Metropolia
University of Applied Sciences

Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences

Evaluation of Indirect Cost Estimation in the Egyptian Construction Industry

Master Thesis

International Master of Science in Construction and Real Estate Management

Joint Study Programme between HTW Berlin and Metropolia UAS

from

Sherif Othman

s0567958

Date:

Berlin, 29.10.2020

1st Supervisor: Prof. Dr.-Ing. Dieter Bunte

2nd Supervisor: MSc. -Ing. Ammar Al-Saleh

[Acknowledgement]

I would like to thank everyone who supported me throughout this journey, especially my first supervisor Professor Dieter Bunte who always guided me, and my second supervisor Engineer Ammar Al-Saleh.

I would also like to thank all the professors in HTW Berlin and Metropolia UAS for their support and patience.

I would like to express my gratitude and appreciation to my mother for her unconditional love, motivation, and encouragement, which are the reasons for bringing my work to light.

Last but not least, my father who supported me during his illness, I cannot find words do thank you for your advice and I hope you get well soon.

Metropolia

htw. Hochschule für Technik und Wirtschaft Berlin University of Applied Sciences

International Master of Science in Construction and Real Estate Management
Joint Study Programme of Metropolia Helsinki and HTW Berlin

HTW Berlin Registration Number: **s0567958**

Surname: **Othman** First Name(s): **Sherif**

Specifications for the Master Examination
according to § 9 and 10 Examination Regulations for the Master Study Programme Construction and Real Estate Management

1. Master Thesis, Examination Commission

(1) Topic of the Master Thesis

Evaluation of Indirect Cost Estimation in the Egyptian Construction Industry

(2) Examination Commission

Chairperson ______

1. Supervisor **Dieter Bunte**

2. Supervisor **Ammar Al-Saleh**

(3) Thesis Period

Date of Issue **27/06/2019**

Closing Date **30/06/2020**
Deadline

Berlin, ______

Signature of the Chairperson of the Examination Board

2. Issue of the Topic of the Master Thesis

Date of Issue **27/06/2019** issued by **Sherif Othman**

I have taken note of the subject and the deadline of the master thesis as well as the formation of the examination commission. It is known to me that I have to submit three (3) written and bound copies of the master thesis alongside a digital version as well as the additional statuary declaration at Metropolia Helsinki or HTW Berlin.

Berlin, **30/01/2020**

Sherif. Othman
Signature of the student

3. Release of the Master Thesis

Date of Release ______ received by ______

Distribution: original for the student, copies for audit file, chairperson and supervisors of the examination commission and chairperson of the examination board

Abstract

In construction projects, site overhead costs are an essential component of the contractor's budget as that what mainly differs a company from another when bidding for a project. The purpose of this research is to reduce the risk, enhance the bid accuracy, and minimize the amount of time and effort in estimating the site overhead costs. This thesis will focus on developing an artificial neural network model that allows a fast, efficient, and accurate estimate for the percentage of site overhead costs for construction projects executed in Egypt to avoid any reduction in the company's profit or incompletion of the project.

This research has three folds; the first fold includes the literature review for the site overhead costs and identifies the significant factors that affect it from previous studies which are project type, duration, location, budget, client type, contract type, company category, extra manpower required, special site requirements, and contractor joint venture. Research articles and publications are included in this section to review the ability of artificial neural networks in cost estimation and other areas related to construction management. Everything in this fold was used as a foundation for the second fold. The second fold is mainly concerned with data collection and analysis through a questionnaire sent to construction companies. Forty projects were collected and analysed to measure the impact and weight of each factor on the site overhead costs. Finally, the third fold includes developing and testing the model using Neural Designer software by coding the data collected and using it as a database to develop the model. The development process included several models with a multilayer perceptron, different activation functions, and several neurons.

The best model was selected based on the lowest RMSE value with 0.2920 and five projects were kept for testing the model and showed a high correlation coefficient of R^2=0.888. The selected model was exported to python language and viewed using Visual Studio software 2019.

Keywords: Site overhead cost, Artificial neural networks, Construction cost management

Table of Contents

Table of Figures

List of Tabulations

List of Abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
LEED	Leadership in Energy and Environmental Design
MSE	Mean square error
OH	Overhead
RMSE	Root mean square error

1. Introduction

1.1. Overview

The construction industry is one of the most competitive sectors. It has many complicated challenges in several areas, such as cost estimation, delays in building activities, tendering, risk analysis, and disputes. (Kulkarni, et al., 2017).

The effect of overhead (OH) cost estimates is essential for the financial position of the construction company. In general, the OH costs of the construction contractor are divided into two categories: OH costs for the site project, which will be the main topic of the research, and office OH costs (Peurifoy, 2014).

Artificial Intelligence (AI) techniques such as artificial neural networks (ANN), case-based reasoning, classifiers and learning processes, fuzzy logic, hybrid approaches, and genetic algorithms are commonly used in construction management to overcome the challenges. During the last two decades of the twentieth century, there has been a rise during publications dealing with Artificial Intelligence techniques and, in particular, ANN in various areas of construction management (Kulkarni, et al., 2017).

ANN approach will be used in this research to build a model capable of predicting the site OH cost in the Egyptian construction market and will be used for data analysis in identifying the impact of each factor on the site OH cost.

1.2. Problem statement

The construction companies estimate the direct and indirect cost of the project at the bidding stage. Most of the companies in the Egyptian market have systems and tools in evaluating the direct costs of the project. However, estimating the OH costs is considered a challenge as it is affected by many factors, which creates a deviation between the estimated and the actual cost in Egypt, especially after the revolution in 2011 and the currency devaluation in 2016.

A wrong estimation for the site OH costs may result in losing the bid if it was high or a profit loss if it was a low estimation which may lead to incompletion of the project. The site OH costs are affected by many factors that will be discussed in detail as project location, budget, duration, and other factors. These factors complicate the site OH costs estimation process; therefore, developing an ANN model is necessary to assist the construction companies in this process.

1.3. Research Objectives

The aim of this research to identify and analyse the factors affecting the site OH cost in the Egyptian construction industry and use the AI in assisting the construction companies to enhance the bid's accuracy and that would lead to the below:

- Reduce the risk and enhance the bid accuracy.
- Minimize the effort and time for the site OH cost estimation.
- Awareness of the impact and weight of each factor on the estimated budget.
- Integrating the AI in solving the construction industry's challenges by developing an ANN model to predict the percentage of site OH cost.

1.4. Research questions

This research aim is to answer the following questions:

- What are the direct and indirect costs?
- What is the site's indirect cost, and it is important in the tendering phase?
- What are the main factors affecting the site's indirect cost estimation?
- What are the weight and impact of each factor?
- Is there an ideal percentage for the site indirect cost estimation from the total budget considering all those factors in Egypt?

1.5. Research Methodology

The methodology used in this research is in the following sequence:

- Identifying the difference between direct and indirect costs
- Identifying the factors affecting the site OH costs from previous studies.
- Literature review of previous works related to the ability of ANN in cost estimation.
- Data collection for construction projects based on these factors.

- Data analysis and identify the impact and weight of each factor on the estimated percentage of site OH costs.
- Illustrating the factors that have a higher or lower impact on the estimation process.
- Explaining the components and structure of the ANN.
- Mentioning the steps to develop a model using Neural Designer software.
- Dataset coding and preparation to be readable by the software.
- Data analysis using the software to verify the results from the previous manual data analysis with showing the actual correlation and weight of each factor on the site OH costs.
- Develop an ANN model to predict the percentage of site OH costs.
- Export the developed model to Python language, which is the primary language for machine learning.

1.6. Scope and limitation of study

In order to achieve the required objective, the research focuses on big construction projects. The scope of this research is limited to the following:

- Construction projects executed in Egypt.
- Construction projects executed by main contractors in Egypt.
- Construction companies, the first and second categories only.

1.7. Thesis organization

This research consists of six chapters. Chapter one discussed the introduction, problem statement, research objective, and research methodology.

Chapter two deals with the principles of cost estimation, direct costs, indirect costs, general OH costs, and that it consists of two types: office OH costs and site OH costs. Identifying the main factors affecting the site OH cost in the Egyptian construction industry. Explaining the concept of ANN and showing from previous works the integration of ANN in the construction field. This chapter is considered a base for chapter three.

Chapter three deals with data collection of construction projects through a questionnaire based on the factors mentioned in chapter two. Also, this chapter includes a method of analysis of the collected data.

Chapter four explains the content of the ANN and steps to develop the model using Neural Designer software.

Chapter five is concerned with developing, selecting, and testing the ANN model based on coding the data collected from chapter three and following the steps mentioned in chapter four. The model objective is to estimate the percentage of site OH costs in construction projects.

Chapter six presents a summary of the research, conclusion, and recommendations with other future measures.

2. Literature review

2.1. Introduction

The objective of this chapter is to establish theoretical information on the idea of site OH cost in construction projects and the implementation of AI using the ANN tool to solve the challenges in the construction industry.

The literature review includes a definition of the cost estimation, direct cost, indirect costs, general OH cost, types of general OH costs, which are office OH cost and site OH cost, which is the main focus of the research, factors affect the site OH cost in the Egyptian construction industry, ANN and previous work related to the study area. The sources used in this chapter are books, academic research, scientific journals, thesis, reports, and papers.

2.2. Definition

2.2.1. Cost Estimation

Experts and Researchers have various definitions of cost estimation. (The Chartered Institute of Building, 2009) defined cost estimating as a process of predicting the construction costs.

(AACE International Recommended Practice, 2019) defined cost estimating as the basic principles for cost control, business planning, and project management.

(The project Management Institute, Inc., 2013) defined cost estimating as an assessment or prediction of the total costs of resources and materials required to complete it.

2.2.2. Direct Costs

Direct costs were defined as the labour costs, equipment, supplies, and materials integrated for project completion (Pratt, 2011).

(AACE International Recommended Practice, 2019) defined the direct costs as the costs of work completion, including resources involved in the physical construction of the project.

2.2.3. Indirect cost

(Consultants Estimating Manual Commonwealth of Massachusetts, 2006) identified the OH costs as the costs that cannot be assigned to an activity in the project. Direct costs are costs that may be specific to an item or activity and cannot be applied to the OH costs. They divided the OH costs into office OH costs and site OH costs.

(Stolz, 2010) started by explaining that cost estimation consists of direct costs, indirect costs, and profit. Direct costs are those that are directly attributable to the implementation of specific project activity. Indirect costs, on the other hand, are expenses that support the project, which is often called OH costs.

(Pratt, 2011) stated that the indirect costs include other things, such as OH, interest for construction durations, inflation, risk contingencies, and which are not included in the direct costs.

(Peurifoy, 2014) mentioned that indirect costs consist of additional items, such as the risk, profit, contingencies, contractor's OH interest for the construction duration. These items are not part of the completed work.

(AACE International Recommended Practice, 2019) defined the indirect costs as costs that are not explicitly related to the execution of activity; however, they are distributed on all project activities. In the construction industry, the indirect costs may include contractor's fees, direct supervision, filed administration, taxes, and insurance

2.2.4. General overhead costs

(The Institute of Cost Accountants of India, 2013) described general OH costs as the costs which are not included in the direct costs such as indirect expenses, indirect employee, and indirect materials.

(Emerging professional‘s companion, 2013)mentioned that the general conditions in the construction industry usually refers to OH costs. These OH costs include consultant engineers, site supervisors, small facilities or tools, and security. It also includes loans, licenses, permits, and the costs of insurance assigned to the project.

2.3. Types of general overhead costs

(Patil & Bhangale, 2014) demonstrated that in a construction company, there are two types of OH costs, which are and site OH and company OH costs. The OH costs are also referred to as general and administrative OH, which includes all the costs needed by the construction company to run the company and keep it in operation and support the business. These costs are not explicitly related to a particular project. The site OH cost could be defined as the expense for site management and operation, which is the cost related to a project and is not included in a specific activity.

2.3.1. Office overhead costs

Office OH cost is commonly referred to as the contractor's administrative expenses for all existing projects to support the business, and these costs are not related to a specific project. Contractors are free to take these factors into account in whatever way they choose. However, they must still use the same framework on all agreements (Zack, 2002).

(Shelton & Brugh, 2002) explained the office OH costs as administrative costs to keep the performance of the company. They mentioned examples for the costs categorized under the office overhead costs such as office rent, office salaries, office furniture, taxes, permits, software costs, and utilities.

In other words, the office OH costs reflect the costs of the operations of the contractor to manage the company and to support different projects. It is challenging to allocate office OH cost to a specific project, and it includes the following (Lorman, 2014):

- Accounting
- Management
- Rent
- Sales and marketing
- Insurance

2.3.2. Site overhead costs

(Chan & Pasquire, 2002) defined the site OH costs as the project's direct cost, which are necessary for the site work and site accommodation.

(Lowe, et al., 2003) defined the site OH costs as costs of administration and managing a particular project.

(Ruf & Ruf, 2007) gave some examples for the site overhead costs; however, they are not limited to the below:

- site office rent
- Site office expenses
- site utilities
- Site security
- Site trash removal
- Insurance
- Safety Supplies
- Telephones

(Dagostino, 2002) stated that the OH costs include products that may be associated with a specific project. The site OH involves costs not explicitly related to a specific activity but necessary to build the project but not equipment, labour, or materials.

(Abdul-Malak, et al., 2002) mentioned that the increased OH costs are also more readily quantified. The contractor shall report the preliminary installation of the site, showing comprehensive costs of all items that are classified as general site items infrastructure of the site, cranes, and general equipment.

(Shelton & Brugh, 2002) stated that the indirect costs for contracts include indirect labour, contract management, machinery and equipment, procurement, quality assurance, testing, insurance, maintenance, servicing, depreciation, amortization, and support costs such as central planning and the elimination of payrolls in some situations.

(Lowe, et al., 2003) discussed the labour OH, a particular chapter in site OH costs. The labour OH is overhead on the salary, which includes holidays, sick leave, unemployment, retirement, medical insurance, and social security.

2.4. Factors that affect Site overhead costs

(Assaf, et al., 2001) identified the factors influencing the OH and discussed how contracting firm's decision-makers do not stick to project estimator figures but rather adjust organization OH levels to higher or lower values and that decision is based on many factors which are the project's size, complexity, contract type, contract requirements, the agreed service level, the financial causes for contractors, the contractor's experience with the customer, the number of contractors that are competing to gain the contract.

(Eksteen & Rosenberg, 2002) discussed the factors affecting the site OH cost and how it is affected by the project's duration, size, nature. Also mentioned that the prediction of site OH costs is mainly based on previous project's data.

