goodness of fit and predictive ability of regression models (Chandanshive & Kambekar, 2019). Wang et al. (2022) defines the formula as shown below:

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{m} (y_i - \bar{y})^2}$$

6. Precision, Recall, and F1-score: These measures are widely used evaluation metrics in binary or multiclass classification tasks to assess the performance of ANNs. Precision focuses on the model's ability to accurately identify positive instances, while recall measures the model's ability to correctly identify all positive instances. The F1-score combines both precision and recall into a single metric, providing a balanced measure of the model's performance. These metrics provide valuable insights into the model's classification accuracy and its ability to handle different classes effectively. (Alpaydin & Bach, 2014; Bishop, 1995) .

It is generally agreed that the choice of performance measure depends on the specific task and the nature of the dataset being analyzed. Different performance measures provide insights into different aspects of the model's performance.

### *Model Sensitivity Analysis*

Sensitivity analysis involves determining a plausible range of values for various elements of project data. The final result is then evaluated for each variation in the data, resulting in a series of "what-if" estimates (Keith & Nii, 2013). Sensitivity analysis involves computing the derivatives of the network's response with respect to each element of the input vector. It aims to identify the importance of individual input elements by examining the magnitude of these derivatives. Elements with small derivatives can be considered less influential and potentially eliminated from the input vector (Hagan et al., 2014). Keith & Nii (2013) articulates that sensitivity analyses are commonly visualized using spider diagrams, which highlight the most sensitive or critical areas that require management attention. However, a limitation of sensitivity analysis is that it treats risks individually and independently. Therefore, caution should be exercised when directly assessing the effects of combined risks using the data.

Haykin (2009) outlines that Sensitivity analysis can be effectively conducted using back-propagation learning, enabling users to explore and analyze the impact of changes in input variables on the output mapping in an efficient manner. He further highlights the sensitivity of an input-output mapping function F with respect to a parameter of the function, denoted by $\varpi$, is defined by the formula below.

$$\boxed{S_{\omega}^{F} = \frac{\partial F / F}{\partial \omega / \omega}}$$

Sensitivity analysis involves temporarily disabling the network learning process, ensuring that the network weights remain unchanged. The fundamental concept behind sensitivity analysis is to systematically modify the input values of the network and observe the resulting changes in the output. These changes are typically quantified as percentages, providing insights into the sensitivity of the network's output to variations in the input values (Günaydin & Doğan, 2004).

According to Hagan et al. (2014), once the training is completed for a multilayer network, it is valuable to evaluate the significance of each element within the input vector. Identifying low-impact (unimportant) elements allows for their elimination, simplifying the network, reducing computational requirements, and mitigating the risk of overfitting. Conducting a sensitivity analysis can be beneficial when there is no definitive method for determining the importance of each input.

**Artificial Neural Network Properties**

Heravi & Eslamdoost (2015) describes ANN in construction network through immense literature background. They outlined insights into the potential applications of neural network architectures in construction engineering and management, highlighting their effectiveness as a management tool for automation in the construction industry. They also highlighted utilization and applicability of artificial neural networks (ANNs) in civil engineering where they stated ANNs function and demonstrated their effectiveness in solving various civil engineering problems. One of the studies in his report developed a neural network model specifically for estimating productivity in excavation-hauling operations. Another study focused on estimating construction productivity for concrete formwork elements, such as slabs, walls, and columns. They explored different neural network structures and found that a three-layered network with a fuzzy output structure provided the most suitable model, considering the subjective nature of the input variables.(Heravi & Eslamdoost, 2015)

Vikas et al. (2011) characterizes the learning process of a neural network by three main events. First, the network is stimulated by environmental input. Second, the network's free parameters are adjusted in response to this stimulation, allowing it to adapt to external stimuli. Finally, the network exhibits a modified response to the environment as a result of the changes that have occurred in its internal structure and function.

Hashemi et al. (2020) highlighted the percentage of different methods applied to construction types. They outlined ANN being used in higher percentage in all types of construction projects compared to other methods of neural network such as fuzzy neural network, Support Vector Machine (SVM), Radial Basis Functions (RBF), and other methods. ANN varies from 35% in road and infrastructure projects, 40% in building projects and can reach up to 100% in water-related construction projects (Hashemi et al., 2020).

### *ANN Problems and Challenges*

There are several problems and challenges associated with using Artificial Neural Networks (ANNs) in various applications. Here are some common ones:

- Overfitting: ANN can be prone to overfitting; it is a significant issue that can arise during the training of a neural network. It occurs when the network becomes too specialized in the training data and fails to generalize well to new situations, leading to reduced performance (Chandanshive & Kambekar, 2019).
- Training time and computational complexity: Training ANNs can be time-consuming and computationally intensive, especially for large and complex networks with a large number of parameters. This can lead to longer training times and higher resource requirements (Chen & Lin, 2014).
- Selection of Appropriate Architecture: Designing an optimal architecture for an artificial neural network (ANN) involves various challenges, such as determining the appropriate number of layers, neurons, and connectivity. The optimal architecture may vary depending on the specific problem and dataset (Haykin, 2009).
- Data Limitations: The performance of a model is primarily determined by the quality and quantity of training data. The amount of data required for a machine-learning learning algorithm depends on the complexity of the problem and the chosen algorithm. Limited or biased data can lead to poor performance and biased predictions (Matel et al., 2019).
- Interpretability: ANNs are commonly regarded as black box models, which implies that comprehending and interpreting the rationale behind their predictions can be difficult. This lack of interpretability may impose constraints on their usability in certain fields (Guidotti et al., 2018; Hashemi et al., 2020).

- ANNs Hyperparameter Tuning: it can include parameters like learning rate, regularization, and network architecture, is a time-consuming task that demands meticulous experimentation to identify the optimal combination (Goodfellow et al., 2016).

**Application of ANN In Cost Estimation**

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

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

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

Günaydin & Doğan (2004) employed an ANN model to forecast the cost of the structural system per square meter during the initial phases of building design. They collected data from 30 residential building projects in Turkey. The ANN model they developed consisted of an input layer incorporating eight parameters that were accessible during the early design stage. The findings demonstrated that the trained ANN model accurately estimated the cost of the buildings, achieving a minimum accuracy of 93% in predicting the cost per square meter.

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

The results indicate that economic factors play a significant role in reducing estimation errors and may even be more influential than project characteristics. The findings have practical implications for stakeholders in the construction engineering and management field, providing insights for decision-making, and contribute to a better understanding of the impact of various influential factors on construction cost estimation.

The studies reviewed, including those mentioned previously, indicate that employing artificial neural networks for early cost estimation in construction projects has significant potential. These findings emphasize the importance of conducting additional research in this field to investigate and improve the utilization of neural networks for estimating construction costs at the initial project phases.

# CHAPTER 3

## Research Methodology

### Chapter Introduction

This chapter outlines the research methodology utilized in this thesis. It begins by discussing the research strategy and provides a visual representation of the research design through a flow chart. The research design involves an in-depth analysis of recent literature to identify key factors influencing cost estimation in building projects.

Subsequently, data collection is conducted through surveys to establish a correlation between these factors and the cost of projects. The collected data is then analyzed using methods presented and explained in this chapter. The end result can be used to enhance the input parameters for the ANN model that will be constructed in the next phase of the study. The model will be employed at the new projects' early stage to produce accurate estimates of costs.

### Research Strategy

The research strategy in this study consists of the following points:

1. Research Approach: The research approach is generally either quantitative, qualitative, or mixed methods. This research is based on mixed methods which allows for triangulation and deeper insights into research questions. Close-ended survey questions with predetermined choices are considered quantitative data that can be evaluated statistically. Afterwards, tools for development of Data collection are used such as interviews with expert reviews. They are considered a qualitative

research method as they entail having candid, in-depth discussions with individuals/organizations in order to assemble comprehensive, specific, and individualized data. In-depth exploration of organizational experiences, perceptions, and insights is made possible by the probing and follow-up questions that can be asked during interviews.

2. Research Design: The research design is well explained in the next paragraph where figure 15 depicts the design of every step of the research. It is based on exploratory and correlational data, where new factors are being determined while the previous literatures are used to correlate the data to close the knowledge gap in order to provide a better understanding of inputs for building exact cost estimation model.
3. Data Collection Methods: The first data collection was based on the analysis of literature reviews which were focused on least investigated factors in building ANN models for cost estimation. Afterwards, a quantitative survey is conducted with a sample of almost ninety organizations. Subsequently, tools for development of Data collection are used to improve survey results as described in the next point.
4. Sampling Strategy: The targeted population for the surveys were chosen based on their profiles and not distributed randomly to reduce misleading results. The population consisted of almost ninety organizations ranging from owners, consultants, and contractors in order to view results from all sides. The population itself was from various job positions but focusing primarily on managerial positions with a high experience to optimum results.
5. Data Analysis Techniques: Data analysis for quantitative surveys generates data that can be analyzed using statistical analysis, enabling numerical examination of

factors such as percentages, averages, and correlations. Interviews are regarded as a qualitative method whereby the outcomes are typically examined through thematic analysis or other qualitative analysis approaches to identify patterns, themes, and trends.

6. Limitations: the results can be considered limited as most of the population experience originated from UAE which makes results not accurate for other regions in the world. Moreover, the results are specific to construction field and in specific, building projects, making it limited to be used in road or infrastructure projects or any other field.

**Research Design**

The objective of this study is to identify crucial factors for early-stage cost estimation and subsequently develop an optimized Artificial Neural Network (ANN) model as an end result.

The subsequent phase of the study will concentrate on enhancing the accuracy of early-stage estimation in building projects through improvement of input factors that can be used in the implementation of the ANN model. Figure 15 depicts the designed procedure of every step of the research.

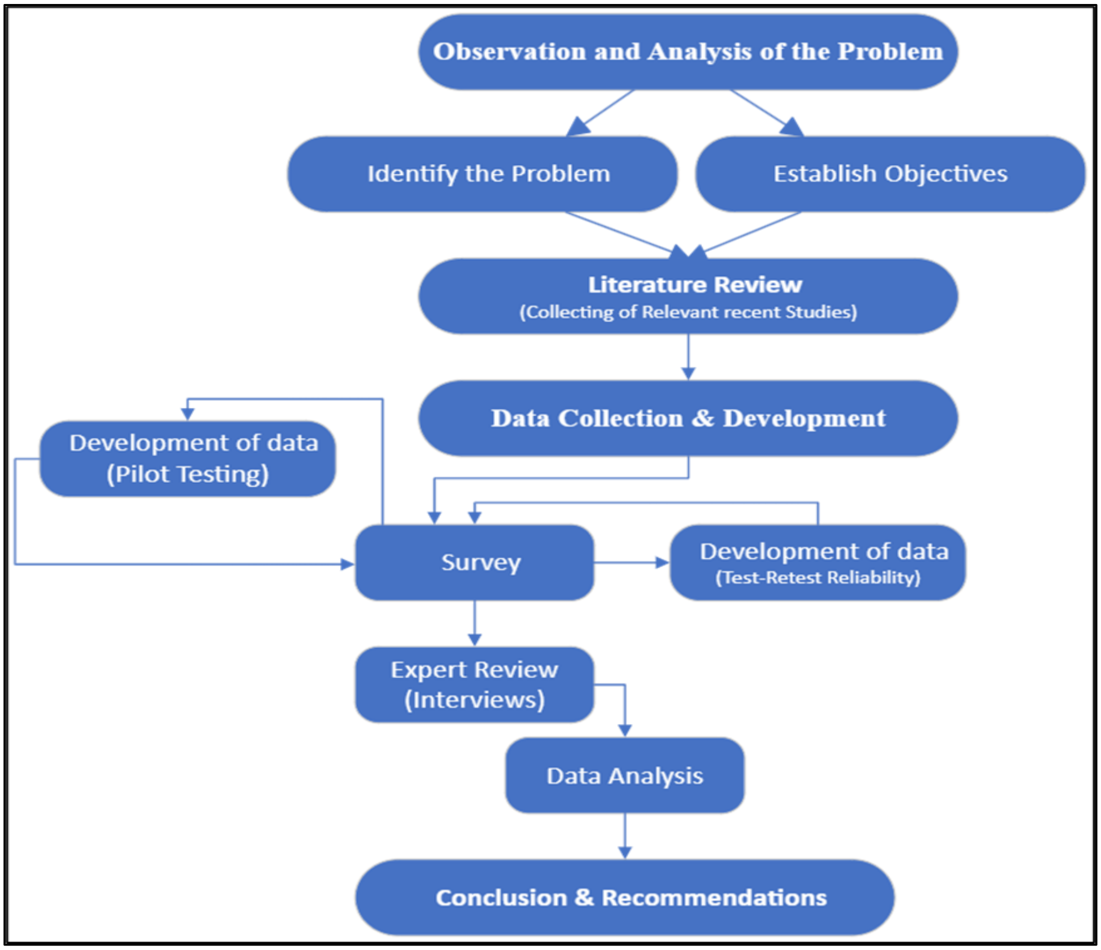



*Figure 15: Research Design.*

The methodology employed for this study, as illustrated in Figure 15, comprises the following stages:

1. During the observation and analysis of the problem, the topic is chosen by identifying the problems, objectives are set, and a research plan is formulated.
2. Literature Review: Relevant literature, including papers, reports, dissertation, and books, focusing on cost estimation in the construction industry, were thoroughly examined. Special attention was given to studies exploring computational artificial neural approaches such as Artificial Neural Network models for cost estimation.