(El-Sawy, et al., 2011) listed most of the factors affecting the site OH costs based on previous studies and reports from 1980 till 2009 from different markets, as shown in Table 1. The estimation of site OH costs is essential and significant. This concern was reflected in the extensive research work to measure, define, and calculate site OH costs of construction projects.

#	Factors affected the site OH costs from 1980-2010
1.	Specialized subcontractor
2.	Amount of subcontractor work
3.	Supervision and consulting
4.	Type of contract
5.	Company need for the job
6.	Client type
7.	Site preparation
8.	Project size
9.	Project duration
10.	Special equipment on site
11.	Project delay
12.	Company experience
13.	Construction law and country policy
14.	The cash flow of the project
15.	Project location
16.	Category of the construction company

Table 1. All Factors affect the site overhead cost percentage from different markets (El-Sawy, et al., 2011)

Analysis of the questionnaires collected revealed differences between the factors that affect the site OH costs in the Egyptian construction market and the general factors summarized above, as shown in Table 1. Many factors in Egypt are not taken into account because of their irrelevance on the local market. However, they contribute significantly to the construction markets in Europe and North and South America. There is a tendency between contractors in Egypt to combine several factors in one key factor, which the academics rejects and defined to be unprofessional because it entirely depends on the person who is doing the task and on her/his experience. Therefore, after analysis of the collected questionnaire obtained by academics and contractors in Egypt, and after appropriate alterations were made, a list of all factors affecting the site OH cost related to the Egyptian construction industry as shown in below in Table 2 (El-Sawy, et al., 2011).

#	Factors affect the site OH costs in Egypt
1.	Project budget
2.	Project type
3.	Project duration
4.	Project location
5.	Special site preparation
6.	Company category
7.	Client type
8.	Extra manpower required
9.	Contractor joint venture
10.	Contractor type

Table 2. Factors affect the site OH costs in Egypt (El-Sawy, et al., 2011)

These ten factors in Table 2, represent the main factors that affect the site OH cost in Egypt were used in this research in the data collection section and in building the model. However, these factors were reduced to seven factors to simplify and fasten the data collection process as people are working from home due to COVID-19, and the response rate will be too less to build the model. The selected seven factors used in this research are:

- Project type
- Project duration
- Project budget
- Project location
- Client type
- Contract type
- Company category

2.5. Artificial Neural Networks

Deep Learning is considered the most vital branch in Machine Learning. It is a technique that teaches machines to do something familiar to humans, which is learning by example. Deep learning is a crucial technology behind driverless vehicles that enable them to recognize or differentiate between pedestrians and a lampstand. It is essential in consumer systems such as smart TVs, telephones, voice control, TVs, and hand-free speakers. Lately, deep learning is getting much attention because it achieved results that were not possible earlier (Chauhan, 2019).

In deep learning, a computer model learning directly from pictures, sound, or text to classify tasks. Deep learning models are capable of achieving cutting-edge precision, exceeding human output often. Models are equipped using many data and several layers of neural network architectures for several complex tasks. Deep Learning can be represented in (Chauhan, 2019):

- Regression and classification by ANN.
- Time Series analysis by recurrent neural networks.
- Computer vision by convolutional neural networks.
- Feature extraction by self-organizing maps.
- Recommendation systems by Auto Encoders.
- Recommendation systems by deep Boltzmann machine.

The first recognized knowledge of ANN dates to 1940, but 40 years later, they were introduced to detect adequate algorithms that greatly expanded their use. There are already many theoretical studies and a growing interest in neural networks, and many universities around the world study them. Neural networks have found practical application in numerous fields and are used as a way of solving several complex and difficult problems (Lazarevska, et al., 2014).

This research will focus on ANN. ANN is a computer model focused on biological neural network structure and function. Information flowing through the networking influences the structure of the ANN. The ANN learns based on the inputs and the type of required output (Chauhan, 2019).

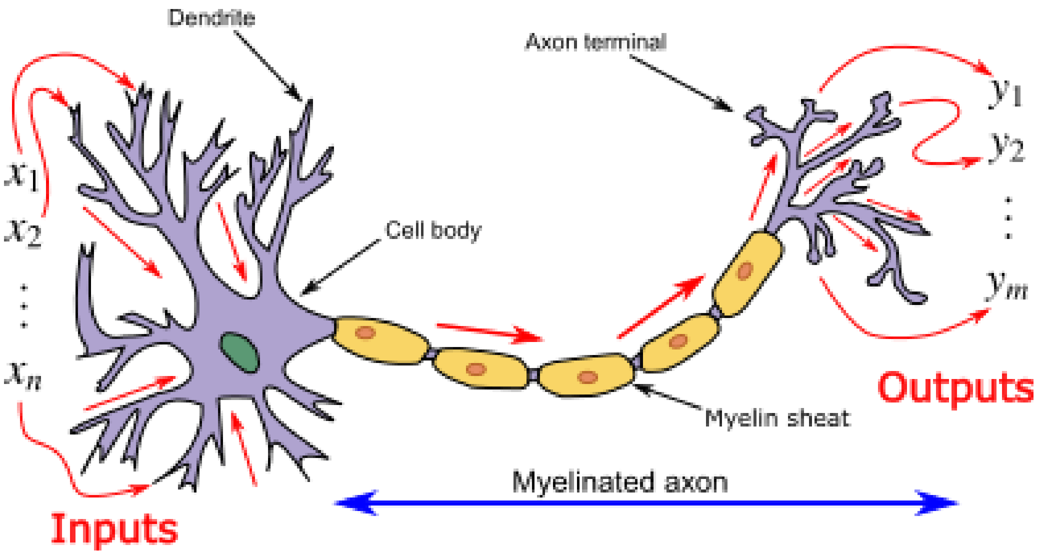

Figure 1. Biological Neuron (Chauhan, 2019)

As shown in Figure 1, the relation between the biological neuron and the ANN concept is represented as follows (Chauhan, 2019):

- Cell body (Soma): The neuronal cell body contains the nucleus and transforms the biochemical processes involved in the neuron's life.
- Dendrites: Every neuron has hair-like extensions in a tree shape around the cell. Their function is to receive the signals.
- Axon: it is a thin tube and acts as a transfer line.
- Synapse: In a complex spatial system, neurons are related to each other. When an axon reaches its final destination, the terminal arborization branches again. At the end of the axon, the structure called synapses is exceptionally complex and specialized. In these synapses, the relation between two neurons exists.
- Dendrites are accessed by other neuron synapses. The Soma processes the incoming signals over time and transforms the value into an output transmitted through the axon and synapses to other neurons.

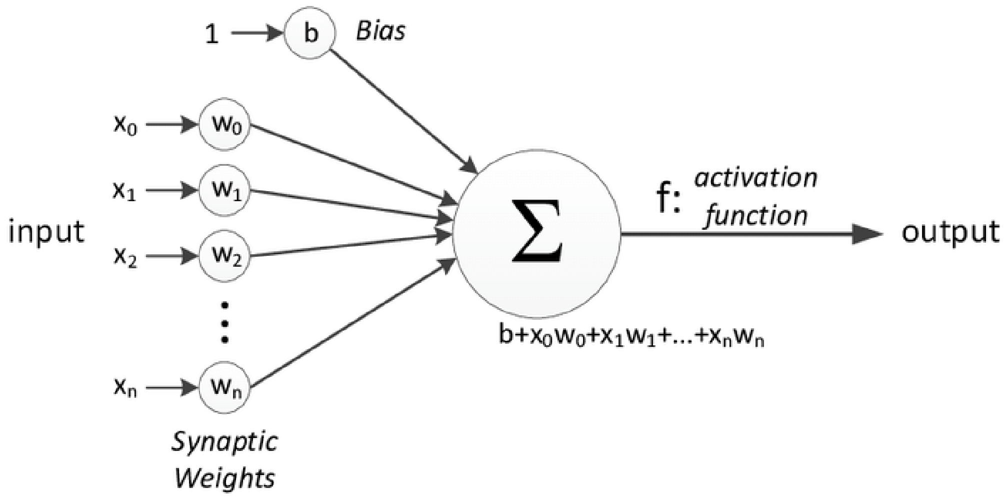

Figure 2. ANN Concept (Chauhan, 2019)

The x_0, x_1, x_2, x_3, x_n describes a number of network inputs for a single observation as shown in Figure 2. The weight or synapse of each input is multiplied. The weight is shown as w0, w1, w2, w3, and w_n. The power of a particular node is seen in weight (Chauhan, 2019).

The ANN model is represented in Figure 3 as an example to explain how it operates. If the model objective is object recognition, the first hidden layer might be for the brightness analysation of the pixels. The second hidden layer might be for the identification of the edges in the image. By that time, complex detectors should be created by the model to identify specific objects or images. Backpropagation is used to correct the mistakes and improve the model to operate without human interference (Dormehl, 2019).

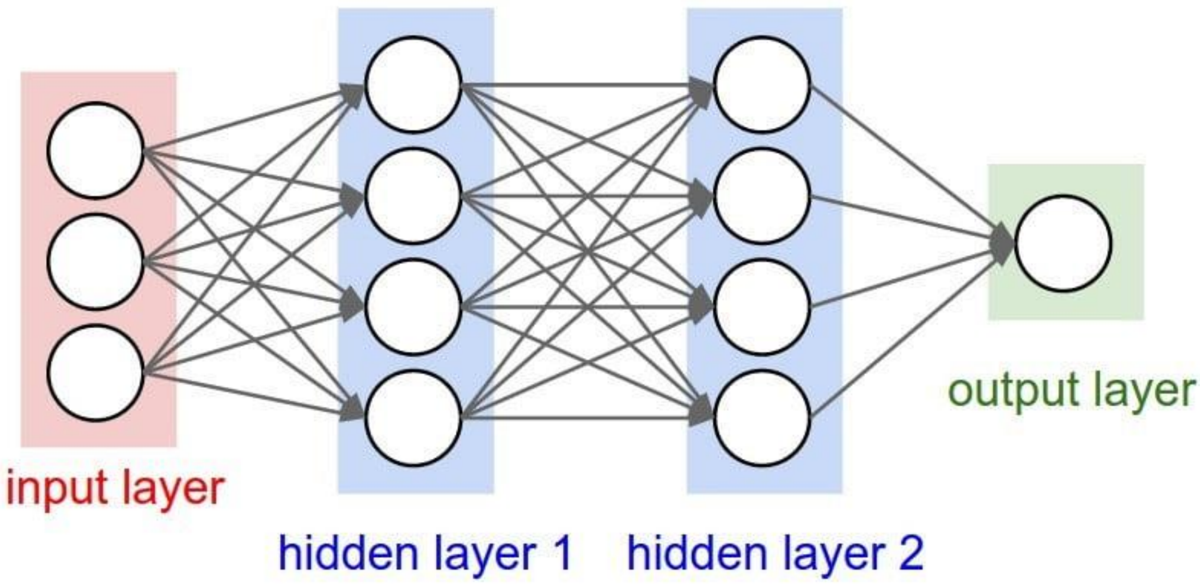

Figure 3. ANN sample (Dormehl, 2019)

Non-linear statistical analysis modeling techniques are known to model complex relations between the inputs and the outputs, and that is the function of ANN models. ANNs are profound models of learning that identify patterns and machine learning. They represent a part of artificial intelligence technology (Techopedia, 2019).

In cases of complex quantitative and qualitative reasoning, ANN is known as being more efficient than conventional statistical and mathematical methods. The ANN models can manage noisy data and to achieve high precision and accurate forecasting. They were successfully tested in many areas in construction management, such as classification, estimation, optimization, decision making, risk analysis, and selection (Waziri, et al., 2017).

(Kulkarni, et al., 2017) discussed and analysed the recent applications of ANN in different areas related to the construction industry productivity, cost, risk analysis, dispute, duration, and unit rate. The results confirm the usability of the ANNs in classification, prediction, modeling, and optimization of a task in the construction field. ANNs are based upon input data and the type of output data required to train the model, and this data set can always be updated by adding new examples to improve the results and minimize the errors. Therefore, ANN has significant advantages as a vital tool in solving complex problems or challenges in the construction industry. However, the experiments on many training algorithms, hybrid methods, and network architectures are still wide-ranging and may lead to higher model results. The ANN could be widely sued in the construction field, but a guideline should be provided for the learning algorithms, network architecture, and input selection.

2.6. Previous work

(Choon & Ali, 2008) explained the type of information that should be known and managed by the cost estimators or quantity surveyors to drive their companies to success. However, in terms of quantity surveying and contracting companies, the cost estimators' job description varies depending on the business strategy, organizations, and functional areas for which cost estimates are applicable. Meanwhile, developments have changed the nature of competitiveness and the methods used by cost estimators and quantity surveyors to achieve success in their operations, including business environment, social changes, client, information technology, globalization, political, organization management, internet, and construction technologies. The conclusion they reached was the cost estimated price could differ, and that depends on the category of the contractor or size of the company. The estimated price is influenced by factors related to a specific project regarding the different literature. This estimation of costs is a rather subjective topic that does not have standard methods of measurement but rather a method of steps and a certain basis for estimating construction costs.

(Luu & Kim, 2009) used the ANN in calculating the total construction costs for projects in Vietnam. The input variable was defined by sending ninety-one questionnaires, but only fourteen were received back. The collected data were analysed and prepared for the training and testing of neural networks. The ANN was designed using MATLAB software. To apply the neural network to practical projects, the software was developed using Visual C++. The findings indicate that the total construction costs for apartment projects in Vietnam can be predicted and that neural networks are used efficiently for cost models.

(Yitmen & Soujeri, 2010) developed an ANN model for the change orders management in the construction project through different phases without the interruption of the operation works. These interruptions usually result from disputes between project parties using MATLAB 7.5 software. The dataset of the project consists of eleven variables as inputs, which are changes dues to force majeure, schedule, changes in the aesthetic, permits, safety, process, Environmental, cost, productivity, purchase order, and modified scope of work. The data collection of the projects was done based on these factors; thirty-five change orders were collected from twenty different projects in North Cyprus. The ANN model was trained by a feed-forward backpropagation algorithm. The model was tested and proved it is the ability to predict the probability of dispute based on these factors with an acceptable margin of an error value.

(Li & Shi, 2010) incorporated the use of ANN, fuzzy logic, and particle swarm optimization in the quality evaluation of building projects. The research used fuzzy logic to describe the elements of the evaluation matrix and developed a building quality assessment model. In the evaluation and efficiency estimation of building projects in China, particle swarm optimization has been adapted to train the model. Simulated results of quality evaluation of the building projects, comparing ANN based on backpropagation and genetic algorithm, show that the training with particle swarm optimization provided greater accuracy in the sums of squares error and faster in the number of iterations and the time for simulation than for backpropagation and genetic algorithm.

(Zheng, et al., 2010) proposed an ANN model using a backpropagation algorithm as an example to estimate the construction costs in highway projects, collecting data from road projects in specific regions built during 2000 and 2002. The data was divided into seventeen standard projects for training and sixteenth for checking and testing. The model was divided into three sections. The first section is the input data, which consisted of nine factors that affect the construction of highway projects, which are highway grade, foundation treatment type, width, height, cross-section type, half-filling, cutting, half digging, and embankment. The collected projects were categorized based on these factors, and this data was transformed from qualitative to quantitative ones to be used as input data for the model. The model showed a relative error of the actual value with the estimated value is less than 5%. The test results indicate that the error is small. It is possible practically to satisfy the requirement for cost estimation and predict highway projects' construction costs. It shows that the model has a strong potential for generalization and accuracy. It is also recommended that it is entirely feasible to follow the ANN technology in the highway projects to estimate the accurate outcome. The ANN has reference value and academic importance, which are very important for adopting a new scientific method.

(Attal, 2010) has tried to create a reliable model of the initial design cost of highway construction and the length of the project based on statistical analysis. The statistical methods employed to represent ANN and the step-by-step analysis of the regression are used to define important parameters and to predict the early design phase and period of highway construction work. The Virginia Department of Transportation collected and stored the input data used to create mathematical models. The data used in this modeling were derived from two sources: project Cost Estimating System and Warehouse Management Information Portal. This department maintained parametric stage data. Furthermore, two different software techniques were used to describe the useful parameters used in these models; the trial and elimination process of ANNs; and sensitivity analysis. Also, significant parameters used in these models were identified. Two distinct statistical methods have been used to evaluate the selected parameters: non-linear ANN and linear regression analysis.

(Arafa & Alqedra, 2011) built an adequate model using ANN to assess the estimated cost of the construction project at an early stage. This model database consists of seventy-one construction projects from the Gaza strip. Several essential parameters were calculated for project structural skeleton costs and however, data and drawings are available at the pre-design phase of the project. There were seven parameters for the input layer in the ANN model: area of the ground floor, area of typical, number of floors, footing's type, columns, number of rooms, number of elevators. The ANN model was built with a hidden layer of seven neurons. The output layer of the ANN model was one neuron that represented the initial cost estimated of the project. The results of the trained models predicted the early cost estimate of buildings using the necessary project details and without the need for additional detailed information or design. They conducted a sensitivity analysis to show the main parameters which affect the cost of the buildings; these parameters are the floor area, number of floors, foundation's type, and number of elevators.

(Aibinu, et al., 2011) developed and trained a three-layer ANN feed-for-ward model that includes a single output node. The database consists of 100 projects completed and cost details. The model's input variables were nine variables, which are Project location, size, type, total project cost, structural material, method of cost estimation used. The criteria of the developed model are 0.2% of MSE, 3% of absolute error, and 73% of the correlation coefficient. The estimate of output did not vary by more than

8.2% from the actual estimated in over 73% of the test cases. This means that the model succeeded in estimating the actual cost for the projects using only nine parameters as input and did not need detailed design or information, which saves effort and time. Cost estimators could use the qualified ANN model to determine the construction costs at the pretender phase. In order to rapidly estimate the error when estimating the new project, the model can be adjusted for new projects. The percentage of error should be added to the output value as a contingency to avoid any risk. The model may also be expanded to estimate a project's actual costs.

(Muqeem, et al., 2011) developed an ANN model to estimate the production rates of formwork of beams by the research objectives were achieved by evaluating output rates of beam formwork by evaluating seven sites for building projects. Also, the study listed the factors affecting the production rate are project location, weather, site conditions, number of workers, and material availability. Through data analysis to determine the impact of each factor on the output result, the availability of materials showed the highest correlation factor, followed by the number of workers. Finally, the ANN model has successfully estimated accurate values of output rates with the integration of the influencing factors. By measuring the MSE of the expected production rates, the output of the model was calculated. The MSE values of 1.82 were obtained. These findings indicate that the ANN model successfully predicted production rate values for bean formwork with the least number of errors.

(Tatari & Kucukvar, 2011) discussed how the construction environment has a significant economic, social, and environmental impact. The environmental impact study of buildings gained considerable significance in the building industry and the growing environmental issues of the construction consequences. An ANN model for LEED-certified buildings was developed to estimate cost premiums depending on LEED categories. Regression analyses were used to check the feasibility of the model. After the validation and testing of the ability selected of the selected ANN model to predict the cost premium of certified LEED green buildings, sensitivity analyses were conducted to understand the relationships and measure the impact of each input variable on the output. The factors with the highest sensitivity in cost premium were identified for Energy and sustainable sites LEED categories. In this research, we aimed to discover the significant ties between LEED and cost premium categories and provide a model to direct owners in estimating the premium of costs based on LEED categories.

(Chou, 2012) created different models using different AI tools to forecast the dispute handling methods specified in the private-public-partnership type. In the research, machine learning techniques as ANN, tree augmented name, support vector machines, Bayesian inference, regression and classification techniques, efficient tree, Quick, unbiased, exhaustive chi-square, C5.0, and interactive detection, all these methods were used to improve the output of the model. All these models were integrated together to compare the output from different combinations. These techniques demonstrated the highest accuracy for classifications of 84.65% in predicting the outcome of conflict resolution, which could be agreements mediation, awards, arbitration, appeals or no conflict, conflicts. The study indicates an efficient procedure for disputes handling.

(Dagbui & Smith, 2012) demonstrated that the ANN information acquisition, general understanding, and forecasting capabilities to build up to final cost estimate models. The database was collected for ninety-eight construction projects built in Scotland between 2007 and 2011. The model consists of ten variables as input and a single variable as output, which is the final cost of the project. Several hidden layers were used in the model design, and the selected training algorithm was Quasi-Newton.

(Vahdani, et al., 2012) developed a computer model called support-vector machines machine was introduced in order to enhance the computational accuracy of cost estimates during the early project life cycle period. A cross-validation technique was used to train the model, and the output was compared with non-linear regression, with back-propagation ANN results.

(Elhassan, et al., 2012) addressed the use of tools to ease decision-making in complicated situations in the construction industry. The study defined methods for optimizing the decision-making process in construction management. The output of the study that there is a lack of integration of ANN specifically or AI generally in the construction industry; however, using these tools will push the industry forward.

(AL-Zwainy, et al., 2012) created an ANN model to estimate the productivity of finishing works for marble floors in the construction industry. The input variables consisted of ten factors, which are construction materials availability, number of labours, age of labours, labour's experience, health conditions of the labour, site security, site conditions, weather conditions, floor area, and size of the tiles. The data set was collected based on these factors from different projects commercial, residential, and educational from projects in Iraq. The ANN model was trained using the backpropagation algorithm and was tested. The model showed accuracy in predicting the productivity of finishing works for marble floors with a rate of 90.9% and a correlation coefficient of 87.55%.

(Pengcheng & Kun, 2012) developed an ANN model using MATLAB to evaluate the risk in the construction industry for the main project participants. In training the model, the backpropagation algorithm was utilized to eliminate subjectivity considerations. The risk factors found through a field survey were evaluated on a Likert scale of 1-5. The results of the network simulation show the model to be appropriate and realistic.

(Polat, 2012) created an ANN model to estimate contingency costs to allow the project managers to consistently evaluate the risk levels of their projects so that they can measure the cost contingency amount more accurately and reliably. The data set collected for training and testing consisted of 195 projects executed by 85 big contractors in Turkey. The results of a statistical study have shown that the model is accurate. It was able to identify the nonlinear relationship between the cost contingency amount, which is included in the bid value, and the risk factors.

(Petroutsatou, et al., 2012) discussed the Underground complexities and risks are involved with the construction of a tunnel. Also, the final construction costs are hard to estimate, particularly during the design phase, when concerns are identified and major design decisions are taken. A system that helps to estimate the initial costs of road tunnels would therefore be of great benefit as it would allow alternative and more feasible solutions. First, the basic parameters that have an impact on the construction costs permanently and temporarily are calculated, which are geometrical, work quantity, and geological. The data collected were used to build the model consisted of projects carried out between 1998 and 2004 of thirty-three Twin Tunnels of 46 kilometers of length built in Greece. The model was subsequently developed using two types of ANN, which are multi-layer feed-forward network and regression neural network. Finally, the output of the two models was compared by calculating the difference between the actual value and the estimated value. In conclusion, the developed models are fit and showed accuracy in estimating the initial construction costs of road tunnels.

(Wang, et al., 2012) developed a model to estimate project success due to early planning. This study used the Project Definition Rating Index as a survey tool. It used data collected from over ninety-two construction projects to develop a model to estimate the construction project costs. ANN and support-vector machines were used to develop and evaluated several models using the collected data as a database. The model results showed that the early scheduling status could be implemented effectively for surveyed sample projects to predict project results using ANN. The support-vector machines machine model offers the best prediction outcomes accuracy of 92% for the prediction of project costs. Simultaneously, the best prediction outcome is the adaptive boosting of ANN models, 80% average accuracy in the success of the project schedule. The studies prove that a successful complete project could be achieved through efficient early planning also proved that AI model techniques are applicable to nonlinear data comparing to traditional techniques.

(Minli & Shanshan, 2012) explored the decision-making problems for a tender offer from a system perspective, focusing on the study of specific past theories and models. Bids are a dynamic decision-making mechanism with several variables, full of uncertainties. For that matter, it is difficult and even impossible to develop accurate mathematical models. Otherwise, fourteen factors were identified that affect the tender offer, which are market condition, location, construction condition, project type, profit status, the following projects, project scale, owner, duration, risks, and complexity. ANN model was developed to solve this issue using MATLAB. The model was trained, and the best model was selected based on the lowest MSE value. The selected model's characteristics are three perception layers the first layer consisted of 14 neurons, the second layer 29 neurons, and the third layer one neuron. This model's output error is 3.816%, which represents the difference between the actual value and the estimated value by the model and is an acceptable percentage. In the future, some new models will become more reliable with the rapid development of ANN, such as noisy evaluation, expert system methods, and genetic algorithms. This will address the limitations of ANN.

(Alqahtani & Whyte, 2013) used ANN to create a new life cycle cost analysis system for building projects. The model used the cost of significant items to calculate the life cycle costs. The model was built using a set of 20 building projects using MATLAB and Excel solver. The results For Excel solver and MATLAB, the accuracy levels were 2% and 1%, respectively.

(Kaushik, et al., 2013) used a backpropagation algorithm in developing the ANN model to predict the cost. Also, developed another model was built to fit and enhance the efficiency using Constructive Cost Model software. It tackles imprecise, irrelevant input and increases the accuracy of software results of cost estimation. The model is checked with three datasets available to the public. The Constructive Cost Model software output was compared to the output of the ANN model.

(Cunningham, 2013) showed that construction costs are significant for most building customers and outlined the key factors influencing construction costs in the Irish market. The study found that the customer's expectations for quality, cost, and time limitations are critical factors to the efficacy of the brief. The design team is considered the key player in this process for determining the nature of the project, and it is development; therefore, they affect the costs of the project. Factors of design that influence the cost of buildings are socioeconomic factors, function, specification, geometry, and legislative constraints. He also mentioned that there are other factors that affect the cost estimate as to the project's location and environmental site conditions. The study also analysed the effect of other factors such as market conditions and the procurement and concluded with a review of the factors that influence the costs of the contractor production site.

(Goh & Chua, 2013) developed an ANN model with accident data from the Singapore building industry in conjunction with the occupational health and safety management system framework audit. The research examined how the ANN model can be used to identify the relationship between safety and occupational health and safety management and to identify the main factors that have the highest impact on the output, which are emergency preparedness, incident analysis, and meetings in a group. This analysis was critical to know how to reduce the accidents in the construction projects in Singapore and minimize the severity of these accidents if occurred, and this happened through having mitigation strategies, learning from previous incidents, and having open communication. The research showed how the ANN approach could provide valuable insights to enhance safety in the construction sites, and it is the ability to analyse the data.

(Kim, et al., 2013) discussed the different regression analysis tools, ANN, and support vector machines used to estimate the total construction cost for school projects. The model design and validation were carried out with 197 cases, while the remaining twenty cases were used to test the model. All three models have established a strong relationship between the predicted costs and actual costs. The three techniques performed well for the application; however, the ANN models provided more precise estimates than the others. The ANN has proved useful for addressing complex issues and developing user-friendly prediction models. They can identify patterns found in the data and provide more opportunities to explore various project management options and techniques. In this comparison, the ANN model is more fitting than the support-vector machines machine model to estimate school building projects.

(Maghrebi, et al., 2014) developed an ANN model focusing on the parameters of the supply chain to estimate the operation time of concrete. The data set used to build the model was collected from real-life projects in the Sydney Metropolitan Area with two hundred trucks and seventeen depots. The model was tested, and the accuracy of the results compared to other studies' results, which only took into account construction parameters for predicting concrete productivity.

(Lhee, et al., 2014) proposed an optimum contingency estimation approach for transport construction projects using the two-stage ANN. The model objective is to provide the best solutions to reduce the risks and optimize the budget decisions

(Hong, et al., 2014) have provided a cost assessment model and use of the construction engineering method. This model was based on the integration of both approaches particle swarm optimization and backpropagation ANN. The initial weights of ANN were optimized by particle swarm optimization. The hybrid approach's key objective is to enhance the convergence rate of ANN and the ability to achieve optimal global performance. This approach showed accurate results and can be used to carry out a quantitative cost assessment.

(Nagappan, et al., 2014) discussed the increase in the prediction accuracy of ANN when trained using swarm intelligence algorithms. For the assessment of the different ANN-swarm intelligence combinations, several models were developed, such as artificial bee colony, particle swarm optimization, firefly, and ant colony optimization have been found in the study's swarm intelligence algorithms. These models were compared with conventional ANN models for their conversion speed and prediction accuracy.

(Jain & Pathak, 2014) discussed the ANN applications in the construction field by reviewing several articles published in different journals. The research focused on the construction management area. This review's output proves the efficiency of ANN when applied to many areas in the construction field, such as risk analysis, optimization of the resources, prediction, and estimation, selection, and classification. Based on the case studies' findings, ANN function better than traditional methods Many challenges are faced in civil engineering circumstances that are very complicated and not well understood. The mathematical models struggle to describe the complicated behaviour of these issues. In comparison, ANNs are based exclusively on the input-output data, which allows training the model. Additionally, ANNs can still be revised to get better results by providing new training examples as new data becomes available. Therefore, ANN has several significant advantages, making it a practical and functional method for solving problems in the construction engineering field and is expected to be more applicable in the future.