3. Data Collection and Development of Data Collection Tools: A structured survey was distributed to the initial group of participants, while closely noting their comments on the survey. The feedback was integrated into the survey before being distributed to the general population. After conducting the survey, some participants were chosen to answer the survey again and their answers were noted down to check reliability of answers (Test-Retest Reliability).
4. Afterwards, the next step in data collection was to conduct expert review method (interviews), to gain closer view on the main parameters affecting cost estimation procedure of building projects in UAE. The participants chosen to perform interviews were senior participants in specialized fields such as: technical manager, estimation, contracts management, planning and general manager.
5. Data Analysis: After analyzing the data collected from surveys and interviews, the data was analyzed statistically at first and interview results were used to confirm the data and draw recommendations and conclusion from the study.
6. Conclusion and recommendation phase: In this stage, the content of the thesis was written, and the research chapters were covered. Moreover, the research was summarized in the conclusion section with many recommendations.

**Literature Studies**

Numerous researchers have examined different parameters and developed their models based on diverse factors. Most of the factors can be classified into three categories: Structural design factors, finishing factors and special circumstances factors such as market index, type of client, type of contract, project delivery method, etc....

The presented study focused on structural design factors and special factors as described in Table 1 below. The table provides a description of each factor along with the corresponding reference from which it was extracted.

*Table 1: Key Influential Factors Examined in Previous Research*

| Number | Factor Highlighted | Previous Literature Reference |
|---|---|---|
| 1 | Project Size | (Elfaki et al., 2014) |
| 2 | Type of Project | (Elfaki et al., 2014) |
| 3 | Soil type/ Conditions | • (Elfaki et al., 2014)<br>• (Chandanshive & Kambekar, 2019) |
| 4 | Type of Client | (Elfaki et al., 2014) |
| 5 | Type of Contract | (Elfaki et al., 2014) |
| 6 | Building Total Area | • (Günaydin & Doğan, 2004)<br>• (Chandanshive & Kambekar, 2019)<br>• (Juszczyk et al., 2018) |
| 7 | Number of Floors | • (Günaydin & Doğan, 2004)<br>• (Chandanshive & Kambekar, 2019) |
| 8 | Foundation Type | • (Günaydin & Doğan, 2004)<br>• (Chandanshive & Kambekar, 2019) |
| 9 | Structure/ Slab Type | • (Günaydin & Doğan, 2004)<br>• (Chandanshive & Kambekar, 2019) |
| 10 | Earthquake zone | (Chandanshive & Kambekar, 2019) |

| | | |
|---|---|---|
| 11 | Weather Conditions | (Najafi & Tiong, 2015) |
| 12 | Management Conditions | (Najafi & Tiong, 2015) |
| 13 | Skilled Labor Availability | (Najafi & Tiong, 2015) |
| 14 | Project Location | (Juszczyk et al., 2018) |
| 15 | Construction Trend value such as Fluctuation | (Wang et al., 2022) |

**Tools for Development of Data Collection**

Tools and methods used to develop surveys results typically include:

1. Pilot Testing: Administering the survey to a small sample group to identify any issues with the survey design, clarity of questions, and response options.
2. Test-Retest Reliability: Administering the survey to a group of participants at two different time points to assess the consistency of responses.
3. Expert Review (interviews): Seeking feedback from subject matter experts or experienced researchers to evaluate the survey's content validity, clarity, and appropriateness. Conducting interviews can be considered a data analysis method. Interviews allow for in-depth exploration of the participants' perspectives and can provide valuable insights and clarification regarding their survey responses. Interviews can help test survey findings by verifying and complementing the quantitative data collected through surveys. The information obtained from interviews can also contribute to a richer understanding of the research topic and enhance the validity of the study's findings.

# CHAPTER 4

## Data Collection and Results

### Chapter Introduction

The objective of this chapter presents study population characteristics, outlines determined factors correlation between literature and study results, depict statistical results of the answers for each of the studied factors, and concludes the results with a description on expert review of study result to assess answers for replies and provide top eight to ten factors that will be the input layer for future ANN model.

### Survey Analysis

The survey has been replied with 87 feedbacks, from a total of 95 surveys distributed, with a response rate of 91 percent. In the below section, the background of the responders, designation, years of experience and location of the experience will be depicted. There was one result that was filled by non-construction responder that were not considered in the final results.

#### *Characteristics of the Responders*

The characteristics of the responders encompass various attributes, including the distribution of employment by work sector, participants designation, and years of experience by market.

##### Distribution of Employment by Work Sector.

As per the results statistical analysis, 69 percent of responders were working in contracting companies, while the second majority were from engineering consultants with 23 percent and seven percent as owners or owner representatives. The following is described in pie chart in figure 16.

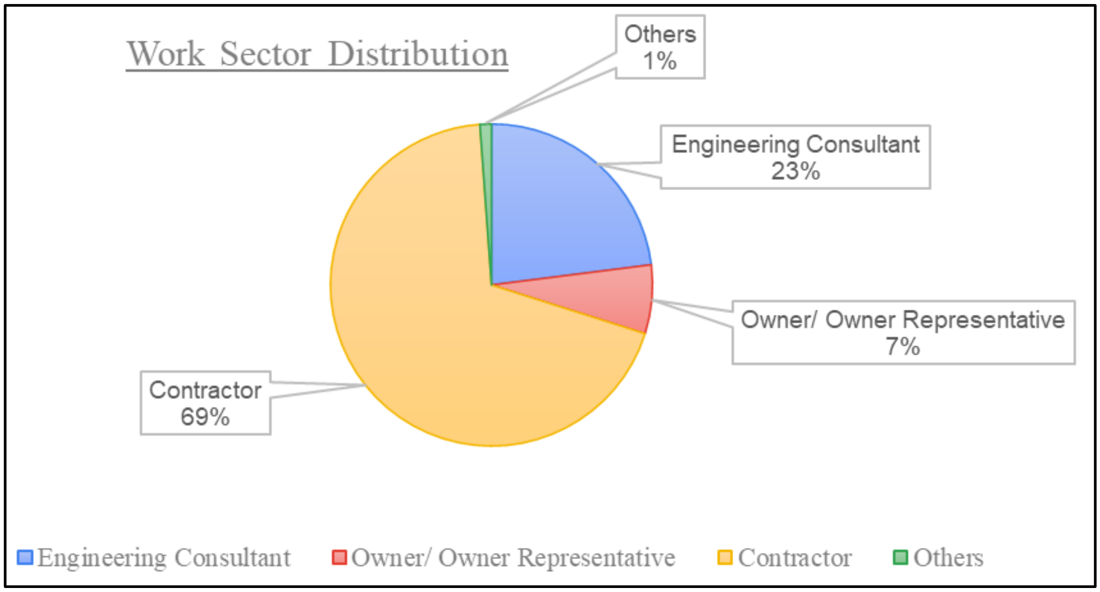



*Figure 16: Work Sector Distribution. (Total Sample of 87)*

**Participants Designation.**

The survey results provided a high return of qualified responders who can provide better feedback. The highest percentage or responders were higher management with over 15 years of experience in the field representing 16% of total participants. The following category were project managers who also can provide a bigger image on cost estimations and factors accuracy with a responder's percentage of 14%. Project engineers represented 11%, designers and consultants 10%, and technical department nine percent. The population also includes planning managers, quantity surveyors (QS), construction managers, owners, quality department, contracts managers, procurement managers, estimation department and HSE managers as depicted in figure 17.

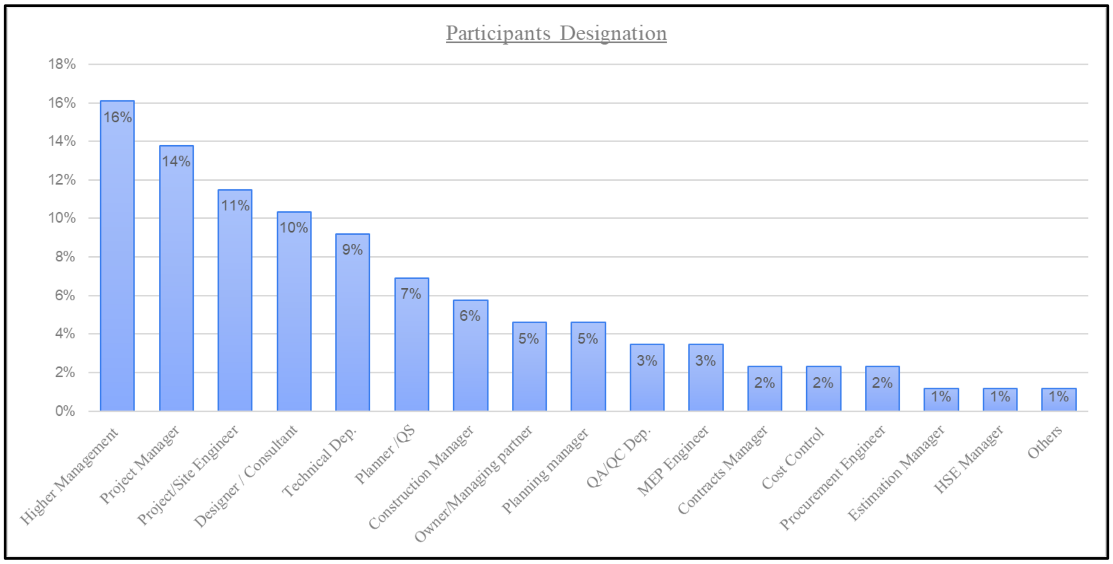
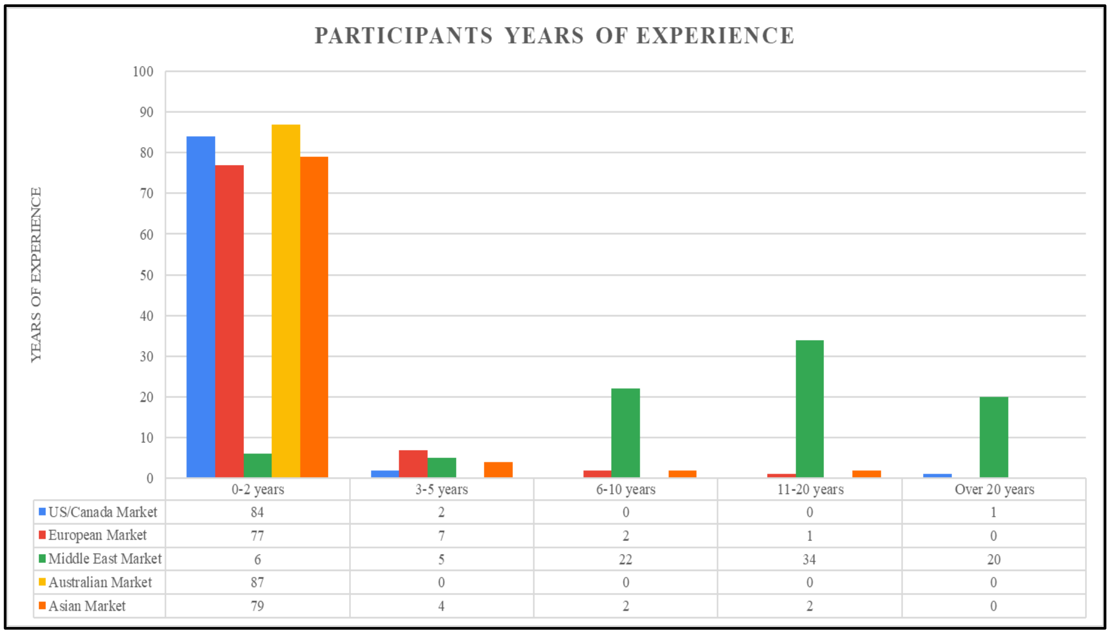



*Figure 17: Population Positions in the Construction Field. (Total Sample of 87)*

**Participants Years of Experience.**

The population years of experience varies from 2 years to over 20 years in the construction field. The study also provides a view on the years of experience by market. As shown in figure 18, most of feedback are from middle east market and this is because the study is focused on UAE market. The American construction market, European market, and Asian market have minor presence in responders experience with total absence of Australian market. The highest percentage of feedback was ranging from 11-20 years with 39% of population, followed by range 6-10 years (25%) and then range of over 20 years (23%).




| | 0-2 years | 3-5 years | 6-10 years | 11-20 years | Over 20 years |
|---|---|---|---|---|---|
| US/Canada Market | 84 | 2 | 0 | 0 | 1 |
| European Market | 77 | 7 | 2 | 1 | 0 |
| Middle East Market | 6 | 5 | 22 | 34 | 20 |
| Australian Market | 87 | 0 | 0 | 0 | 0 |
| Asian Market | 79 | 4 | 2 | 2 | 0 |

*Figure 18: Population Years of Experience in the Construction Field by Market. (Total Sample of 87)*

### Factors Affecting Cost of the Building Projects

Theoretical key factors have been determined from past literatures, other factors that have been identified through expert recommendations from past experiences and both were used to build the survey. Factors of cost estimation are the main component in an ANN model where they are the input layer of the model. Inaccurate estimation can be noted as one of reasons for project cost overruns. Figure 19 shows the percentage of survey replies rating if inaccurate estimation is a reason for cost overruns in projects, where 87% of replies confirmed that inaccurate estimation is a main reason for project cost overrun.