(Lazarevska, et al., 2014) developed an ANN model to estimate the fire resistance of concrete reinforced columns. The database used to build the model consisted of eighty samples, which were used to train the model, and twenty-seven samples were not included in the training and were used for testing the selected model. The main aim of this study was to explain how easy and beneficial the use of ANN is in order to solve engineering issues. The output after the comparison between traditional numerical methods and ANN shows that ANN offers an effective tool for developing a model that can then be used to assess fire resistance of concrete reinforced columns, mainly when numerical results are not available.

(El-Sawalhi & Shehatto, 2014) used Neuro Solution software to create a model for determining building projects cost with a high degree of precision and without the need for detailed details or drawing by using the Artificial Neural Network (ANN) to develop a model that can enable contractors or owners to obtain total costs of the project at a very early stage with few information available without the need of detailed design. ANN is a modern cost estimation method that can benefit from experience and observations to solve non-linear problems. It could manage tasks that have missing information, fuzzy, complicated problems, or incomplete data sets. Qualitative and quantitative approaches have been used to define essential factors that affect the project costs in the structural and architectural stage to create this model. A database of 169 construction projects has been gathered from the Gaza Strip building industry. The architecture of the model was eleven main factors as input variables which affect the project cost. Showed that the neural network managed adequately to estimate the cost of construction projects without detailed information and drawings with an acceptable error of 6%. The study also included sensitivity analysis to show or measure the effect of every input on the output and prioritize them. The number of floors and areas were the highest influential factors on the project cost.

(El-Sawah & Moselhi, 2014) have provided a study in preliminary cost estimates using ANNs. The selection and configuration of the ANN model had a high impact on the output of the model and the accuracy of the preliminary cost estimate. The model used backpropagation in training the network. Models for the cost estimation of low-level steel building and timber bridges have been created. A database of seventy buildings was collected and used to train the model to estimate the actual cost for these projects using the regression model. The analysis was carried out using the built regression and ANN models on current data for seventy low-rise structural steel buildings, and their respective cost was estimated. The same algorithm and design of models were used to estimate the costs of timber bridge projects.

(Lyne & Maximino, 2014) created an ANN model using MATLAB to estimate the construction project's total cost in the Philippines. The model used a feed-forward backpropagation algorithm to train the network and select the model. The data set consisted of 30 projects, and these projects were divided as follows: 60% training, 20% validation, and 20% testing. The model included six input variables, which are reinforcing steel weight, concrete volume, number of floors, formwork area, and number of basements, which were considered the main factors or parameters affecting the construction cost in the Philippines. The selected model architecture was six variables as input, one hidden layer with seven neurons, and a single output node, which is the project cost. The model succeeded in estimating the structural cost of the buildings after training and testing.

(Liu & Guo, 2014) developed a project risk assessment model using ANN and rough sets. The ANN model was developed using MATLAB 7.0. The dataset consists of 15 residential projects in Ganzhou. These data were categorized into four categories of risk factors, which are human factors, material factors, method factors, and environmental factors. The model was tested and showed high accuracy results with 73.56%.

(Patil & Bhangale, 2014) addressed the concept of OH costs, which represented in the percentages of OH costs, understanding of OH costs, the reasons that lead to increase OH costs, adjustments in OH costs, the estimated OH costs, OH cost controls. The study also discussed that it is difficult for contractor companies to settle on the optimal amount of OH prices in an unstable market. This optimal amount is what the contractors need to win over and manage massive contracts while at the same time not exhausting the company financially. The parameters that affect the OH cost include office OH costs, financing, labour costs, and equipment costs. Also, the research discussed the reason behind OH costs increase, which could be inflation, late payments, public regulations, and the shortage of new projects.

(Janani, et al., 2015) did professional interviews and project data collection, which contribute to understanding when bidding and efficiently manage financial resources. They said that the project initiated must be completed without delay, delay of materials delivered to the construction site must be avoided or controlled, salary delays for the employees, Project supervision, cost monitor and control system must be used and regular checking on reports in order to manage the OH costs of the project. The research also discussed the percentage of OH of the contract, the effect of OH on the income, the factors that affect the OH costs, identifying the factor with the highest impact, control of OH costs, contractors and engineers awareness of OH costs.

(Oduyemi, et al., 2015) developed a model using ANN in order to estimate the operational and maintenance costs of existing projects. The database used was collected from the office in Penllergaer's business park. After training and testing the ANN model, the output, which is the total cost of operation and maintenance of the buildings, was accurate, which shows how the ANN could be integrated and improve the life cycle estimation process.

(Patel & Jha, 2015) developed a model to predict safety climate in a construction project based on an ANN. The study listed ten factors that affect the safety climate for construction projects, which are communication, supportive environment, worker involvement, appraisal of physical work, work pressure, commitment, safety rules, supervisor environment, risk appreciation, and competence. These factors were used as inputs, and the safety climate of the project is used as an output for the ANN algorithm. This study collected a total of 250 responses through a questionnaire survey across the country. A three-layer feed-forward backpropagation neural network (10-18-1) has been utilized for the analysis. The sensitivity has been calculated for each safety climate construct (input) based on its mean value and the score of the safety climate of the projects. The developed model predicts the safety climate of a construction project reasonably well. Based on sensitivity analysis, commitment, and supervisory environment are proposed as the most significant out of 10 constructs of safety climate. Moreover, projects are ranked to their safety climate using this model, and outlier projects could easily be identified through the standard probability technique.

(Heravi & Eslamdoost, 2015) developed an ANN model in Iran to estimate the productivity of labours based on feed-forward multi-layer neural networks trained with back-propagation algorithm that deals with a complicated nonlinear relationship between the factors and the productivity rate of the labours. Bayesian regularization and early stoppage were carried out and contrasted with avoiding overfitting the model and enhancing generalization. The findings show that Bayesian regularization works better at early stops. The models produced are implemented in two power plant projects to display the predictable performance of this model. A sensitivity analysis was carried out to identify the impact of each factor, which are the input variables on the output, which is the productivity rate of labours. The output was to focus on the installation of concrete foundations in Iran. This study contributes to the knowledge of construction management by identifying the influential factors that affect the labour productivity and developing an ANN model predicting work productivity.

(Mwiya, et al., 2015) used the ANN approach to evaluate the critical elements that influence the cost estimate, especially in unit rate. A questionnaire was done to collect information about the factors affecting the cost estimate in road construction projects. The output was twenty-five factors that were listed and were reduced to eight factors using the SPSS software. ANN was used to evaluate the eight variables, which are contractor capacity, project feasibility, project location, corruption perception, political issues, profit, and OH financial delays and the availability of the material. The evaluation showed that the political issues and contractor capacity have the highest impact on the unit rate cost estimation. In contrast, project feasibility and financial delays have the lowest impact. The research proved that using the ANN approach will increase the cost estimates accuracy for infrastructure projects and create a benchmark that helps to match new workers' job and training practices, enabling them to develop other best practices for cost estimation.

(Golizadeh, et al., 2016) developed an ANN model to predict the duration of major structural activities of concrete frame projects. The study focused on four activities, which are beam reinforcement, bean concreting, column reinforcement installation, and column concrete. Four ANN models were developed for these four activities. These models were integrated together to develop a more accurate tool for estimating the duration of these activities.

(Aswed, 2016) developed an ANN model to estimate the productivity rate of the labours using thirty factors as input variables. These factors are site security, mortar type, site conditions, the thickness of the wall, height of the wall, length of the wall, availability of materials, salaries, labour health, labour, age, labour experience, and labour number. The data set was collected based on these factors from different project types. The model was trained and tested with an accuracy of 86.28%, which proves that the model can estimate the productivity rate of the labours in different types of projects such as residential, educational, and commercial through using these 30 factors as input for the model.

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

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

(Jiang & Shi, 2016) examined the various positions of OH costs and entry costs in the production firms. It also explores the repercussions of the overall economy's capital distribution. They demonstrated that reducing entry costs would increase the cumulative firm efficiency by facilitating additional entries using an analytically tractable model, thus lowering OH costs by decreasing the selection level adversely and the possibility of increasing output resources allocation. The two costs were estimated in a six-digit United States manufacturing data model, and their future impacts are quantified. Their results indicate that different policies to lower costs for creating new companies are more relevant than policies to reduce operating costs simply as it indicates a 1% decrease in entry costs affect the output by an increase of 0.27%. In comparison, the same amount in OH cost reduces production by just 0.048%.

(Yadav, et al., 2016) created cost estimation ANN model that can estimate the structural construction cost of residential projects by using different parameters as input variables. The data set used in developing the model was collected from the Schedule rate of the book and included projects for the past twenty-three years. The researcher used Nero XL (2.1) to create the ANN model. After training and testing the model, the architecture of the selected model was eight input variables, which are costs of sand, aggregates, cement, steel, masonry, non-skilled workers, skilled workers, and contractor per feet square. The resulting ANN model accurately estimated the total structural construction costs with R^2 of 0.9905.

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

(Mirahadi & Zayed, 2016) proposed to improve the accuracy of estimating the productivity for building operations with a hybrid model. The proposed paradigm model impacts the output of the productivity rate of both the and quantitative, qualitative, and variables and the model's complex structure based on inherent data characteristics. A new model of the ANN driven fuzzy framework was created for this purpose using Fuzzy reasoning. The dataset of 131 data items was divided into 117 samples for training the model fourteen samples Validation and testing. With respect to MSE, the proposed ANN model showed 83% of accuracy. This study allows researchers to build estimating models using the data features.

(Glinskiy, et al., 2016) presented a solution by ANN to determine the safety of territorial environmental entities. ANN technology enables the lack of input knowledge to be resolved and a fair environmental protection assessment for territorial entities to be performed. Also, the flexibility and the ability to deal directly with new data, tolerance, and approximation are the benefits of using ANN. This technique allowed the following tasks to be solved: to carry out a typology of Russia's subjects in terms of environmental protection, to develop scenarios to increase their protection, to carry out a forecast of modifications in their development, and to take environmental factors into account. These factors were identified in the study which are production waste, waste usage, investments in environmental protection, costs of land rehabilitation, cost of natural areas protection, costs of prevention of climate change, economic production index, industrial production index, pollution emissions, air pollutants, freshwater usage, amount of used recycled water, wastewater discharge, cost reduction of materials, carbon dioxide emissions, raw material usage, pollution reduction, waste recycling, the share of environmental protection, amount of progress, amount of air shipping and truck shipping. The selected model had all these twenty-six factors as input variables, and one hidden layer consists of one neuron. The output is to determine if this area is considered environmentally safe or not for construction works.

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

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

(Kulkarni, et al., 2017) discussed how the construction industry has several uncertainties, which are mainly related to safety, cost, time, and quality. Such uncertainties make the whole process of building too complicated. It thus falls within the framework of ANN, in which the hazy knowledge can be efficiently interpreted to draw logical conclusions. The ANN application in construction management was discussed in a different area as risk management, tender bode, safety, labour productivity, cost estimation, and equipment productivity. The results indicate that the ANN was extremely helpful in interpreting an insufficient input of information correctly. It was shown that most researchers used the ANN form of feed-back propagation; however, if a single ANN was found to be inadequate, therefore using a hybrid modeling in combination with other machine learning techniques such as support vector machine or genetic programming was very useful and enhanced the accuracy of the model.

(Nani, et al., 2017) created a model for predicting the duration of prefabricated steel Bridge construction in Ghana using ANN. Data were provided by the Department of Feeder Roads for eighteen bridge construction projects. The collected data had all the information that contained the quantity and duration of all work items per project. Another questionnaire was used to collect information concerning the pre-manufactured steel bridge components of each project chosen. The factors were first reduced to a lesser number, which is the formwork and the span of the bridge. These two factors were used as the input variables for the model to predict the output, which is the duration of the project. The findings showed that the duration required to construct a bridge is closely linked to the formwork and the bridge span. The model developed provided precise results with R2 = 0.998 and accuracy of 95.95% on average during validation, which proves the efficiency of the ANN model in predicting the duration of bridge construction projects.

(Elevado, et al., 2018) discussed using flying ash as a substitute for Portland cement type 1 and replacing gravel as ground aggregates with ceramic waste tiles. Also, specimens were subjected to different curing days to test their improvement of strength. ANN was considered because a large range of data was available. The goal of this study was to provide an ANN model that would be able to forecast the concrete's compressive strength while using waste ceramic tiles as a replacement for coarse aggregates while varying the quantity of fly ash as a partial cement replacement. The model architecture consisted of three input variables, which represent the days of compressive strength 7 days, 28 days, and 56 days, and also two output variables, which are the percentage of ceramic and the percentage of fly ash. The ANN model was validated and tested to ensure accuracy.

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

(Raghed, et al., 2018) developed a model based on ANN that supports construction companies to evaluate and predict the outcomes of their current projects. This is what improves the performance of the construction company and enhance it is the capability to compete in the local and international market. The proposed ANN model included public construction projects only. Twelve factors were identified from the previous literature review: project scope, external constraints, time urgency, interdependency between elements, project resources, project budget, project parties, resource availability, level of interface, project client, and several elements. After analysing these factors, time urgency and external constraints had the highest impact while project client and number of elements had the lowest impact. Data from 30 previous project projects developed in Egypt, the Gulf, and North Africa were collected. These collected projects included malls, hotels, airports, educational, offices, and airports. The ANN model was developed using IBM SPSSS software. Twenty-five projects were used for training the network using a backpropagation algorithm, and five projects were kept for testing the model. The model predicted four projects correctly, which is equal to 80% accuracy.

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

(Golnaraghi, et al., 2019) developed a model to estimate the productivity of formwork labours. ANN techniques using supervised learning algorithms have proven more effective than techniques of statistical regression considering factors such as ease of modeling and prediction precision. A variety of ANN techniques have been used to compare their respective effects, including radial base function neural network, Adaptive Neuro-Fuzzy inference System, general regression neural network, and backpropagation neural network. The findings show that the backpropagation neural network succeeds than other modeling strategies to estimate the productivity of construction labour. In addition, the model proved the ability to deal with a nonlinear dataset. In order to have labour productivity modeling, the behavior of the data sets must be analysed before it is fed with any AI techniques.

(Bhosale & Konnur, 2019) discussed how the construction managers utilize expert project management strategies and tools to control, manage, design, and execute the construction project. The aim of construction management is to control the cost, time, quality, and safety of the project, and hence the management of the building is concerned with many uncertainties. Therefore, the ANN approach is beneficial in this situation. It utilizes any type of information, even if it is missing, achieving the best possible solution based on the input. This paper discussed ANN application in several construction areas as risk management, safety, cost estimation, work environment, and productivity rate. This analysis showed the efficiency of ANN in solving complicated issues in the construction industry. Also, it was mentioned that most of the researchers used the feed-forward backpropagation algorithm in developing the ANN models. However, if the output of a single ANN model was not enough, then it could be integrated with hybrid modeling or other machine learning tools to improve the results. The user's skills and data accuracy have a significant impact on the efficiency of the model.

2.7. Summary

This chapter discussed the literature of cost estimation, direct costs, indirect costs, the types of indirect costs which are divided into office OH costs and site OH costs, factors that affect the site OH costs generally and specifically in the Egyptian market, the definition of ANN and the previous work done related to the integration of ANN technology in the construction industry. In the next chapter, the questionnaire will be prepared and send to construction companies for data collection of real-life projects. These project data will be the first step in developing the ANN model. This model objective is to enhance the bid accuracy in the Egyptian market, which will lead to:

- Awareness of the main factors affecting the site OH costs
- Minimize the effort and time required in the cost estimation process
- Providing an accurate estimation for the site OH cost percentage
- Improve the Egyptian construction industry through the integration of AI represented in ANN technology.

3. Data collection and analysis

3.1. Introduction

The initial part of the research included identifying and explaining the problem. According to that, the structure and the work frame of the master thesis were designed. In the construction industry in Egypt, there is a limitation in the available database, and appropriate cost estimation methods are hardly used. The methods and techniques used are mainly traditional old ones. This research objective is to integrate AI by using the ANN approach in the construction industry.

The second part of the research included the literature review, which discussed what other researchers wrote about this topic. The output was to identify the general factors affecting the site OH cost for construction projects generally and in Egypt, mainly. Several studies related to cost estimation, indirect costs, OH cost estimation, and implementation of ANN in the construction industry were discussed and shown in the literature review.

This chapter mainly discusses the methodology and technique used in this study. The key objective of this chapter is to collect data on site OH percentage cost of construction projects in Egypt.

A questionnaire was designed based on the factors listed in chapter two and sent to construction companies. These data were analysed to illustrate the impact of each factor on the percentage of site OH costs and determine the factors which have the highest impact, and the factors have the lowest impact.

3.2. Questionnaire design

This section includes data collection based on the output of chapter two, which identifies the factors affecting the site OH cost estimation in Egypt. These factors were used in designing the questionnaire.

The questionnaire design was wisely selected as the design is important to get exact results. The question’s design was specific, direct, and to select between identified given answers. As shown in Appendix A, the questionnaire form was divided into three parts:

- Personal information, which mainly included questions about the respondent's information as name, position inside the company, and contact details.
- Organization data, which mainly included questions about the company as the category of the company and another question to identify the experience of the company in construction projects.
- Project data, which mainly included questions related to the project according to the factors affecting the site OH cost estimation that was identified earlier in chapter two. These factors were listed by given answers to choose from starting with the project contract value, project duration, project type, type of contract, owner type, and percentage of site OH costs

3.3. Manual Data analysis

This section includes a comparative comparison with the help of forty completed construction projects between the different factors affecting the site OH costs and their impact on the percentage of site OH costs. These tasks were carried out. The collected data was from different places in Egypt. The comparison is based on the cost impact on the percentage of site OH for each factor separately, such that the relationship between each factor and the percentage for site OH costs is known and understood.

The adapted building technology collected was a typical standard reinforced concrete system in all projects as this technology accounts for more than 95% of construction projects in Egypt. therefore, if any other technique or technology used in the project construction is not included in this study and should be estimated separately by the engineers or cost estimators.

The forty collected projects represented the different classification according to the factors that affect the site OH cost such as different project types were included residential projects, hotels, office buildings, banks and educational, projects located inside the city and others outside the city, projects with a different budget starting from below thirty million Egyptian pounds to above one billion Egyptian bounds, projects executed by different categories of construction companies which reflects the first and second categories, project duration staring from less than eighteen months until projects more than forty-eight months, different types of clients were included private as well as public client and projects and different types of contracts. All these variables will be analysed in detail in the next part. The collected projects were collected from several construction firms using the questionnaire attached in Appendix A.

It should be mentioned that the percentage of site OH costs are calculated in this study by dividing the total site OH costs by the total project value. All the collected projects were listed in Appendix B, and no information regarding the name of the projects or the construction companies were mentioned to ensure confidentiality.

3.3.1. Impact of project type on-site overhead cost

The project type is a basic definition of the client's demands and requirements that need to be fulfilled by the project, and this can be represented through the followings:

- The project's specifications.
- Architecture designs.
- Structural drawings.

The collected forty projects were selected to cover most of the project types in Egypt, which are banks, Office buildings, educational buildings, residential buildings, and hotels. The collected data, as shown in Figure 4, consists of 20% banks, 20% administration buildings, 13.33% educational buildings, 20% residential buildings, and 26.67% hotels.

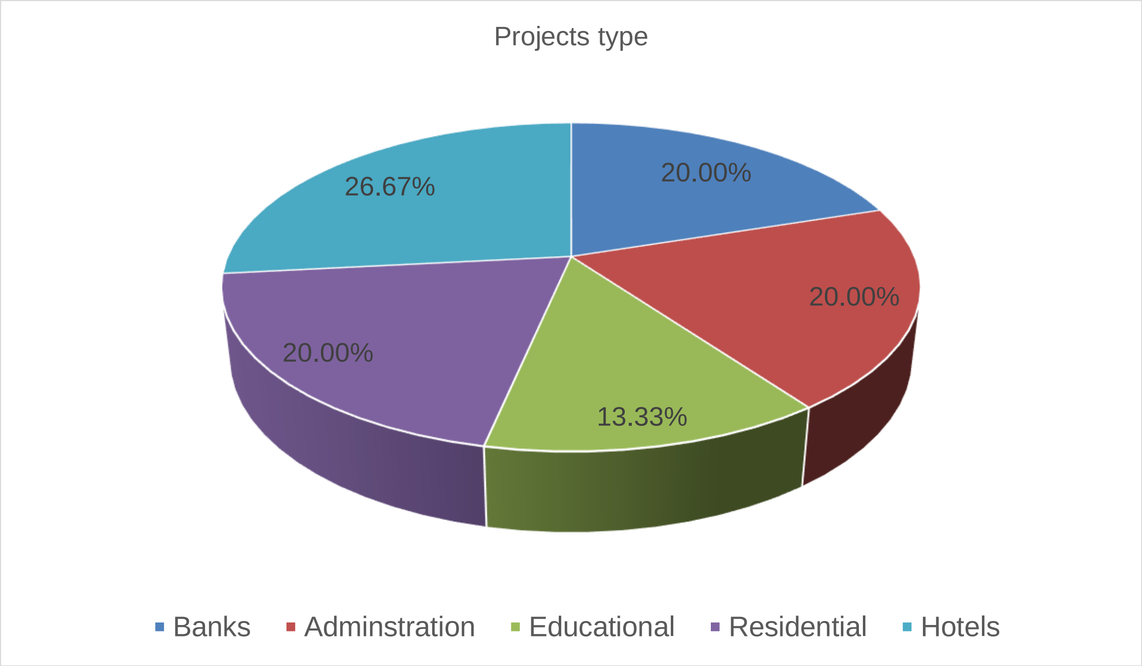

Figure 4. Classification of the projects according to projects type (Author)

The impact of project type on the percentage of site OH cost and the total project value is shown in Figure 5. Through the analysis of this data, it was noticed that the percentage of site OH costs changes with the change of project type. Each category consists of the minimum, average, and maximum percentage of site OH cost. The administration buildings showed the lowest average rate, with 9.5%, and the hotels showed the highest average percentage with 11%. The relationship between the project type and the site OH costs be a non-homogeneous relationship that cannot be represented through a formula or equation.

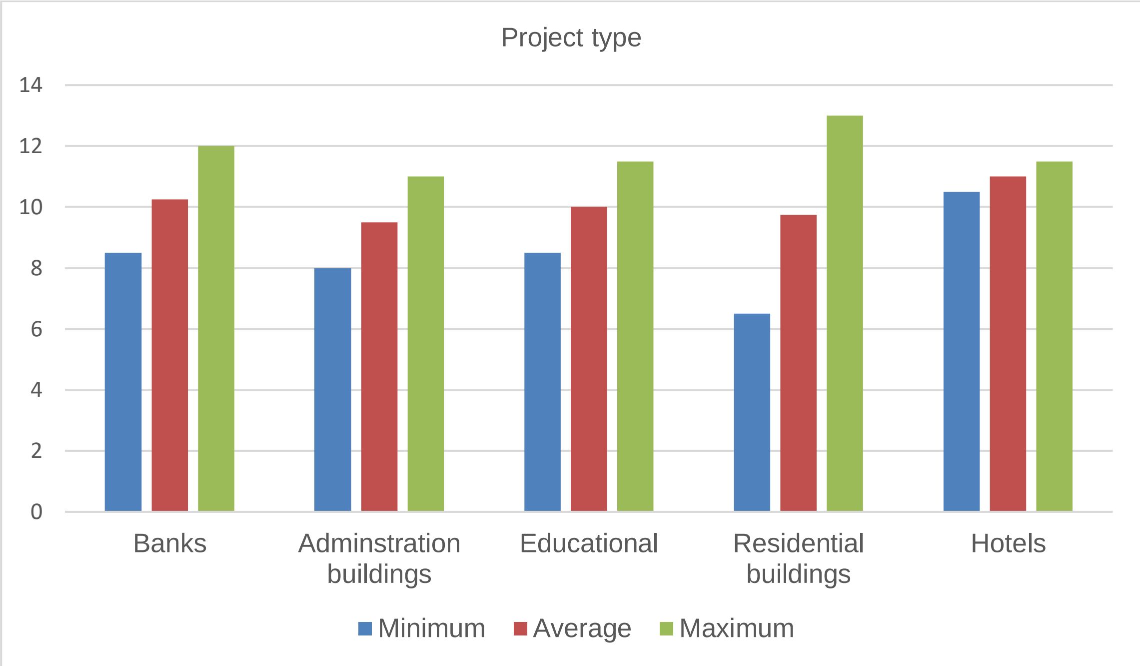

Figure 5. Impact of project type on-site overhead cost percentage (Author)

The output of the analysis showed that the relationship between the project type and the site OH costs be a non-homogeneous relationship that cannot be represented through a formula or equation.

3.3.2. Impact of project duration on-site overhead cost

The project duration is considered one of the critical factors for the construction companies in Egypt. The collected projects were categorized as per their actual duration into six different categories, as shown in Figure 6. The collected data consists of 11.11% of the projects that are less than 18 months, 20% are less than 24 months, 4.4% less than 30 months, 15.56% less than 36 months, 31.10% less than 48 months, and 17.78% more than 48 months. The number of projects is shown with a relative percentage.

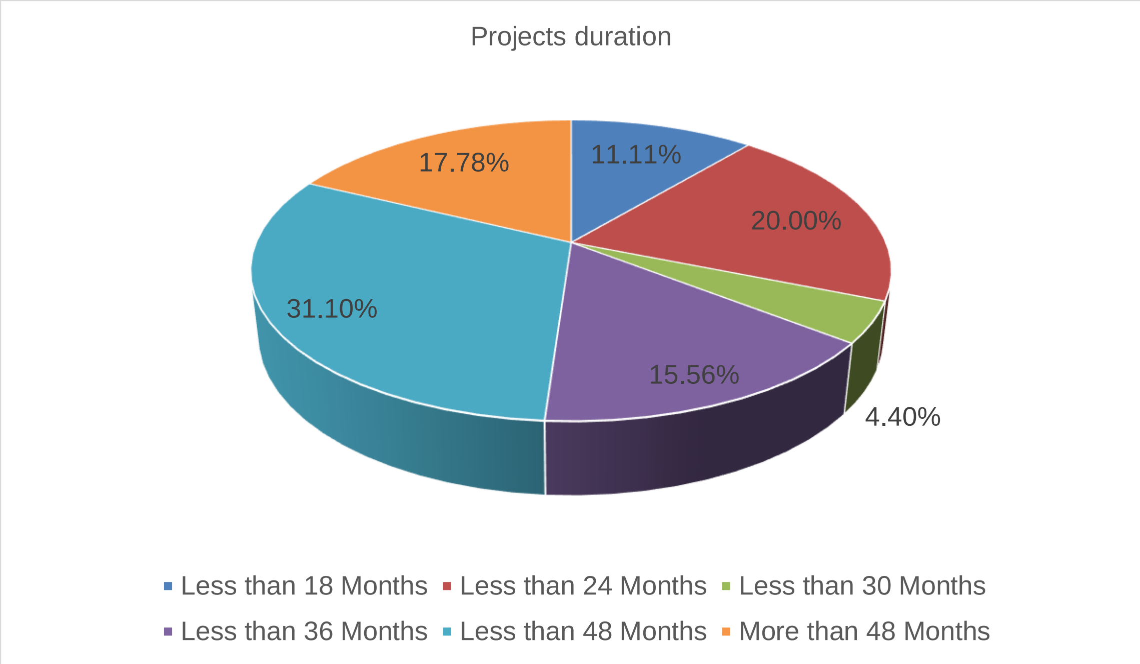

Figure 6. Classification of the projects according to projects duration (Author)

Figure 7 represents the impact of project duration on the percentage of site OH cost. Each category consists of the minimum, average, and maximum percentage of site OH cost. Projects with a duration of less than 18 months showed the lowest average percentage rate with 7.5%, and with a duration of more than 48 months showed the highest average percentage rate with 12%.