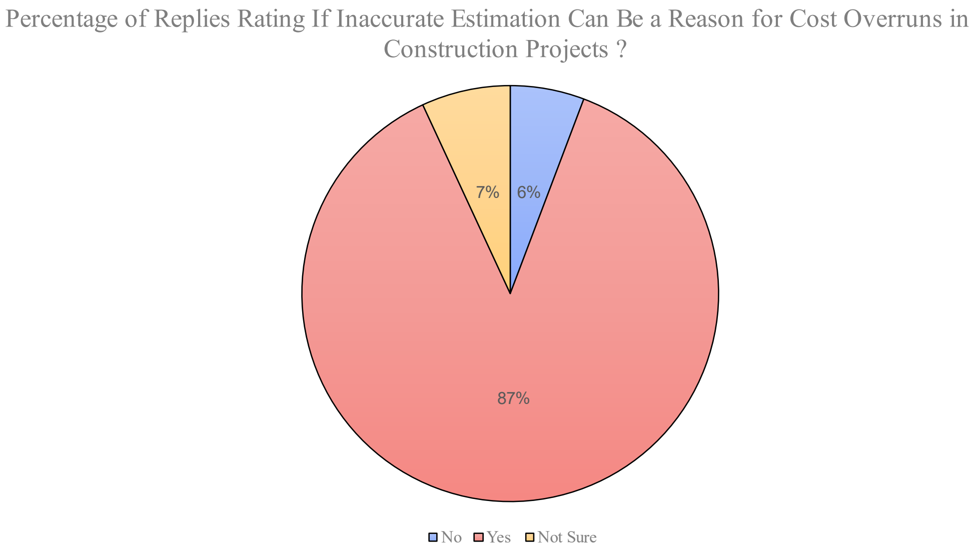



*Figure 19: Percentage Rating If Inaccurate Estimation Can Be Identified As Reason for Cost Overruns in Construction Projects (Total Sample of 87).*

Figure 20 illustrates total replies stating percentage of projects in the company with cost overrun due to inaccurate estimation. It shows that 21% of replies state that only 1/10 of projects face overruns, while 32% of the replies state that a quarter of projects face overruns. 34% of the population replied that half or more than half of projects face overruns due to the same issue.

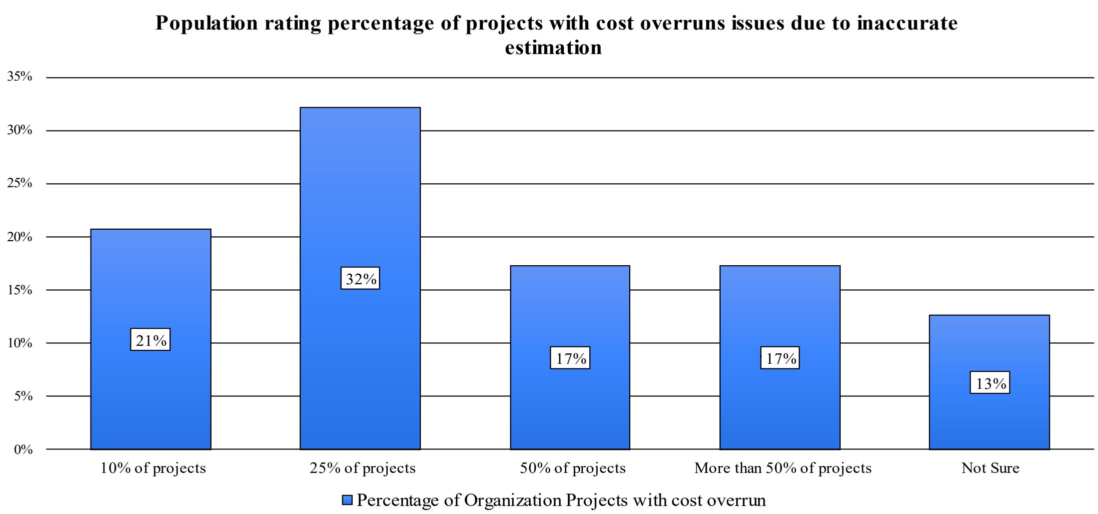



*Figure 20: Population Rating Percentage of Projects With Cost Overruns Issues Due to Inaccurate Estimation (Total Sample of 87).*

### *Data Collection Tools Development*

In this section, data collection tools developed to improve the survey will be described. After problem observation and literature review, the survey was conducted based on the latter output. After that, the first group of 19 people were pilot tested, as illustrated in research design in figure 15, where they were administered the survey to identify any issues with the survey design, clarity of questions, and response options. Their feedback was input, and the survey was enhanced before major distribution. After the final distribution, some of the early 19 population were asked to undertake survey again on efforts of (Test-Retest Reliability), where their responses were checked assess the consistency of their replies. The last step was conducting interviews (Expert Review) with management roles to obtain more understanding of the cost estimation process.

### *Data Results*

The survey shown in appendix A starts by collecting the percentage of advanced technologies in estimation department and in projects, followed by their experience on cost overrun correlation with the inaccurate estimation and percentage of projects facing overruns. Subsequently, 25 identified factors are introduced to be weighted on cost estimation based on their frequency and severity. 15 factors consist of structural design while the last 10 factors are various key parameters that are part of estimation process such as usage of BIM, trend value (fluctuation), safety requirement, type of client, etc.

A comprehensive description of each of the 25 factors influencing cost estimation is provided, highlighting the key considerations associated with each factor, aiming to enhance the reader's understanding of these factors and their significance in the estimation process. At the end, the factors are weighted, and the top 10 factors are presented in table 2 and compared to expert review to validate the study efforts. The rating of each factor in the survey consists of rating severity of impact of factor on cost estimation and the result can range from very low impact to very high impact on cost estimation.

**Type of Used Foundation in the Building.**

The type of foundation is a major package to be considered in project estimation in terms of risk, material quantity, labor required and special equipment to be procured. The package can be assigned to a subcontractor that will have a risk shifting and decrease the risk on the contractor, but at the same time, this will lead to a higher tender price and in a highly competitive market, it may impact the final benefit. As shown in figure 21, 39% of the population rated this factor as high, and 25% of replies were moderate.

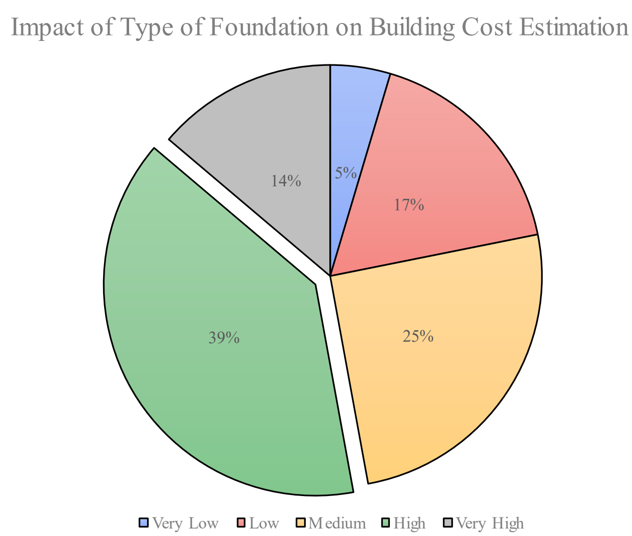



*Figure 21: Population Rating Impact of Type of Foundation Factor on Cost Estimation (Total Sample of 87).*

**Type of Structure Material (Steel, Concrete, Prestressed, Precast …).**

The type of structure material for the construction of the project has a major impact on project planning, type of equipment, skilled team and labors and risk. For example, concrete reinforced projects will have longer project span, but it is a proven science with low risks and low requirement for skilled professionals and special equipment. Steel structure have relatively shorter time frame but require skilled labor which will change the cost estimations. Precast projects are fast executing and have low risk and overruns because of their factory environment, but they have a higher cost and require higher logistics and special equipment. 70% of the replies rated type of structure as moderate to high, with 42% showing high impact as depicted in Pie chart in figure 22.

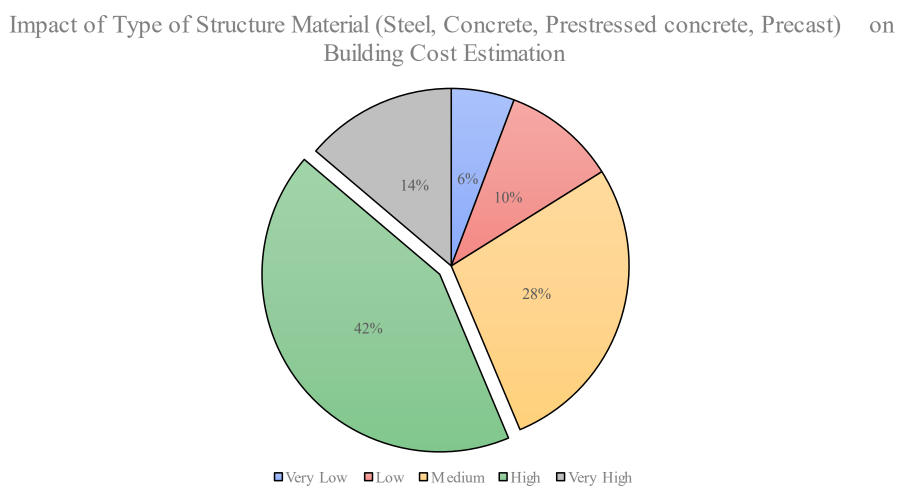



*Figure 22: Population Rating Impact of Type of Building Structure Material Factor on Cost Estimation (Total Sample of 87).*

**Area of Typical Floor.**

The area of the projects' ground impact on cost estimation is rated almost equally from low to high as it depends on the company scale. Small companies will consider its impact higher as they will require to procure material and equipment, recruit new people, and engage sometimes in jobs they have never executed before. Bigger companies have different type of consideration as bigger scale project decrease the overhead and indirect cost per square meter (better ratio of people to area of project leads to less cost of staff, as for example, a 1000-metersquare will require one engineer while a 100-meter square will also still require an engineer, but overhead will increase in the latter). As shown in figure 23, 30% of replies were high, followed by 27% as moderate and 25% as low impact; showing that its impact varies by other related conditions.

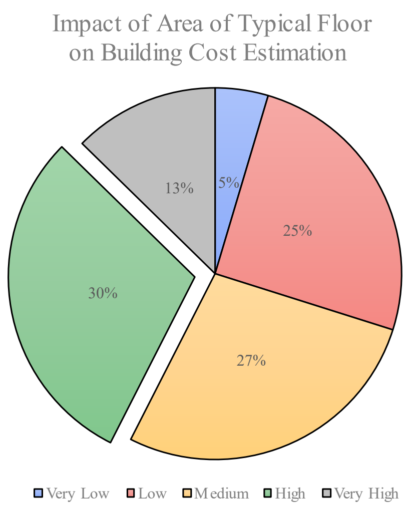



*Figure 23: Population Rating Impact of Floor Area Factor on Cost Estimation (Total Sample of 87).*

**Number of Floors.**

The number of floors has a different perspective if it can be considered a major factor or not. If the number of floors is small, then, the impact difference between Ground and five levels (G+5) or a G+10 is low, other than the quantities that are considered. When the project becomes a high-rise, the overhead and indirect cost will rise, and the factor becomes high rated because manlifts will be added, customized special cranes and concrete might be considered, plumbing and sewer systems needs to be created at certain levels for crew needs, risks are higher and will require more expensive insurances, higher safety requirements, crew training, etc.… Replies were distributed mainly from moderate to very high with highest rating was "High" with 31% of replies followed by "Very high" and "Moderate" at 23% as shown in figure 24.