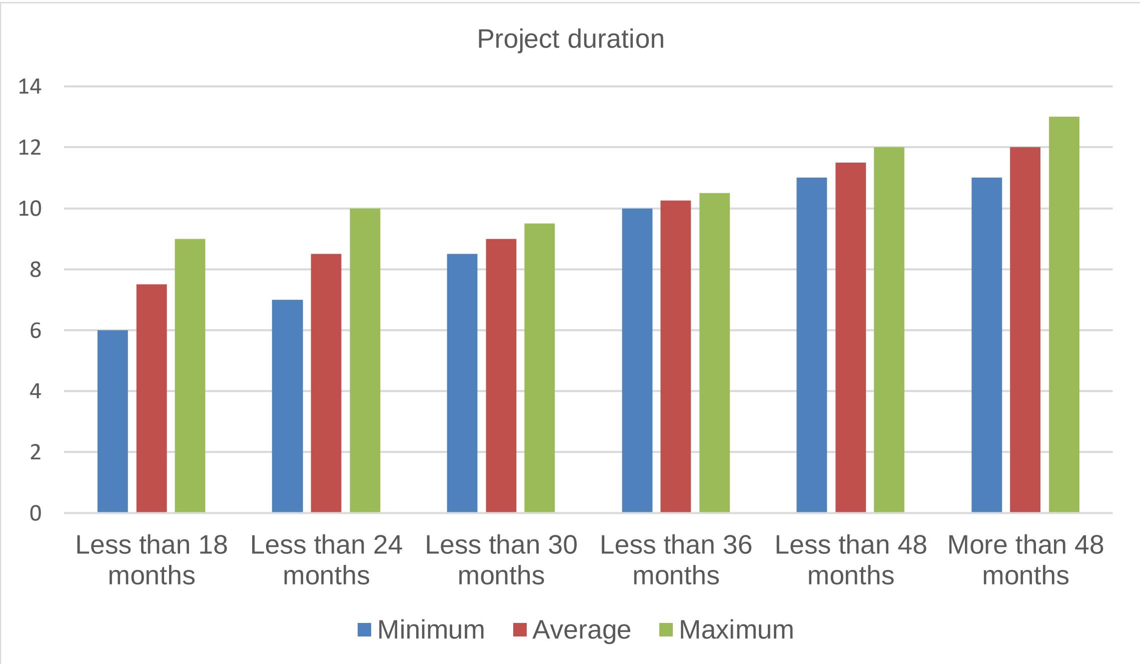

Figure 7. Impact of project duration on-site overhead cost percentage (Author)

Through the analysis of the above data, it was obvious that the average percentage of site OH costs increases with the increase of the project duration, which proves that there is a linear relationship between the project duration and the percentage of site OH costs.

3.3.3. Impact of project budget on-site overhead cost

The collected projects were categorized as per their actual contract value into eight categories, as shown in Figure 8. The collected data consists of that 11.10% less than 30 million EGP, 8.89% less than 50 million EGP, 20% less 100 million EGP, 28.89% less 200 million EGP, 6.67% less than 350 million EGP, 4.44% less than 500 million EGP, 4.44% less one Billion EGP and 15.56% more than One Billion EGP.

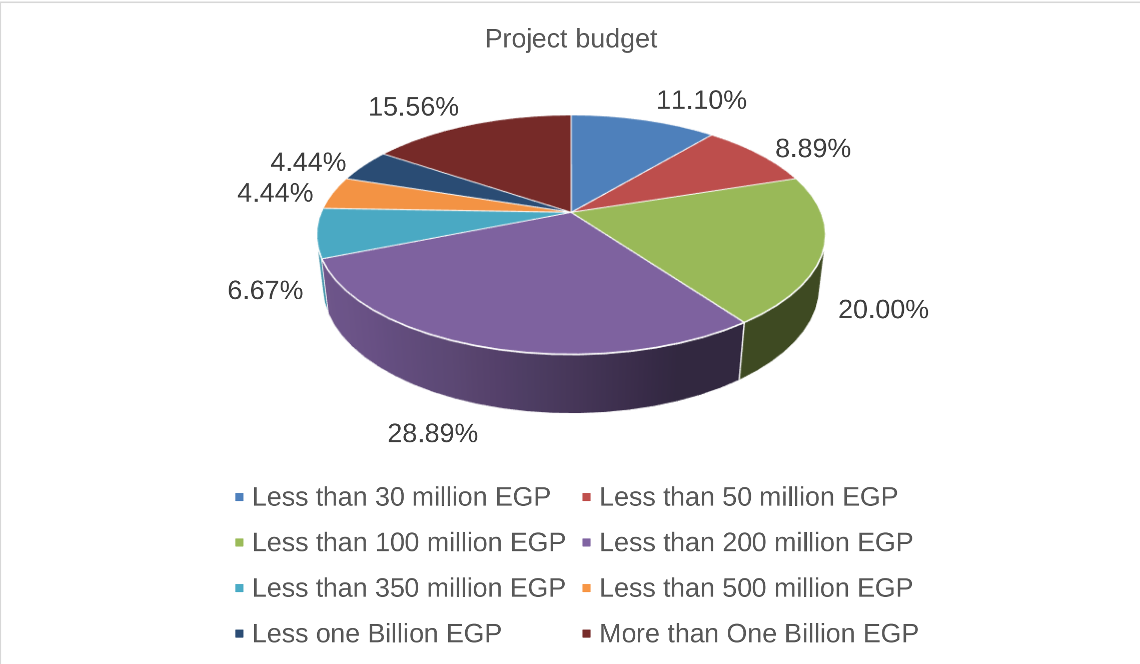

Figure 8. Classification of the projects according to projects budget (Author)

Figure 9 shows the impact of the contract value or project budget on the contractor percentage of site OH cost. Each category consists of the minimum, average, and maximum percentage of site OH cost. Projects with a project budget of less than 30 million EGP showed the lowest average rate with 7.5%, and with a project budget, more than one billion EGP showed the highest average rate with 12%.

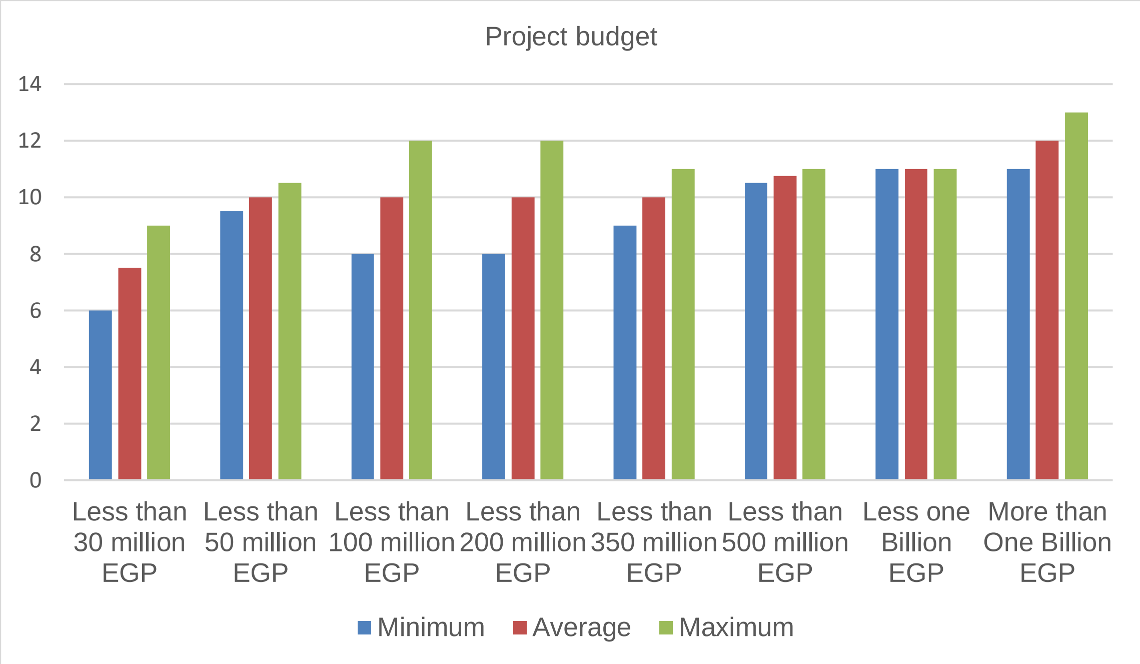

Figure 9. Impact of project budget on-site overhead cost percentage (Author)

Through the analysis of the above data and figures, it was very significant that there is a linear relationship because the average site OH percentage increase with the increase of the project budget.

3.3.4. Impact of project location on-site overhead cost

The project location is considered one of the main essential factors in the construction industry because it affects procurement, transportation, labours, and many other things while estimating a new project. The forty projects were divided into two categories, as shown in Figure 10, projects in the city, and projects into rural areas to check the difference and measure the impact of the project location on the percentage of site OH costs. The classification of the collected projects showed 46.67% of the projects are located inside the city, and 53.33% of the projects are in rural areas.

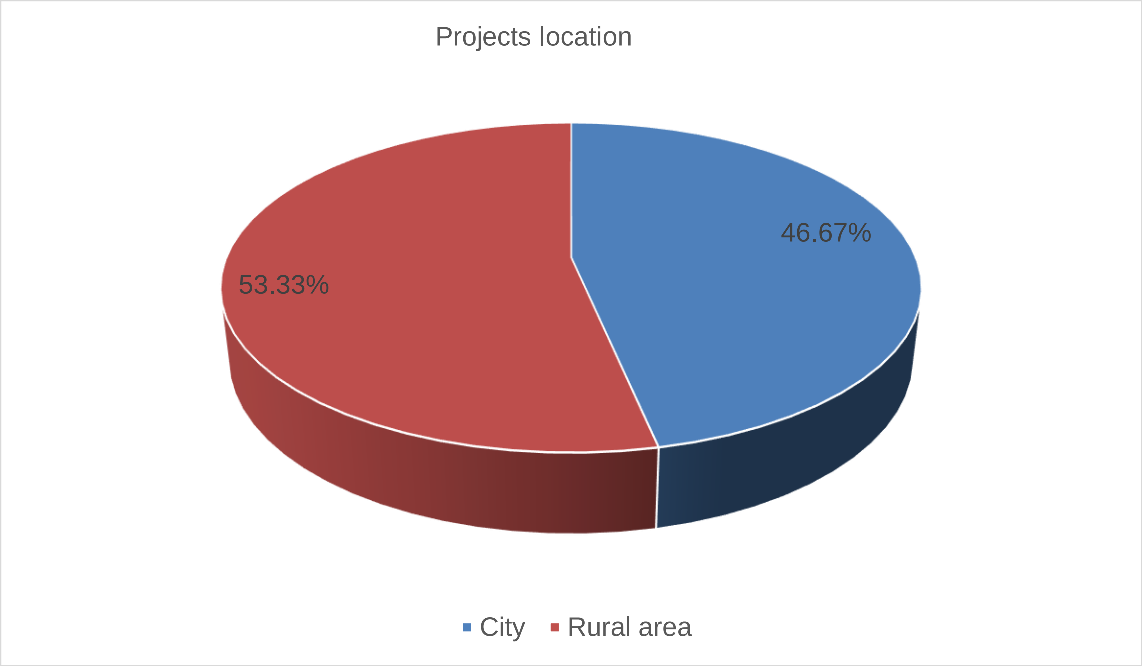

Figure 10. Classification of the projects according to projects location (Author)

Figure 11 represents the impact of project location on the percentage of site OH cost. Each category consists of the minimum, average, and maximum percentage of site OH cost. Projects which are located inside the city have a lower average percentage rate of 9% than the projects located in rural areas with an average percentage rate of 10%.

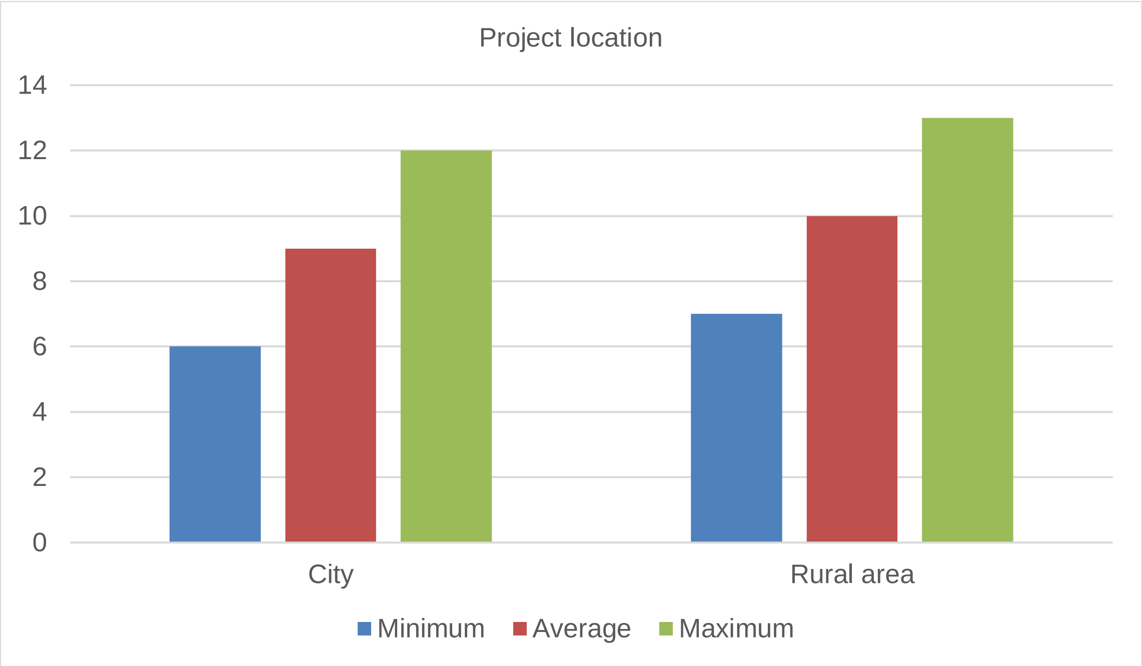

Figure 11. Impact of project location on-site overhead cost percentage (Author)

Through the analysis of the above data and figures, It shows that the percentage of site OH costs is affected by the project location as projects located in rural areas as countrysides or desert areas have a higher average percentage rate than the projects located in the city surrounded by most of the facilities.

3.3.5. Impact of client type on-site overhead cost

The collected projects were divided into two categories, as shown in Figure 12, based on the client type, which are private clients and public clients represented in the government. It shows that 93.3% of the projects are private clients, and 6.67% is a public or governmental client.

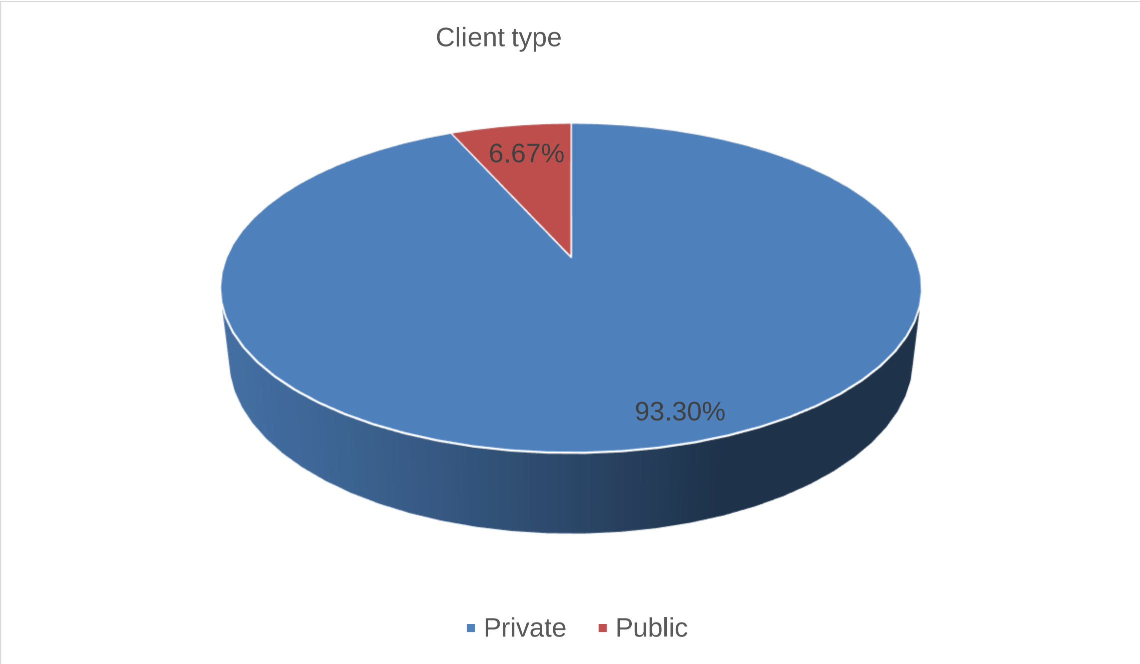

Figure 12. Classification of the projects according to client type (Author)

Figure 13 shows the impact of the client type on the contractor site OH cost, and it is categorized to private and public clients. Each category consists of the minimum, average, and maximum percentage of site OH cost. Projects for private clients have a higher average rate of 9.5% than the projects for a public client with an average rate of 9.25%.

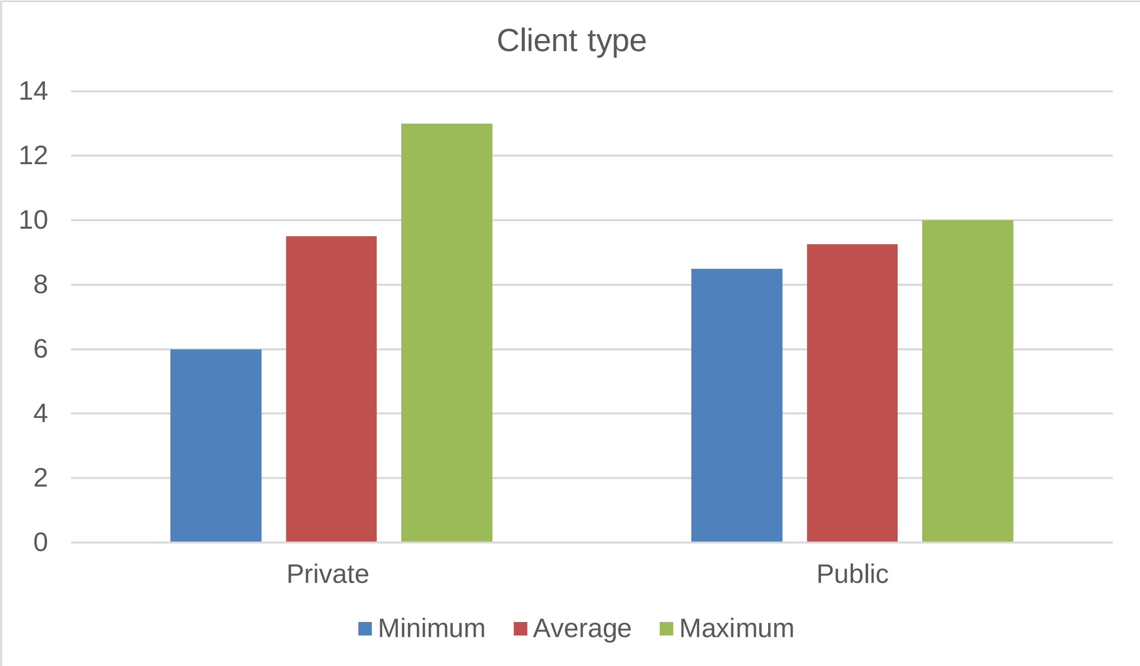

Figure 13. Impact of client type on-site overhead cost percentage (Author)

Through the analysis, the type of client had a low impact on the percentage of site OH cost with a difference of 0.25% between the private and public clients.

3.3.6. Impact of contract type on-site overhead cost

The contract type is considered an essential element in the construction industry as it saves the contractor rights toward the client and depends on the contract type, changes should be done in the prices and terms and conditions. The contract types included in this research are lump sum, cost-plus, and unit rate contracts. The collected data were categorized as shown in Figure 14, according to the type of contract where 6.67% of the projects are lumpsum contracts, 15.56% are cost-plus contracts, and the unit rate contracts have the highest share with 77.78%.

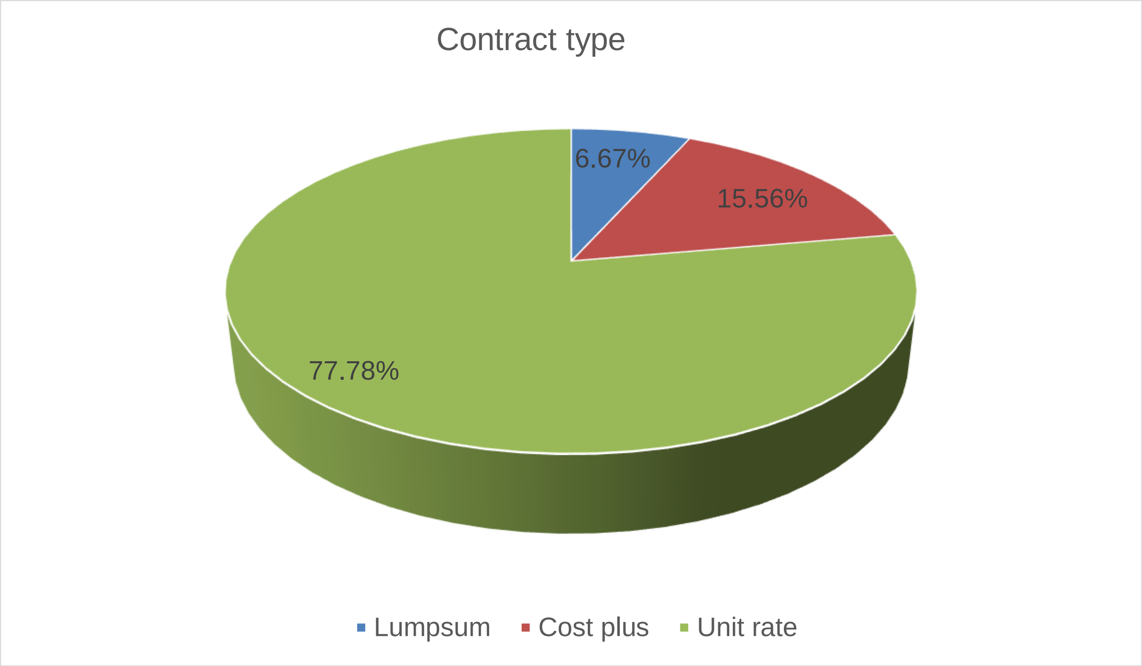

Figure 14. Classification of the projects according to contract type (Author)

The impact of the contract type on the percentage of site OH cost is represented in Figure 15. Each category consists of the minimum, average, and maximum percentage of site OH cost. The projects with lumpsum contracts have the highest average percentage rate of 10.5%. The cost-plus contracts have an average percentage rate of 9.5%, and the unit rate contracts have the lowest average percentage rate of 9%.

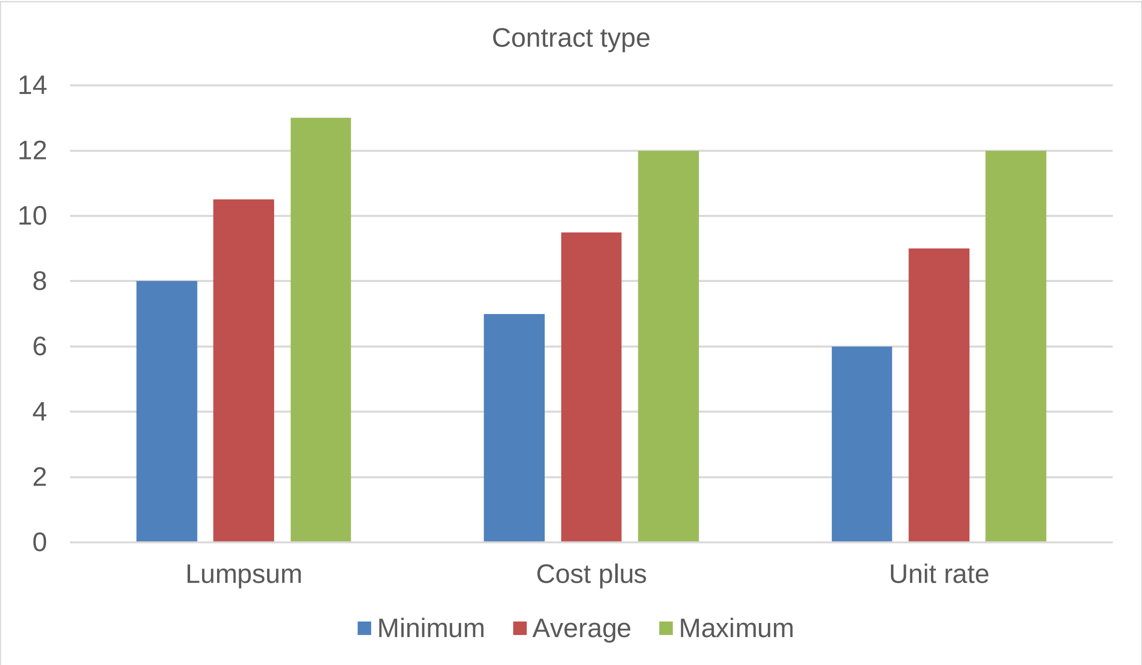

Figure 15. Impact of contract type on-site overhead cost percentage (Author)

The output of the analysis showed that most of the contracts in Egypt are unit rate contracts with 77.76%, and from the analysis proves the contract type has an impact on the percentage of site OH costs as it differs with the change of contract type.

3.3.7. Impact of company category on-site overhead cost

It is common that clients define the category or the grade of the construction company in the tender documents, and in Egypt, the construction companies are categorized into several categories. These categories indicate the company's experience, quality, and financial capability. The collected data was limited to the first and second categories only, as shown in Figure 16 because these two categories have construction management processes and costs estimation tools. The classification of the collected data as follows: 84.44% of the projects are for first-grade companies and 15.56% for second-grade companies.

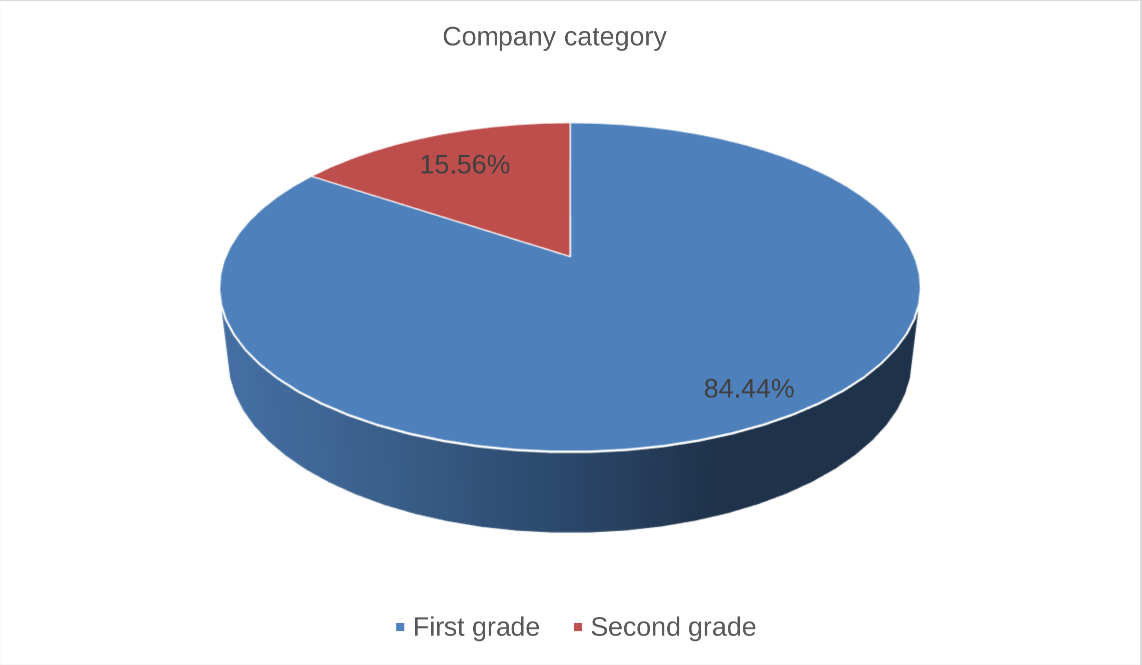

Figure 16. Classification of the projects according to company category (Author)

The impact of the company category on the contractor site OH cost is shown in Figure 17. Each category consists of the minimum, average, and maximum percentage of site OH cost. The projects that belong to first-grade companies have a higher average percentage rate of site OH cost of 9.5% than the projects for second-grade companies with an average percentage rate of 8.75%.