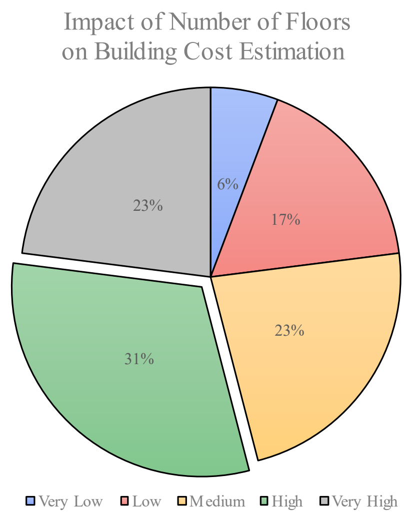



*Figure 24: Population Rating Impact of Number of Floors Factor on Cost Estimation (Total Sample of 87).*

**Type of Slab (Solid, Ribbed …).**

The type of slab is generally related to the factor stated previously "Type of Structure Material". Its impact will not be important comparing main structure material, hence, the replies rating the factor were distributed from low to moderate to high with 17%, 32% and 29% respectively as illustrated in figure 25. As stated previously, the impact will vary on size of organization and availability of material and equipment, usage of subcontractor and many other factors.

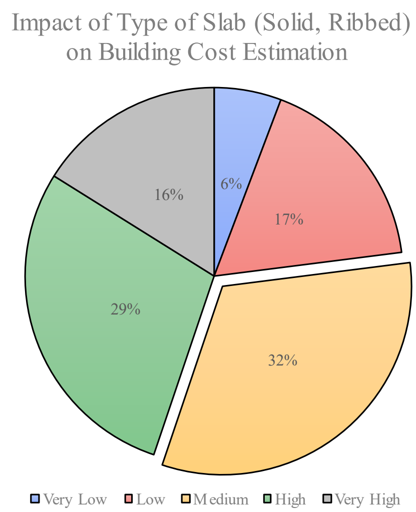



*Figure 25: Population Rating Impact of Type of Slab Factor on Cost Estimation (Total Sample of 87).*

### Number of Staircases in the Building.

The number of stairs can be considered a low impact factor as shown in figure 26. Third (33%) of the replies were considered low, followed by 26% moderate and only 24% considered the factor as high. The factor impact consists of material availability, risk of execution (risk of formwork opening) and time frame for execution.

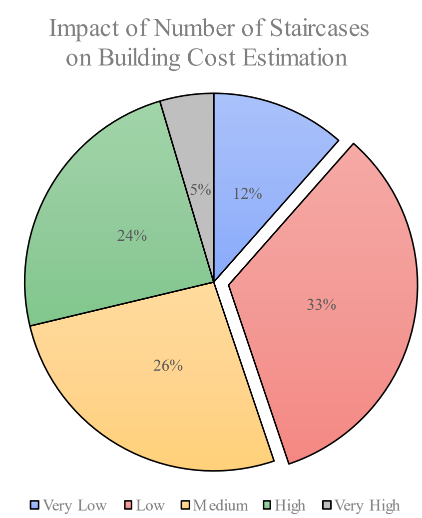



*Figure 26: Population Rating Impact of Number of Staircases Factor on Cost Estimation (Total Sample of 87).*

**Type of Contract (Lump Sum, Remeasured, Etc...).**

The type of contract is one of the most important factors in cost estimation of projects. One main point in contracts is whether the conditions are fair and balanced or fixed with no possibility for variations and changes. The Lumpsum contract will have a higher price for the same project classified under unit-price contract (called remeasured in U.A.E) because of the risk shifting from client to contractor. Another influence that has a direct relation with type of contract is the method of delivery of the project as it defines the risk for which party and this what basically defines the estimation. Figure 27 shows 37% replied factor as "High", then 29% as "Moderate" and 11% as "Very high".

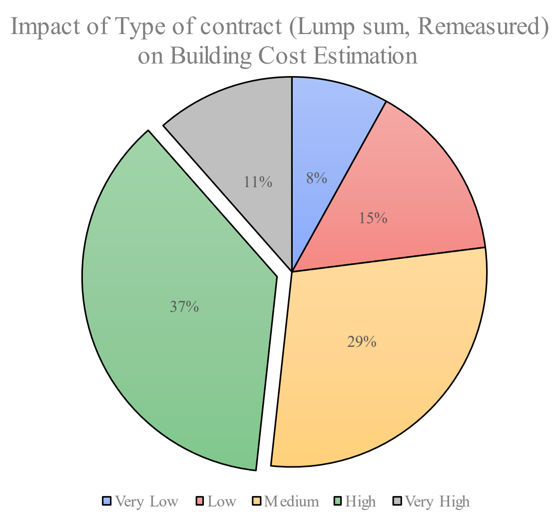



*Figure 27: Population Rating Impact of Type of Contract Factor on Cost Estimation (Total Sample of 87).*

**Length of Spans Between Columns.**

Distance of spans have a direct effect on operations, and because most of the population background was operation, it can be noticed that this factor has been defined ranging from low to high with highest rating (36%) as moderate effect as shown in figure 28 below. Its effect from expert review perspective is minor and can be defined as “Very low”. Third of the replies was “Moderate” with quarter of replies rating “Low” and another quarter rating “High”.

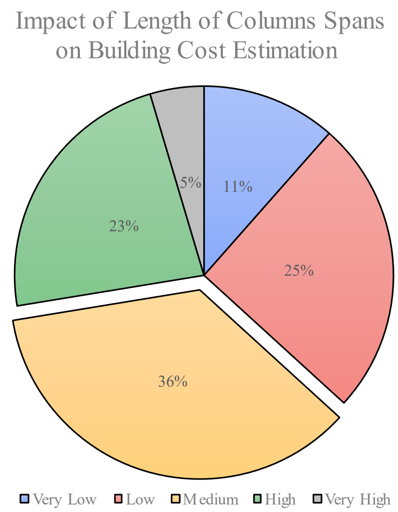



*Figure 28: Population Rating Impact of Type of Contract Factor on Cost Estimation (Total Sample of 87).*

**Area of Shear Walls in Project.**

Area of shear wall depend on three factors mainly: availability of formwork and labor, risk of formwork opening, and time frame for execution. Third of the survey (30%) population rated it moderate which gives it accurate description, and half of replies was divided between low and high as shown in figure 29. Experienced people classified as low, operation people and unexperienced opted for higher classification.

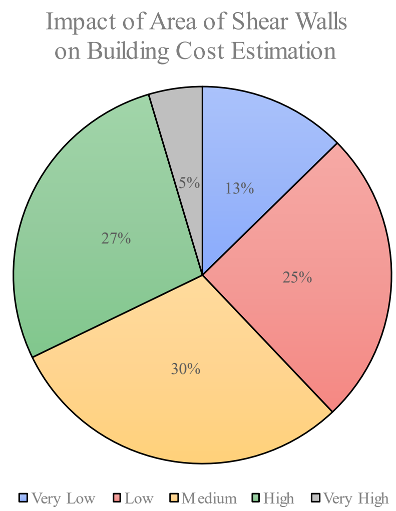



*Figure 29: Population Rating Impact of Area of Shear Walls Factor on Cost Estimation (Total Sample of 87).*

**Location of Project (Country, City).**

The location of the project can be classified in the top five in terms of impact on the cost estimation. It is linked to most of the other important factors and its impact will consist of the below points to be able to provide a decent cost estimate.

- Is the project location in an urban, suburb or in the city?
- Is it located in an isolated area or near material sources?
- Is the location previously recognized or new to the company?
- How's the political situation and customs and regulations if foreign country?

All the factors above will contribute to a noticeable rise in cost because of unknown risk and in procurement of insurances to mitigate risks. It can include overhead and indirect costs such as staff housing if isolated area, challenges in transportation, might require building of temporary roads and service networks, raw materials (gravel, cement, sand, and others) might need higher time to deliver or need to look other suppliers with higher prices, area labor might be scarcer, etc. Figure 30 shows

34% of replies were “Moderate” followed by 30% as “High” and another 30% divided between “Very high” and “Low”.

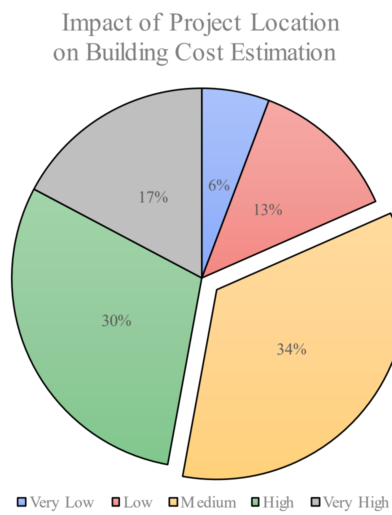



*Figure 30: Population Rating Impact of Location of Project Factor on Cost Estimation (Total Sample of 87).*

**Usage of Building.**

The usage of building factor can be classified into two main categories: governmental (public) or private project. Governmental projects will have a higher standard, regulations, and Leed requirements, opposed to private where the budget will determine the rest. Subcategories such as commercial, schools, residential or facility can further impact the estimation, but the main impact will always whether governmental or private. Figure 31 below shows that 32% of replies were “High”, followed by 29% “Moderate”, and 16% as “Low”.

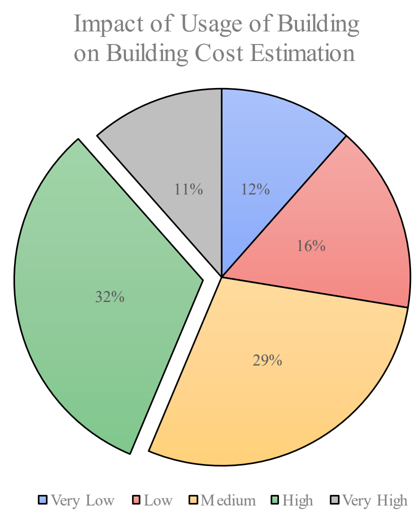



*Figure 31: Population Rating Impact of Project Building Usage Factor on Cost Estimation (Total Sample of 87).*

**Requirement for Dewatering System.**

The usage of dewatering is a moderate factor because its risks are known and mitigable. Figure 32 illustrates 33% of replies as “Moderate”, followed by 29% as “High” then 21% as “Low”.

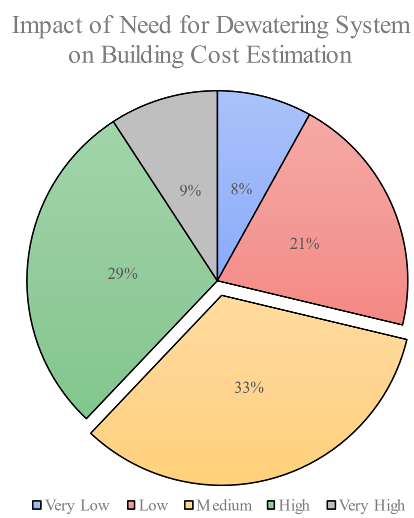



*Figure 32: Population Rating Impact of Need for Dewatering Factor on Cost Estimation (Total Sample of 87).*

### Earthquake Zone.

The replies on the survey were based on UAE factors, this explains the general distribution between weights, because some replies are based on local market while rest based on international market. Earthquake factors are rarely applied in the region because there are no seismic zones. In general, seismic zones project will have a higher specification, more rigid structure system, more complex building system, and detailed method statements more than that of normal projects. Also, it will require special insurance in case an earthquake hits while the project is in construction and will require more safety regulations. Figure 33 depicts 27% for each of moderate and high weight and 21% for each of very low and very high weight.

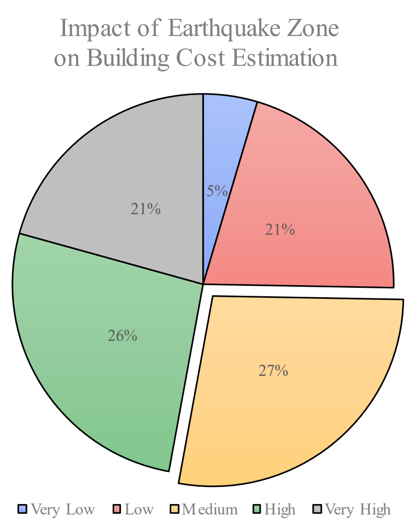



*Figure 33: Population Rating Impact of Earthquake Zone Factor on Cost Estimation (Total Sample of 87).*

### Soil Type.

Soil type is a major factor; it is the base of design, type of foundation, and many other factors. Soil type is a risky factor in the case of a design and build (D-

B), if the contract is lumpsum, if the contract conditions are imbalanced. One example is a construction in Abu Dhabi, UAE where the project where D-B and the soil report was not ready at the time of estimation; the project land was full of cavities and concrete volume for piles were underestimated. The contract was a fixed lump sum causing a loss of profit for the contractor. Figure 34 states that 32% of replies were “high”, followed by 32% “Moderate” and 22% as “High”.

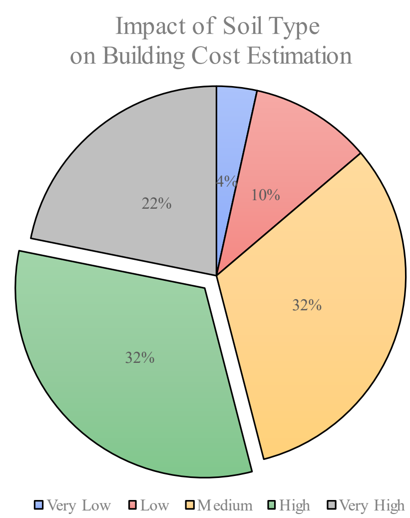



*Figure 34: Population Rating Impact of Type of Soil Factor on Cost Estimation (Total Sample of 87).*

**Complexity of the Project.**

Complexity of project was rated highest impact on cost estimation. As can be noticed in Figure 35, 40% rated as “Very high”, followed by 31% as “High”, then 20% as “Moderate”, building total percentage to 90% as moderate or higher. The main reason for this high percentage of inaccurate estimation analyzed from the study led from complexity of drawings and because of the low experience of some of the estimation team, quantities may be estimated wrong or even drawings may not be understood. Complex projects may also require new execution methods that are hard to predict cost, higher non-mitigable risks, etc. This is a conclusion of the study, but complexity as a factor is a wide subject where the factor can be better defined, and more conclusions can be withdrawn from deeper studies on the factor impact on the cost estimate.

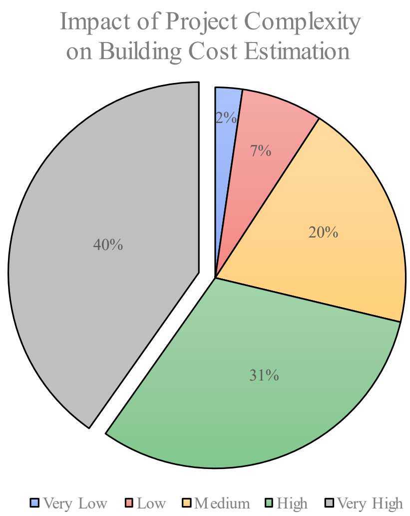



*Figure 35: Population Rating Impact of Project Complexity Factor on Cost Estimation (Total Sample of 87).*

**Usage of Building Information Modelling (BIM).**

The usage of advanced techniques such as Bim has improved the quality of cost estimation process by providing accurate automatic quantity survey for different material and enhance the understanding of complex projects. It also improved the coordination between different trades, enabling to detect clashes before happening which decreased abortive work. Figure 36 shows that more than 76% of replies stated that they use BIM in half or more in their project. this proves that it is widely used. Dubai projects, in the UAE, are now obliged to use BIM for the handover as a requirement from the client.

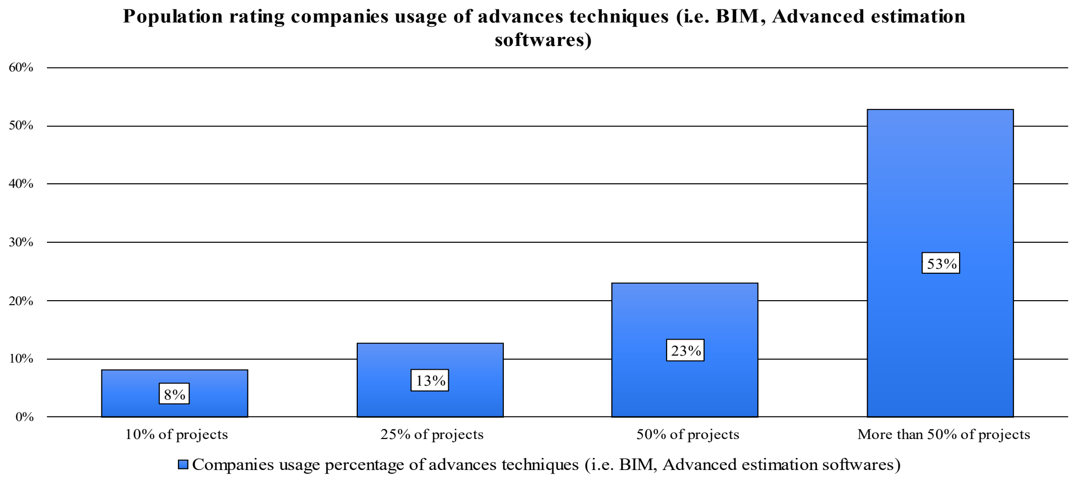



*Figure 36: Population Rating Their Organization Usage of Advances Techniques (I.E., BIM) (Total Sample of 87).*

Figure 37 highlights the importance of usage of BIM as 34% where replies gave a moderate rating on impact with 25% rating impact high and 20% rating very high. This means that 80% of replies agreed impact vary from moderate to high.

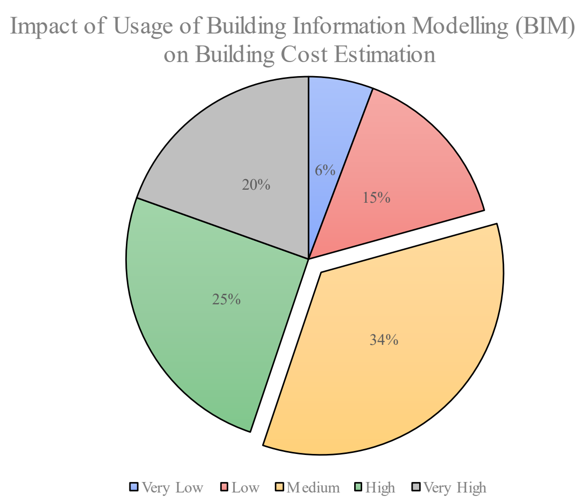



*Figure 37: Population Rating Usage of BIM Impact on Cost Estimation (Total Sample of 87).*

**Weather Conditions (Severe Fog, Sandstorm, Rainfall).**

The impact of weather conditions on cost estimation varies on the case presented. Some cases can be extreme and unknown, and others can be mitigated and known no matter how extreme the case is. For example, marine projects are extreme and unknown, risk can be mitigated by monte Carlo simulation of the tides and sea movement, but the possibility of sea change cannot be determined, and high safety measures should be in place. Another example where the case is extreme but known is working in harsh conditions such as Canada freeze zones where people can only 3-6 months a year which will impact the time frame of execution. Generally, weather will be accounted on a risk factor known to the organization, where in UAE, severe fog will impact some days where crew will not be able to arrive to work, heavy rainfall in London will impact concrete pouring schedule. Figure 38 highlights the factor impact mainly as moderate with a percentage of 34%, while 29% rated it high and 22% rated the impact low.

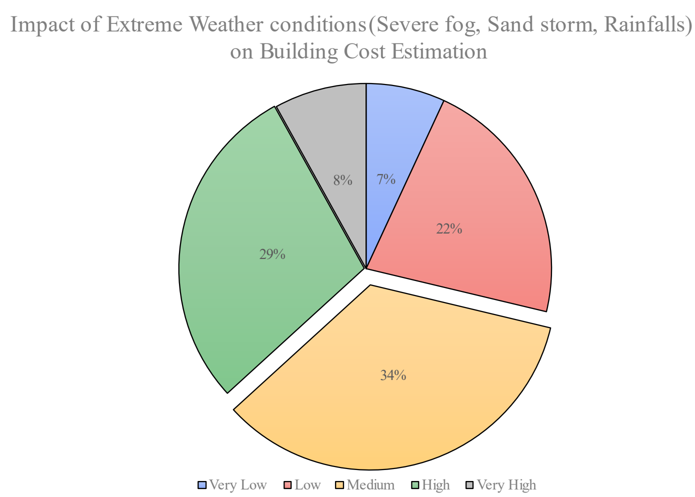



*Figure 38: Population Rating Impact of Weather on Cost Estimation (Total Sample of 87).*

**Management Conditions.**

A good Management is an important factor for project execution but not as important to cost estimation process. A well-established management will have better contingency plans, better project execution methods, less risks and abortive works and all of the mentioned will lead to less risk of project budget overrun.

However, estimators will consider the points above, however, will not put unwanted risk factors because of improper project management. Figure 39 shows that 41% of replies noted it as moderate, 24% as high and 15% as low.

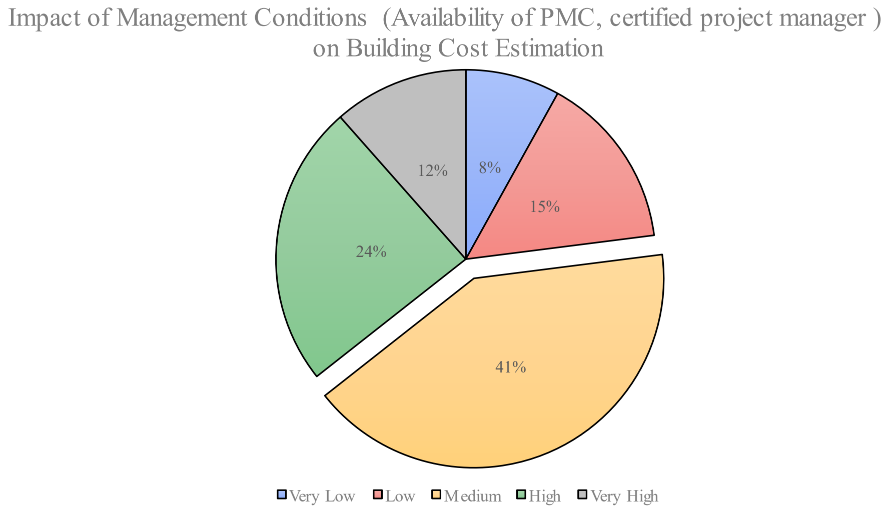



*Figure 39: Population Rating Impact of Management Factor on Cost Estimation (Total Sample of 87).*

**Availability of Skilled Labor.**

Availability of skilled labor is a critical factor in cost estimation. Locations with scarce labor may delay project execution and might even reach to stop the works. Recruitment in such zones will be more expensive and might require the organization to outsource labors from abroad, which in turn costs housing, meals, transportation, and higher salaries because of expatriation. Another issue related to availability of skilled labor that can be applied to international market but not in UAE, is that outsourcing labors may create issues with local labor unions and local authorities. In Figure 40, 36% of the replies rated the impact as high, followed by 29% rating moderate, then 24% rating very high impact and none rated it very low.

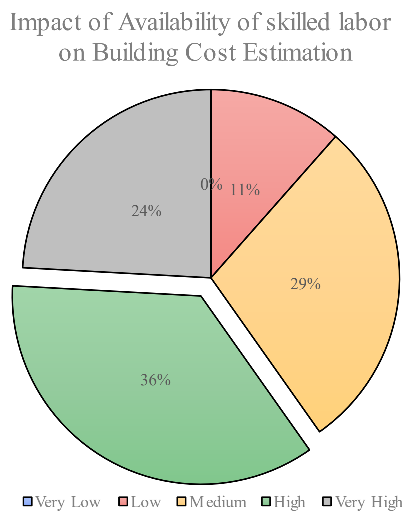



*Figure 40: Population Rating Impact of Availability of Labor Factor on Cost Estimation (Total Sample of 87).*

**Characteristics of Client/Shareholders.**

The type of client is an important factor in cost estimation. It defines how the client will interact with the contractor, define regulations, cashflow, and requirements. It is not considered a major factor because the type of client is considered known and can be considered and taken into estimation. Figure 41 depicts 38% as moderate, 33% as high, and both low and very high rated as 13% and 15% respectively.

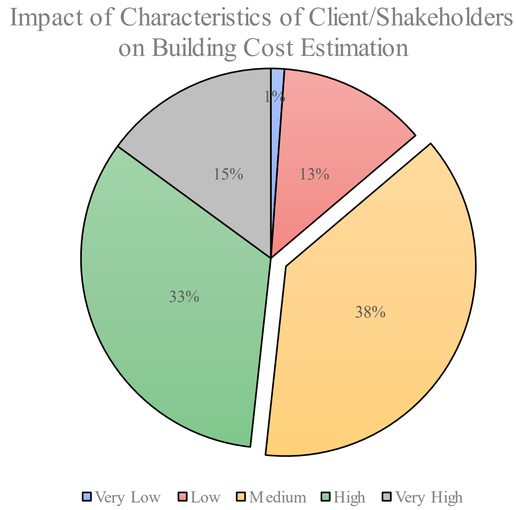


Figure 41: Population Rating Impact of Type of Client Factor on Cost Estimation (Total Sample of 87).

**Trend and Value of Market Indices (Stock Market Index, Construction Index, Inflation).**

The trend of market and inflation can have a big impact on cost estimation. For example, the economic boom of projects in Saudia Arabia in 2022 till date caused a big impact on prices of steel, aluminum, bitumen, and many other materials. Moreover, it led to a market shortage in skilled labor and engineers, increasing prices which affects the cost estimation especially for fixed cost projects. Most of the replies highlighted the impact as moderate to high with percentage of 33% and 31% respectively as shown in Figure 42. nevertheless, the risk price should not be raised high as this will decrease the chances of winning the tender of any future projects, this is why 30% replies rated very low to low as they can run estimations to decrease the impact on estimation.

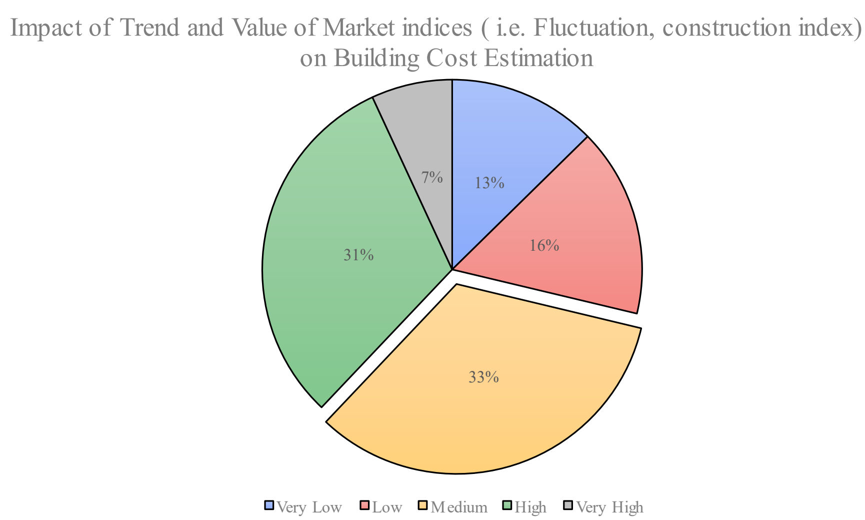



Figure 42: Population Rating Impact of Market Trend Factor on Cost Estimation (Total Sample of 87).

### External Social Conditions (I.E. Diseases, Wars and Conflicts).

A factor that is sometimes counted on force majeure and its impact will highly depend on the contract type is external social conditions such as conflicts and wars and the impact on the price of raw materials. Another example is diseases, although it is slightly considered now, it has not been considered prior to global spread of Covid-19 and caused many litigations and project delays. Figure 43 shows an almost equal distribution of 20% between all five ratings as opinions vary depending on the experience and designation. It is observed that most of the experienced managers opted for a rating of moderated or higher.

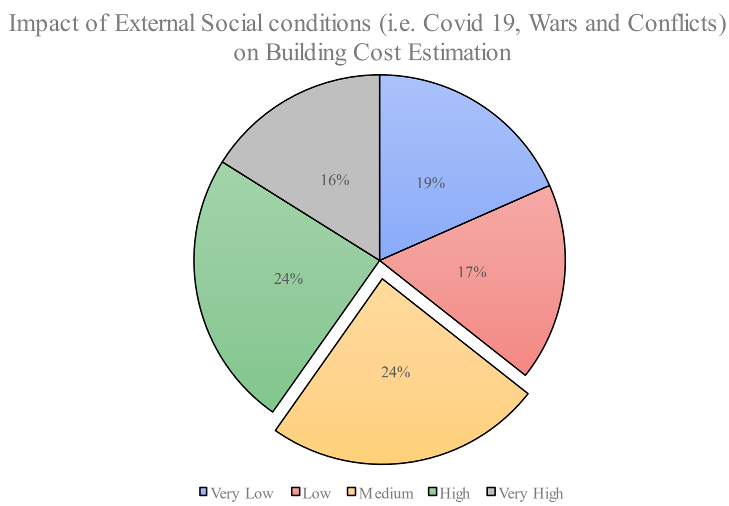



*Figure 43: Population Rating Impact of External Social Conditions Factor on Cost Estimation (Total Sample of 87).*

### Sustainability in Design and Construction (I.E Leed Building).

Project sustainability plays an important role in determining cost estimation. It has a strong impact on cost as it will require a more sophisticated design, greener materials (more expensive), require waste management and recycling process.

Figure 44 illustrates that 38% of replies show the impact as moderate, 30% as high, with 16% and 11% for low and very high respectively.

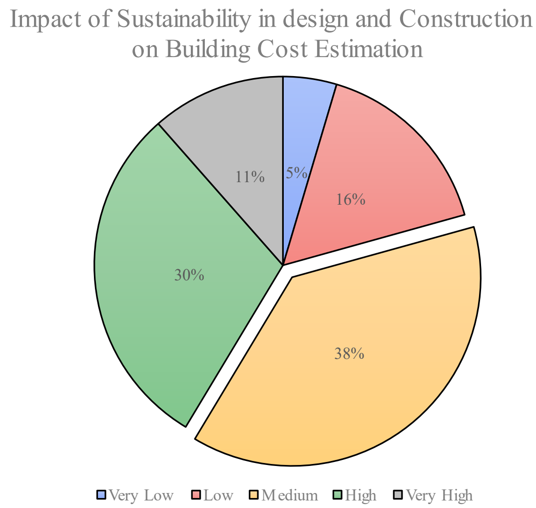



*Figure 44: Population Rating Impact of Project Sustainability Level on Cost Estimation (Total Sample of 87).*

**Construction Safety Regulations.**

A general rule of thumb in construction is that safety should never be price competitive. Safety is a primary cost that should not be mitigated or decreased. However, some projects might require high safety regulations, or the type of client might require a high standard of safety, this will have an increase of cost, but this cost impact is known and will be calculated in the initial estimate. This can be depicted in Figure 45 where 47% of replies were moderate, 26% as high impact and 15% as low.

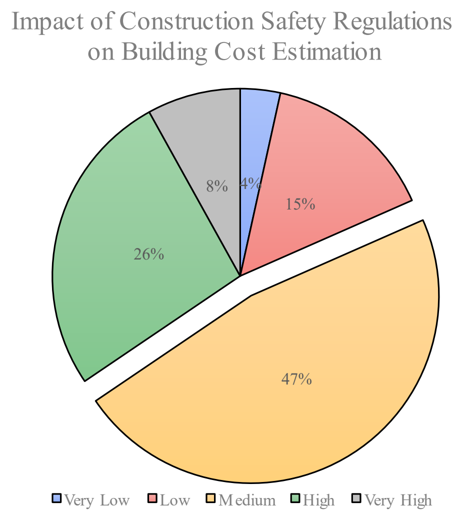



*Figure 45: Population Rating Impact of Safety Requirements Factor on Cost Estimation (Total Sample of 87).*

**Labor Union (Demand of Increase in Wages, Strikes, Etc.…).**

Labor union and syndicates can lead to increase in wages and can agitate workers to strike in extreme cases, where salary increase is rejected by the organization. This is not the case in the UAE, where expatriates know the work conditions beforehand and are being paid higher than their home countries, so they tend to accept their wages and rights. This explains why Figure 46 shows that factor is represented as moderate to low impact, where 35% replied as moderate, 18% as low and 23% as very low impact on cost.

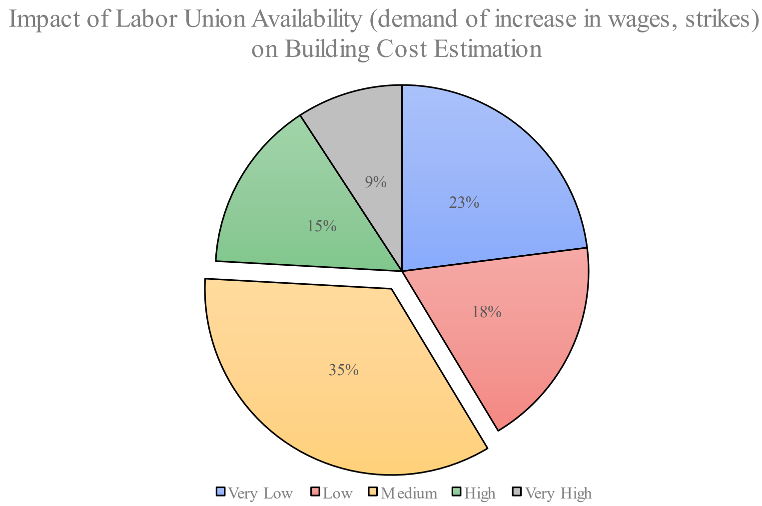



*Figure 46: Population Rating Impact of Labor Union Availability on Cost Estimation (Total Sample of 87).*

### *Data Analysis*

Table 2 below presents the top 10 factors, out of the 25 factors described above, that have the most significant impact on cost estimation based on the study results. The factors were determined based on the weighting assigned to each factor, derived from the total score for responses of the study population. The weighting for each factor was calculated by adding the total score then dividing it by the total number of replies.

The identified top 10 factors provide valuable insights into the key considerations for accurate cost estimation in the context of this study. However, to further analyze and develop results of these factors and ensure their relevance, the next step involves comparing them with the answers obtained from conducting interviews. This process will help confirm the importance and applicability of these factors in the next step of building the cost estimation model using ANNs. By analyzing and developing these factors, the objective is to enhance the accuracy and reliability of the cost estimation model, ultimately providing a more robust and effective tool for estimating project costs in the construction industry.

*Table 2: Top 10 Factors for Impact on Cost Estimation.*

| FACTORS | TOTAL SCORE | Weight | Ranking |
|---|---|---|---|
| Complexity of the Project | 348 | 0.800 | 1 |
| Availability of skilled labor/crew | 324 | 0.745 | 2 |
| Soil Type | 312 | 0.717 | 3 |
| Type of structure material (Steel, Concrete, Prestressed concrete, Precast) | 303 | 0.697 | 4 |
| Number of floors in the building | 303 | 0.697 | 4 |
| Characteristics of client/Shareholders | 303 | 0.697 | 4 |
| Type of used foundation in the building | 296 | 0.680 | 5 |
| Location of project (country, city) | 296 | 0.680 | 5 |
| Earthquake zone | 294 | 0.676 | 6 |
| Usage of Building Information Modelling (BIM) | 294 | 0.676 | 6 |
| Type of slab (Solid, Ribbed) | 289 | 0.664 | 7 |
| Type of contract (Lump sum, Remeasured) | 286 | 0.657 | 8 |
| Sustainability in design and Construction (i.e., Leed building) | 285 | 0.655 | 9 |
| Area of typical floor | 279 | 0.641 | 10 |
| Construction Safety regulations | 279 | 0.641 | 10 |

**Expert Review**

To enhance the comprehensiveness of factors related to cost estimation, expert reviews are conducted as part of the data development process through interviews. To gain a holistic understanding of cost estimation processes, the expert review in this study included professionals such as estimation managers, contracts managers, planning managers, technical managers, and a general manager. Each of them is involved in project tendering, offering unique perspectives and inputs. Consequently, interviews were conducted with all of them to capture a comprehensive view of the cost estimation process. The results of the reviews were integrated in the data results explaining factors of cost estimation, and their conclusion highlighting most important factors are described below.

***Interview 1: Contracts Manager***

For the first review, the cost estimation factors depended on how they are integrated in the project and how deeply they affect cost estimation in case they are miscalculated. The most important factors extracted from study are:

1. Type of Contract
2. Complexity of the Project
3. Location of the Project
4. Number of Floors
5. Soil Type
6. Type of Client
7. Safety Regulation
8. Sustainability Requirements

***Interview 2: Estimation Manager***

For the estimation manager review, the replies highlighted better what are the risk in each of the factors, how interconnected they are and where risk contingency should be placed. The most important factors extracted from study are:

1. Type of Contract
2. Complexity of the Project
3. Location of the Project

4. Number of Floors
5. Soil Type
6. Area of Typical Floor
7. Type of the Structure Material
8. Trend Value of the Economy (Inflation, Construction Index)

In addition to the factors above, it is highlighted that duration and delivery method of the project are rated as high on cost estimation.

***Interview 3: Technical Manager***

For the third review, the replies highlighted the technical side where each of the factors can impact, highlighting design and material choice on the impact on cost estimation. The most important factors extracted from study are:

1. Type of Contract
2. Type of Foundation
3. Type of Client
4. Number of Floors
5. Soil Type
6. Area of Typical Floor
7. Type of the Structure Material
8. Usage of BIM

***Interview 4: General Manager***

During the review conducted with the general manager, particular emphasis was placed on the interconnected relationship between the type of contract, project delivery method, management conditions, and planning. These factors collectively influence the execution of the project and aim to minimize cost overruns. Notably, the fairness of contract bonds emerged as a prominent aspect of consideration. Based on the review, the following factors were identified as the most important:

1. Type of Contract
2. Complexity of the Project
3. Location of the Project
4. Management Conditions
5. Soil Type
6. Area of Typical Floor
7. Trend Value of the Economy (Inflation, Construction Index)
8. Number of Floors

In addition to the factors above, it is highlighted that duration and delivery method of the project are rated as high on cost estimation.

***Interview 5: Planning Manager***

The review conducted with the planning manager focused on how the factors can impact the duration of the project and hence impact the cost of the project. Based on the review, the following factors were identified as the most important:

1. Complexity of the Project
2. Type of Contract
3. Type of Foundation
4. Type of Structure material

5. Number of Floors
6. Type of slab
7. Area of Typical Floor
8. Location of the Project

### *Interview Final Analysis*

After consulting the results from all interviews, each with their cost estimation defined from their experience and judgement providing clearer understanding. The most common factors from all of the reviews are as noted down.

- Type of Contract.
- Complexity of the Project.
- Location of the Project.
- Number of Floors.
- Soil Type / Type of foundation.
- Area of Typical Floor.
- Trend Value of the Economy (Inflation, Construction Index).
- Type of Structure Material.
- Usage of Advanced Technologies such as BIM.

## Chapter Conclusion

The factors extracted from the literature, results from the replies of the survey shown in table 2 in page 93 and analysis from the expert reviews can point out the final factors to be considered for the next step of the study for building an ANN model as follow:

- Type of Contract.
- Complexity of the Project.
- Location of the Project.
- Number of Floors.
- Soil Type / Type of foundation.
- Area of Typical Floor.
- Trend Value of the Economy (Inflation, Market Index).
- Type of Structure Material.
- Usage of Advanced Technologies such as BIM.
- Sustainability Requirement.

# CHAPTER 5

## Conclusion and Recommendations

### Conclusion

In conclusion, this thesis aimed to determine the key factors that influence cost estimation in construction projects through the discussion and analysis of various aspects related to cost estimation. This has shed light on several important points. It is evident that accurate cost estimation plays a critical role in the success and profitability of construction projects. Through an extensive literature review, survey analysis, and interviews with industry professionals, several influential factors were identified and validated.

The findings of this research highlight the significance of considering various factors such as resource allocation, project complexity, market conditions, and technological advancements in cost estimation. These factors contribute to the overall accuracy and reliability of cost estimates and play a crucial role in project planning and decision-making.

The findings have contributed to enhancing the understanding of these factors and their significance in the construction industry, particularly in the context of the UAE market.

It is important to note that this research has its limitations, including its focus on specific factors and the unique characteristics of the UAE market. Further research is needed to explore other markets and expand the scope of factors considered. Additionally, ongoing advancements in technology and data availability provide opportunities for future studies to refine and improve cost estimation models.

The contributions of this thesis are twofold. Firstly, it provides a comprehensive understanding of the factors that influence cost estimation, thereby assisting practitioners in making informed decisions and improving cost management practices. Secondly, the determined factors will enable to develop ANN model in the PhD study that will offer a practical solution for accurate cost estimation, which can lead to enhanced project outcomes and improved profitability.

The development of the ANN model offers a promising approach to enhance cost estimation in construction. By considering key factors and training the ANN model on relevant data, the uncertainty associated with cost estimation can be reduced. This can lead to more informed decision-making, better project planning, and ultimately, improved project outcomes.

Overall, this thesis contributes to the body of knowledge in cost estimation by providing insights into the factors influencing costs, highlighting how each factor can affect the estimation and what are the most important points to consider while reviewing the factors highlighted in the study. It also offers an effective weighting for most important factors to be considered as input layer for the utilization of ANN models. The outcomes of this research have practical implications for construction professionals, enabling them to make more informed decisions and optimize cost estimation processes for successful project execution by more comprehensive explanation on each of the factors in the study.

**Recommendations**

Based on the findings of this master's study on determining factors that impact cost estimation, and considering the next step to builds an ANN model in the PhD study, the following recommendations are suggested:

Integration of Project Delivery Systems: In the development of the ANN model, consider incorporating various project delivery systems, such as design-bid-build, design-build, or construction management. This will enable a comprehensive analysis of how different delivery systems influence cost estimation and enhance the model's applicability in diverse project contexts.

Consideration of Project Duration: Include project duration as a significant factor in the ANN model. Analyze how the duration of a project impacts cost estimation and explore ways to account for the temporal aspects of construction projects in the model. This can be done through the inclusion of relevant variables or techniques to capture the time-related influences on cost.

Integration of Advanced Technologies: Incorporating advanced technologies such as BIM and data analytics can significantly improve the accuracy of cost estimation. The integration of these technologies should be further investigated with case studies on site, as it can improve the quality of input for the ANN model which will provide more comprehensive and reliable cost estimates.

Expansion to More Regions: To enhance the generalizability of the finding, collect data from multiple regions, not limited to the UAE. Incorporate data from different countries or regions to capture variations in construction practices, cost factors, and market dynamics. This broader dataset will improve the study robustness and broaden its application potential.

Continual Data Collection: As the construction industry evolves, it is crucial to continuously collect and update data on cost factors. This will ensure that the study results remain relevant and capable of capturing the latest trends and variables that influence project estimate.

Diverse and Representative Population: Ensure a diverse and representative population for data collection. This will require including a wider range of professionals involved in tendering and cost estimation to capture different perspectives and expertise. Consider including professionals from different organizations and backgrounds to improve the model's versatility.

Sensitivity Analysis: Conducting sensitivity analyses on the identified factors can provide clearer insights into their relative impact on cost estimation. This analysis can help prioritize and allocate resources more effectively, considering the factors that have the most significant influence on project costs.

By incorporating the recommendations into the forthcoming PhD study, it is envisaged that the resulting model will embody a paradigm shift in cost estimation within the construction industry. This endeavor seeks to culminate in the development of an innovative and cutting-edge ANN model, characterized by its enhanced sophistication and refined accuracy. Through this scholarly pursuit, the aim is to elevate the standards of cost estimation, thereby empowering industry professionals with a state-of-the-art tool for informed decision-making and optimized project planning.

REFERENCES


Dey, P. K. (2013). Project management: a systems approach to planning, scheduling, and controlling. CRC Press.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. Proceedings of COMPSTAT'2010, 177–186.

Nguyen, D., & Widrow, B. (1990, April). Neural networks for self-learning control systems. IEEE Control Systems Magazine, 10(3), 18–23. https://doi.org/10.1109/37.55119

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. Bradford Books.

Bishop, C. M. (2011, April 6). Pattern Recognition and Machine Learning. Springer. https://doi.org/10.1007/b9479810.1007/978-0-387-45528-0

Aggarwal, C. C. (2018, September 13). Neural Networks and Deep Learning: A Textbook. Springer. https://doi.org/10.1007/978-3-319-94463-0

Chakraverty, S., Sahoo, D.M., Mahato, N.R. (2019). Artificial Neural Network Terminologies. In: Concepts of Soft Computing. Springer, Singapore. https://doi-org.ezproxy1.lib.asu.edu/10.1007/978-981-13-7430-2_10

Kwak, Y. H., & Ibbs, C. W. (2000). Development of early cost estimating models for building projects. Journal of construction engineering and management, 126(3), 171-177.

Meredith, J. R., & Mantel Jr, S. J. (2012). Project management: a managerial approach. John Wiley & Sons.

Project Management Institute. (2017). A Guide to the Project Management Body of Knowledge (PMBOK® Guide) (6th ed.). Project Management Institute.

Karim, A., & Bubshait, A. (2013). Cost estimation using artificial neural networks in construction projects: a case study of Saudi Arabia. International Journal of Engineering Business Management, 5(1), 1-9.

LeCun, Y., Bengio, Y., & Hinton, G. (2015, May 27). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Haykin, S. (1999). Neural networks: a comprehensive foundation. Prentice Hall.

Bishop, C. M. (2006). Pattern recognition and machine learning (Vol. 4). Springer.

Construction Industry Institute. (1996). CII Best Practices Guide: Estimating for Construction. The University of Texas at Austin.

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

Davé, B. U. (2003). A Delphi approach to increase constructability of highway projects. Construction Management and Economics, 21(2), 133-144.

Fleming, Q. W., & Koppelman, J. M. (2016). Earned value project management. Project Management Institute.

Professional Guidance Document No. 01 - Guide to Cost Estimate Classification Systems. (Revised 2022). Library.aacei.org. https://library.aacei.org/ pgd01 /pgd01.shtml

Rad. (2002). Project estimating and cost management (1st edition). Management Concepts.

Juszczyk, M., Leśniak, A., & Zima, K. (2018). ANN Based Approach for Estimation of Construction Costs of Sports Fields. Complexity, 2018, 1–11. https://doi.org/10.1155/2018/7952434

Kraus, William E,P.E., C.C.E. (2007). Tools for Developing Mutual Understanding on Classification of Construction Costs. AACE International Transactions, ES181-ES185.

Ruwanpura, J. Y., & Bandara, Y. M. (2009). Analogy-based parametric cost estimating model for construction projects. Journal of Construction Engineering and Management, 135(7), 611-618.

Apanavičienė, R., & Daugėlienė, A. (2011). New Classification of Construction Companies: Overhead Costs Aspect / Naujas Statybos Imoni Uklasifikavimas: Pridėtinių Išlaidų Aspektas. Journal of Civil Engineering and Management. https://doi.org/10.3846/13923730.2011.625629

Nguyen, T. T., Deeds, D. L., & Carpenter, B. (2014). Analogy-based cost estimation in a complex domain: A study of defense acquisition. Journal of Business Research, 67(11), 2360-2366.

Hakami, W. (2017). A Qualitative –Quantitative Model of Cost Estimate for Construction Projects in Yemen [Doctoral Dissertation, SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY].

Mendes, E., Mosley, N., & Counsell, S. (2003). Analogy-based effort estimation: A review of the literature. Journal of Systems and Software, 66(2), 129-147.

Khosrowshahi, F. (1994). Analogy-based estimating: Past, present, and future. Construction Management and Economics, 12(5), 433-447. doi: 10.1080/01446199400000048

Krogh, A. (2008, February). What are artificial neural networks? Nature Biotechnology, 26(2), 195–197. https://doi.org/10.1038/nbt1386

Olawale, Y. A., & Sun, M. (2010). Cost and time control of construction projects: Inhibiting factors and mitigating measures in practice. Construction Management and Economics, 28(5), 509-526. doi: 10.1080/01446191003674519

Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. (2018). Introduction to Research in Education (10th ed.). Cengage Learning.

Jha, K. N., & Bhandari, S. (2016). Construction project cost estimation: An integrated approach. Routledge.

Kelton, W. D., Sadowski, R. P., & Sturrock, D. T. (2010). Simulation modeling and analysis (4th ed.). McGraw-Hill Education.

Al-Sabah, A. I., Abdul-Rahman, H., & Zayed, T. (2016). Monte Carlo simulation in construction projects: a review paper. Journal of Engineering and Applied Sciences, 11(3), 1946-1955.

Lim, C. S., & Mohamed, M. Z. (2013). A review of studies on parametric estimating in construction. International Journal of Civil Engineering and Geo-Environmental, 4(1), 117-124.

Chan, D. W., & Kumaraswamy, M. M. (1997). A comparative study of causes of time overruns in Hong Kong construction projects. International Journal of Project Management, 15(1), 55-63.

Artificial Neural Network. (n.d.). Artificial Neural Network. https://www.saedsayad.com/artificial_neural_network.htm#:~:text=Four%20types%20of%20transfer%20functions,%2C%20piecewise%20linear%2C%20and%20Gaussian.

AACE. (2020). *Recommended Practice No. 18R-97: Cost Estimate Classification System as Applied in Engineering, Procurement, and Construction for the Process Industries. TCM Framework: 7.3 - Cost Estimating and Budgeting*. 1–7. https://web.aacei.org/docs/default-source/toc/toc_18r-97.pdf

Alcineide, P., Gean, S., Luiz, M. F. M., Felipe, C. A., & Débora, de G. S. (2021). Cost forecasting of public construction projects using multilayer perceptron artificial neural networks: A case study. *Ingenieria e Investigacion*, *41*(3), 1–11. https://doi.org/10.15446/ing.investig.v41n3.87737

Alpaydin, E., & Bach, F. (2014). *Introduction to Machine Learning*. MIT Press. http://ebookcentral.proquest.com/lib/asulib-ebooks/detail.action?docID=3339851

Alqahtani, A., & Whyte, A. (2016). Estimation of life-cycle costs of buildings: regression vs artificial neural network. *Built Environment Project and Asset Management*, *6*(1), 30–43. https://doi.org/10.1108/BEPAM-08-2014-0035

Amade, B., Okon, E., & Akpan, P. (2015). *Project Cost Estimation: Issues and the Possible Solutions*. *May 2014*.

BCIS. (2012). Elemental standard form of cost analysis. *Bcis*, 108.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Clarendon Press.

Chakraverty, S., Deepti, ·, Sahoo, M., Nisha, ·, & Mahato, R. (2019). *Concepts of Soft Computing Fuzzy and ANN with Programming*.

Chandanshive, V. B., & Kambekar, A. R. (2019). Estimation of Building Construction Cost Using Artificial Neural Networks. *Journal of Soft Computing in Civil Engineering*, *3*(1), 91–107. https://doi.org/10.22115/SCCE.2019.173862.1098

Chen, X.-W., & Lin, X. (2014). Big Data Deep Learning: Challenges and Perspectives. *IEEE Access*, *2*, 514–525.

Chollet, F. (2018). *Deep learning with Python* (1st edition). Manning Publications Co.

Elfaki, A. O., Alatawi, S., & Abushandi, E. (2014). Using intelligent techniques in construction project cost estimation: 10-Year survey. *Advances in Civil Engineering*, *2014*. https://doi.org/10.1155/2014/107926

Elmousalami, H. H. (2020). Artificial Intelligence and Parametric Construction Cost Estimate Modeling: State-of-the-Art Review. *Journal of Construction Engineering and Management*, *146*(1), 03119008. https://doi.org/10.1061/(asce)co.1943-7862.0001678

Feng, G., & Li, L. (2014). Application of Genetic Algorithm and Neural Network in Construction Cost Estimate. *Advanced Materials Research*, *756–759*. https://doi.org/10.2991/iccia.2012.254

Geisser, S. (1975). The Predictive Sample Reuse Method with Applications. In *Source: Journal of the American Statistical Association* (Vol. 70, Issue 350).

Geron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : concepts, tools, and techniques to build intelligent systems* (2nd ed). O'Reilly.

Goodfellow, I., Bengio Yoshua, & Courville Aaron. (2016). *Deep Learning*.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018). *A Survey Of Methods For Explaining Black Box Models*. https://doi.org/10.48550/arxiv.1802.01933

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

Hagan, M., Demuth, H., De Jesús, O., & Beale, M. (2014). *Neural Network Design 2nd Edition*.

Hanna, A. S., Camlic, R., Peterson, P. A., & Lee, M.-J. (2004). Cumulative Effect of Project Changes for Electrical and Mechanical Construction. *Journal of Construction Engineering and Management*, *130*(6), 762–771. https://doi.org/10.1061/(asce)0733-9364(2004)130:6(762)

Haykin, S. S. (2009). *Neural networks and learning machines*. Prentice Hall/Pearson.

Heravi, G., & Eslamdoost, E. (2015). Applying Artificial Neural Networks for Measuring and Predicting Construction-Labor Productivity. *Journal of Construction Engineering and Management*, *141*(10), 04015032. https://doi.org/10.1061/(asce)co.1943-7862.0001006

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/https://doi.org/10.1016/j.ijforecast.2006.03.001

Ibrahim, A. H., & Elshwadfy, L. M. (2021). Factors affecting the accuracy of construction project cost estimation in Egypt. *Jordan Journal of Civil Engineering*, *15*(3), 329–344.

Jinisha, R., & Jothi, R. (2019). Construction Prediction of Cost Overruns. *International Journal of Innovative Research in Science, Engineering and Technology - IJIRSET*, *8*(3), 2695–2699. https://doi.org/10.15680/IJIRSET.2019.0803160

Juszczyk, M., Leśniak, A., & Zima, K. (2018). ANN Based Approach for Estimation of Construction Costs of Sports Fields. *Complexity*, *2018*. https://doi.org/10.1155/2018/7952434

Karaca, I., Gransberg, D. D., & Jeong, H. D. (2020). Improving the Accuracy of Early Cost Estimates on Transportation Infrastructure Projects. *Journal of Management in Engineering*, *36*(5), 04020063. https://doi.org/10.1061/(asce)me.1943-5479.0000819

Keith, P., & Nii, A. (2013). *Construction Cost Management Second edition*.

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

Kohavi, R. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. http//roboticsStanfordedu/"ronnyk

Kwon, H., & Kang, C. W. (2019). Improving Project Budget Estimation Accuracy and Precision by Analyzing Reserves for Both Identified and Unidentified Risks. *Project Management Journal*, *50*(1), 86–100. https://doi.org/10.1177/8756972818810963

Len Holm, J. E. S. (2021). *Construction Cost Estimating*. Routledge. https://doi.org/https://doi.org/10.1201/9781003023494

Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H. (2019). An artificial neural network approach for cost estimation of engineering services. *International Journal of Construction Management*, *0*(0), 1–14. https://doi.org/10.1080/15623599.2019.1692400

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Najafi, A., & Tiong, R. (2015). Construction productivity and cost estimation using artificial neural networks. *Measuring Construction: Prices, Output and Productivity*, 224–236. https://doi.org/10.4324/9781315882925

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Samphaongoen, P. (2010). A Visual Approach to Construction Cost Estimating. *Masters Thesis, Marquette University*, 1–113.

Shehatto, O. (2013). *Cost Estimation for Building Construction Projects in Gaza Strip Using Artificial Neural Network (ANN)*.

Shiha, A., Dorra, E. M., & Nassar, K. (2020). Neural Networks Model for Prediction of Construction Material Prices in Egypt Using Macroeconomic Indicators. *Journal of Construction Engineering and Management*, *146*(3), 04020010. https://doi.org/10.1061/(asce)co.1943-7862.0001785

Hashemi, S. T., Ebadati, O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. *SN Applied Sciences*, *2*(10), 1–27. https://doi.org/10.1007/s42452-020-03497-1

Tijanić, K., Car-Pušić, D., & Šperac, M. (2020). Cost estimation in road construction using artificial neural network. *Neural Computing and Applications*, *32*(13), 9343–9355. https://doi.org/10.1007/s00521-019-04443-y

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, *7*. https://doi.org/10.1186/1471-2105-7-91

Vikas, Kuldeep, & Samiksha. (2011a). Cost prediction using Neural Network Learning Techniques. *International Journal of Computer Science and Management Studies*, *11*.

Vikas, Kuldeep, & Samiksha. (2011b). Cost prediction using Neural Network Learning Techniques. *International Journal of Computer Science and Management Studies*, *11*.

Wang, R., Asghari, V., Cheung, C. M., Hsu, S. C., & Lee, C. J. (2022). Assessing effects of economic factors on construction cost estimation using deep neural networks. *Automation in Construction*, *134*(December 2021), 104080. https://doi.org/10.1016/j.autcon.2021.104080

# APPENDIX A

## SURVEY SAMPLE USED TO COLLECT STUDY RESULTS

# Survey: Influential factors on Construction Cost Estimation

My name is Salem Al Saber, and I am a graduate engineering student at Arizona State University (ASU) working on my MSc thesis. This survey is conducted in partial fulfillment of the requirements for the Degree Master of Science in Construction Management and Technology at Arizona State University.

All the information you provide for this survey shall be kept confidential to protect your interests and the interests of your company. All the information collected by this survey shall be used for the sole purpose of research at Arizona State University.

The objective of this survey is to find out the major influences on cost estimation in construction projects through this quantitative approach, using the questionnaire, that could help produce a cost estimation technique using Artificial Neural Networks (ANN).
This survey is divided into 2 (two) sections as follows:
Section I: Demography
Section II: Influential factors on the estimation process in the construction

Completion of the survey will take approximately 15 minutes.

* Indicates required question

1. Email *

2. Name (optional)

Section (1) Demography

3. 1- You are representing *

*Mark only one oval.*

- Owner/ Owner Representative
- Engineering Consultant
- Contractor
- Other: ____________________

4. 2- Your current Role regarding construction is *

*Mark only one oval.*

- Owner/Managing partner
- Higher Management / Project Manager
- Construction Manager
- Designer / Consultant
- Project/Site Engineer
- Technical Dep.
- Planner /QS
- Other: ____________________

5. 3- Please indicate your years of industry experience *

*Mark only one oval per row.*

| | 0-2 years | 3-5 years | 6-10 years | 11-20 years | Over 20 years |
|---|---|---|---|---|---|
| **US/Canada Market** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **European Market** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Middle East Market** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Australian Market** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Asian Market** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |

6. 4- How much do you rate the usage of advanced technologies in your organization (i.e. BIM, Advanced estimation softwares) *

*Mark only one oval.*

- ⬭ Not Sure
- ⬭ 10%
- ⬭ 25%
- ⬭ 50%
- ⬭ More than 50%

Section (2) Influential factors on the estimation process in the construction

7. 1- Do you think inaccurate estimation is one of the reasons for cost overruns in Construction projects ? *

*Mark only one oval.*

- Yes
- No
- Not Sure

If yes, please answer question 2; if not please choose Not Sure

8. 2- How common do projects face cost overruns or overbudget issues due to inaccurate estimation *

*Mark only one oval.*

- Not Sure
- 10%
- 25%
- 50%
- More than 50%

9. First : Factors related to the structural work construction phase (building the core structure) *

This means that as much you see below items, you will expect a rise in price estimation

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval per row.*

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Type of used foundation in the building** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Type of structure material (Steel, Concrete, Prestressed concrete, Precast ...)** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Area of typical floor** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Number of floors in the building** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Type of slab (Solid, Ribbed ...)** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Number of staircases in the building** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Type of contract (Lump sum, Remeasured etc)** | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |
| **Length of spans** between | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ |

| | | | | | |
|---|---|---|---|---|---|
| Area of Shear walls in project | ○ | ○ | ○ | ○ | ○ |
| Location of project (country, city) | ○ | ○ | ○ | ○ | ○ |
| Use of building | ○ | ○ | ○ | ○ | ○ |
| Need for Dewatering System | ○ | ○ | ○ | ○ | ○ |
| Earthquake zone | ○ | ○ | ○ | ○ | ○ |
| Soil Type | ○ | ○ | ○ | ○ | ○ |
| Complexity of the Project (i.e. Burj Arab) | ○ | ○ | ○ | ○ | ○ |

Second : Other Various Factors impacting construction projects economy

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

10. 1. Usage of Building Information Modelling (BIM) *

Think of the following point from positive and negative side to weight the effect on cost estimation
Does having a BIM model for the project improve estimation for the project ?
Does having a BIM model required in scope of work increase the cost of the Tender?
Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

Very low to none

1 ⬭

2 ⬭

3 ⬭

4 ⬭

5 ⬭

Very high

11. 2. Weather conditions (Severe fog, Sand storm, rain…) *

If the project is located in a country with severe (Harsh) weather conditions, such as high summer temperatures, severe cold degrees, heavy rain in winter... Does this affect Cost Estimation of the project? Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

Very low to none

1 ⬭

2 ⬭

3 ⬭

4 ⬭

5 ⬭

Very high

12. 3. Management Conditions (Availability of PMC, certified project manager ...) *

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

| | |
|---|---|
| | Very low to none |
| 1 | ○ |
| 2 | ○ |
| 3 | ○ |
| 4 | ○ |
| 5 | ○ |
| | Very high |

13. 4. Availability of skilled labor/crew *

Does project estimation change if the area have little labor available or many projects are open limiting labor to hire ?

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

| | |
|---|---|
| | Very low to none |
| 1 | ◯ |
| 2 | ◯ |
| 3 | ◯ |
| 4 | ◯ |
| 5 | ◯ |
| | Very high |

14. 5. Characteristics of client/Shareholders *

Does a client or shareholder could have an impact on the project estimation?

Number for Authority inspection (ADM approval) for each stage?

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

| | |
|---|---|
| | Very low to none |
| 1 | ⬭ |
| 2 | ⬭ |
| 3 | ⬭ |
| 4 | ⬭ |
| 5 | ⬭ |
| | Very high |

15. 6. Trend and Value of Market indices ( i.e. Stock market index, daily wages, construction index) *

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

Very low to none

1 ⬭

2 ⬭

3 ⬭

4 ⬭

5 ⬭

Very high

16. 7. External Social conditions (i.e. Covid 19) *

What was the impact on project estimation during the Pandemic Era (2020-2023) and does it still have any impact on estimations? Rate your answer based on effect during this time line and in future not on past projects

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

| | |
|---|---|
| | Very low to none |
| 1 | ⬭ |
| 2 | ⬭ |
| 3 | ⬭ |
| 4 | ⬭ |
| 5 | ⬭ |
| | Very high |

17. 8. Sustainability in design and Construction (i.e Estidama/ Leed building) *
How does Sustainability affect project estimation. Does having sustainable design increase or decrease the project cost?

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

| |
|---|
| Very low to none |
| 1 ◯ |
| 2 ◯ |
| 3 ◯ |
| 4 ◯ |
| 5 ◯ |
| Very high |

18. 9. Construction Safety regulations *

Does projects with high safety standards increase project estimation? (Cost of 5 points PPE, Fall preventive measures, intensive training requirements, etc...)

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

Very low to none

1 ⬭

2 ⬭

3 ⬭

4 ⬭

5 ⬭

Very high

19. 10. Labor Union (demand of increase in wages, strikes, etc ...) *

Degree of effect on cost estimation 1=very low to none 2=low 3 =moderate 4= high 5= very high

*Mark only one oval.*

Very low to none

1 ⬭

2 ⬭

3 ⬭

4 ⬭

5 ⬭

Very high

20. If you have any recommendations for improving the cost estimation process , please list them below: (Optional)

APPENDIX B

IRB FORM

EXEMPTION GRANTED

Kristen Hurtado
IAFSE-SEBE: Construction, Del E. Webb School of
-
Kristen.Hurtado@asu.edu

Dear Kristen Hurtado:

On 7/17/2023 the ASU IRB reviewed the following protocol:

| | |
|---|---|
| Type of Review: | Initial Study |
| Title: | Investigation of Factors Impacting Construction Cost Estimate to Develop Construction-Driven Artificial Neural Network (ANN) |
| Investigator: | Kristen Hurtado |
| IRB ID: | STUDY00018291 |
| Funding: | None |
| Grant Title: | None |
| Grant ID: | None |
| Documents Reviewed: | • IRB Requirement question 8.pdf, Category: IRB Protocol; |

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2)(ii) Tests, surveys, interviews, or observation (low risk) on 7/17/2023.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

Sincerely,

IRB Administrator

cc: Salem Al Saber