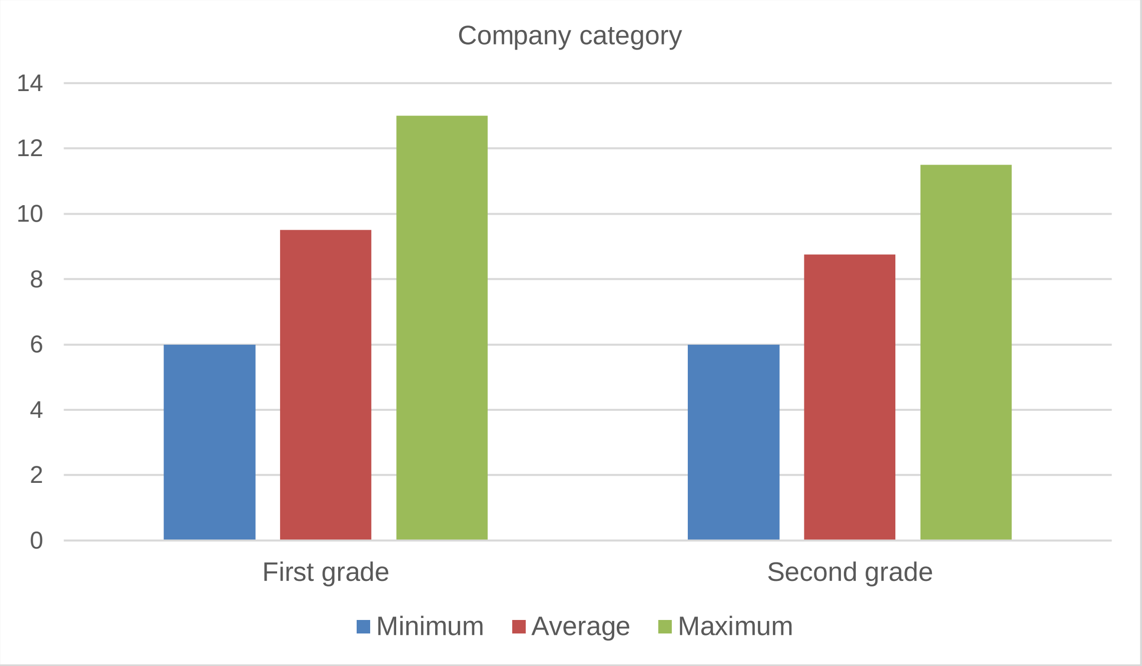

Figure 17. Impact of company category on-site overhead cost percentage (Author)

Through the analysis of the above data, it was significant that the grade of the construction company has a high impact on the percentage of site OH costs in the Egyptian construction industry.

3.4. Summary

The study outlined some facts that were important to explain and recognize the site OH costs for construction projects in Egypt. These facts will be the main starting point for developing the ANN model for determining the percentage of site OH costs. These facts can be easily summed up as follows:

- From the literature review view ten factors that affect the percentage of the site OH costs for Egypt's building projects. These factors were simplified to 7 factors to ease and fasten the data collection process and developing the model.
- Data collected for forty construction projects in Egypt. This data was analysed based on the above mentioned seven factors and illustrated the impact of each factor on the site OH costs. The top two factors affecting the percentage of site OH costs are project duration and project budget.
- Client type and project type showed the lowest factor that affects the site OH costs.

4. Artificial Neural Network guideline

4.1. Introduction

One of the purposes of this study is to create an ANN model to predict the site OH costs percentage for construction projects. The decision-makers could benefit from this during the tender submission in the Egyptian construction industry.

This chapter explains the ANN components, and the steps followed using Neural Designer software to develop the model to estimate the site OH cost percentage in Egypt. In chapter two, All the factors that affect the site OH cost in Egypt were identified and listed were used in categorizing the collected projects as input variables in creating the dataset file to develop the ANN model. The output variable of the ANN model is site OH costs a percentage of the project's contract value.

Neural Designer software was used in developing the model, and the software's manual was used as a guideline. Furthermore, the test and error method was carried out to validate and select the model. The ANN model consists of the below sections: (Neural Designer, 2020)

- Identify the application type of the ANN classification or approximation.
- Data set
- Network architecture
- Train the network
- Model selection
- Testing analysis

4.2. Application type

A model can be described through mathematical principles as a representation of the real-world system or method. The mapping between input and output variables is typically shown. ANN is used in this regard to identify relationships, predict, recognize patterns and classify data associations, and they are classified into two types, which are approximation and classification (Neural Designer, 2020).

The ANN could be used in classification and prediction through solving challenges in different areas of construction management as risk analysis, construction costs estimation, operation, and maintenance cost, productivity rate, and it could assist the decision-makers as contractors or managers to take an accurate decision (Kulkarni, et al., 2017).

In this research, the approximation type was selected. Here the ANN learns from the information of a data set consisting of instances with variables of input and target. The problem of fitting a data function be an approach. In an approximation problem, the fundamental objective is to model one or more target variables depending on the data input (Neural Designer, 2020).

The backpropagation algorithm is the most used in ANN. It is a feed-forward network. This model provides a continuous method for mapping between such input and output quantities. No connections between neurons in the same layer are enabled in a backpropagation network. Each layer neuron gives the next layer input to each neuron. This network employs regulated learning, which includes knowledge of both input and output patterns. The error measured at network output is propagated to change weights through the neuron layers. As shown in Figure 18, the X represents the inputs that go through different paths to identify their weights. The networks calculate the target output from the input layer to the perceptron layers reaching the output layer. The model calculates the error by subtracting the actual output from the desired one, then travel back to the perceptron layer and input layer to adjust the weigh to decrease the error. The model repeats this process until it achieves the desired output (Guru99, 2020).

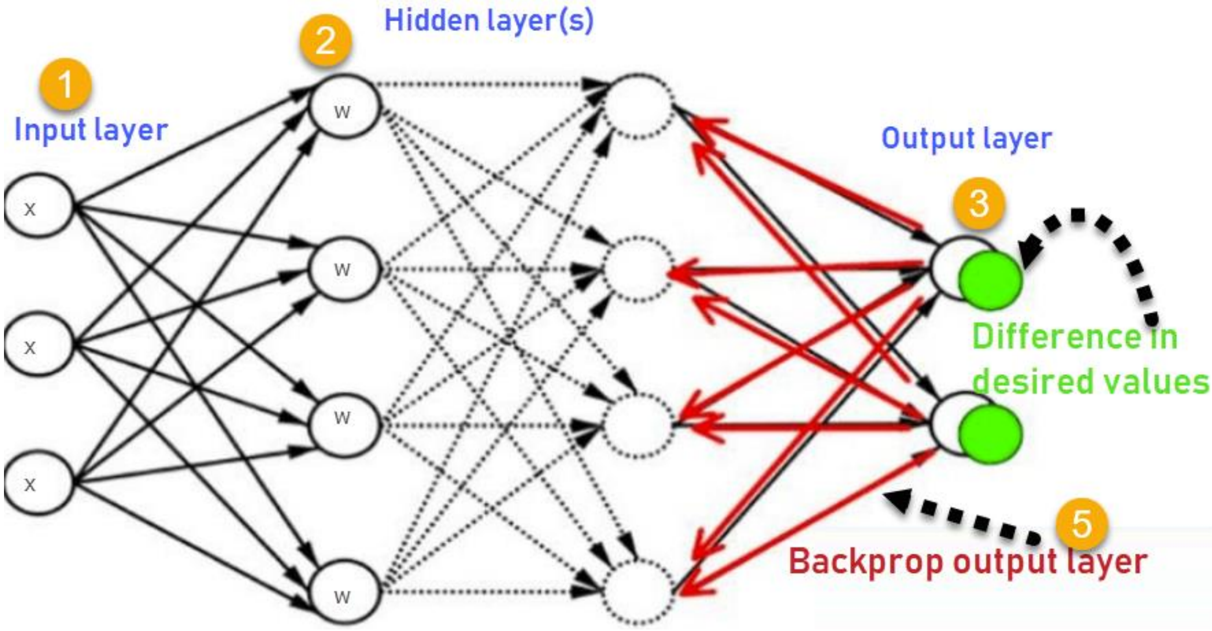

Figure 18. Backpropagation algorithm (Guru99, 2020).

4.3. Data set

The Neural Designer software accepts data from files as Comma-separated values, Excel, and Open Office or from databases as MySQL, Oracle, and SQLite. In this research, the data set was prepared using Excel and was imported (Neural Designer, 2020).

The data set consists of columns and rows. The columns represent the variables, and the rows represent the projects. These variables are classified to input variables that are independent, and output variables are dependent (Neural Designer, 2020).

Datafiles	Databases
• CSV. • Excel. • OpenOffice Calc. • Etc.	• Oracle. • MySQL. • SQLite. • Etc.

Figure 19. Types of data source (Neural Designer, 2020)

4.4. Network architecture

The neural network is inspired by the biological brain as it consists of artificial neurons all connected in a network. The structure of this network has parameters that should be adjusted to be able to perform. One of the most critical parameters is the perceptron layer (Neural Designer, 2020).

The perceptron layer is also known as dense layers that are responsible for learning the network. As shown in the below figure, the perceptron neuron receives inputs in numerical form combined with weights and bias to produce output in numerical form (Neural Designer, 2020).

- Inputs (X_1…., X_n)
- Weight ((W_1….,W_n)
- Bias (b)
- Output (y)

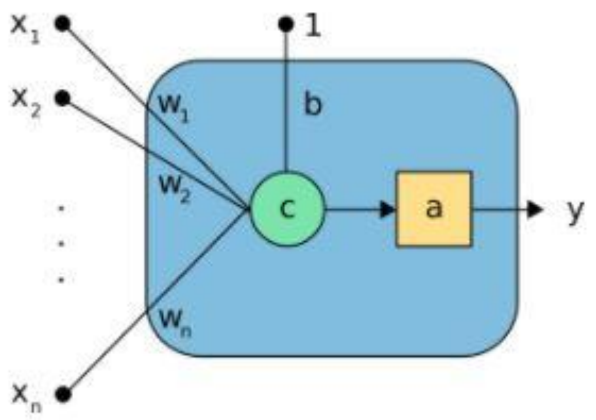

Figure 20. Perceptron neuron (Neural Designer, 2020)

The combination function, as shown in Figure 20, is represented by (C), which mainly responsible for producing one input value, as shown in Equation 2.

$$combination = bias + \sum weights \cdot inputs$$

Equation 1. Combination function (Neural Designer, 2020)

This single input value is used in the activation function which is represented by the letter (a) as shown in Figure 20, to define the output as shown in Equation 3. The activation function identifies the function represented by the neural network, which is a hyperbolic tangent function, sigmoid (logistic) function, and linear function.

$$output = activation(combination)$$

Equation 2. Activation function (Neural Designer, 2020)

4.4.1. Hyperbolic tangent function:

The hyperbolic tangent is a sigmoid function and considered as one of the common activation functions used. It ranges between -1 and +1, as shown in Figure 21 (Neural Designer, 2020).

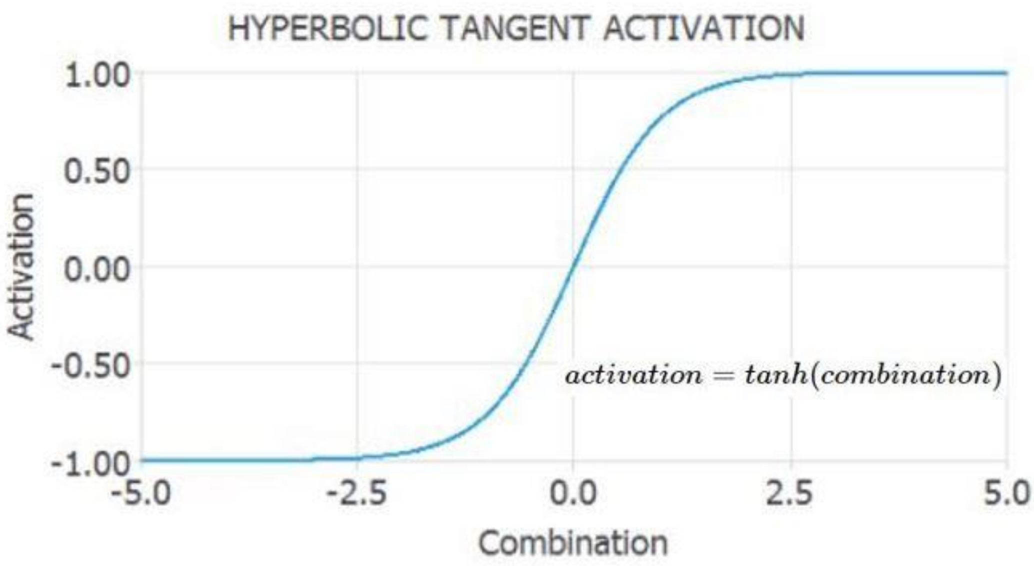

Figure 21. Hyperbolic tangent function (Neural Designer, 2020)

4.4.2. Logistic function:

Logistic is also a common activation function which is commonly used, and it is a different type of sigmoid functions. It ranges between -1 and +1, as shown in Figure 22 (Neural Designer, 2020).

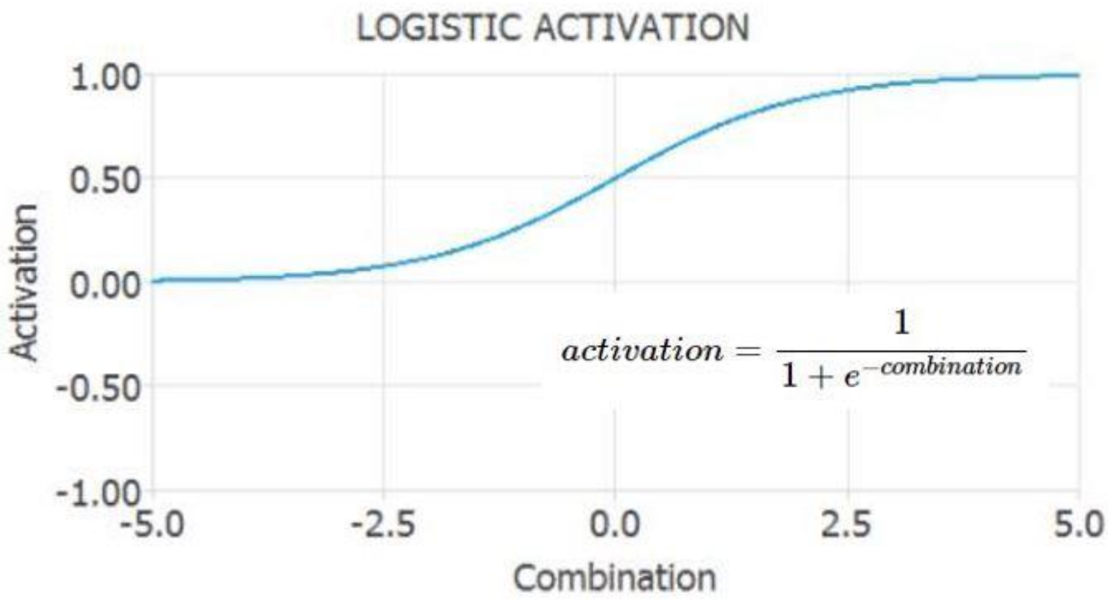

Figure 22. Logistic function (Neural Designer, 2020)

4.4.3. Linear function

The linear function is the output or target, which is equal to the combination of this neuron, as shown in Figure 23 (Neural Designer, 2020).

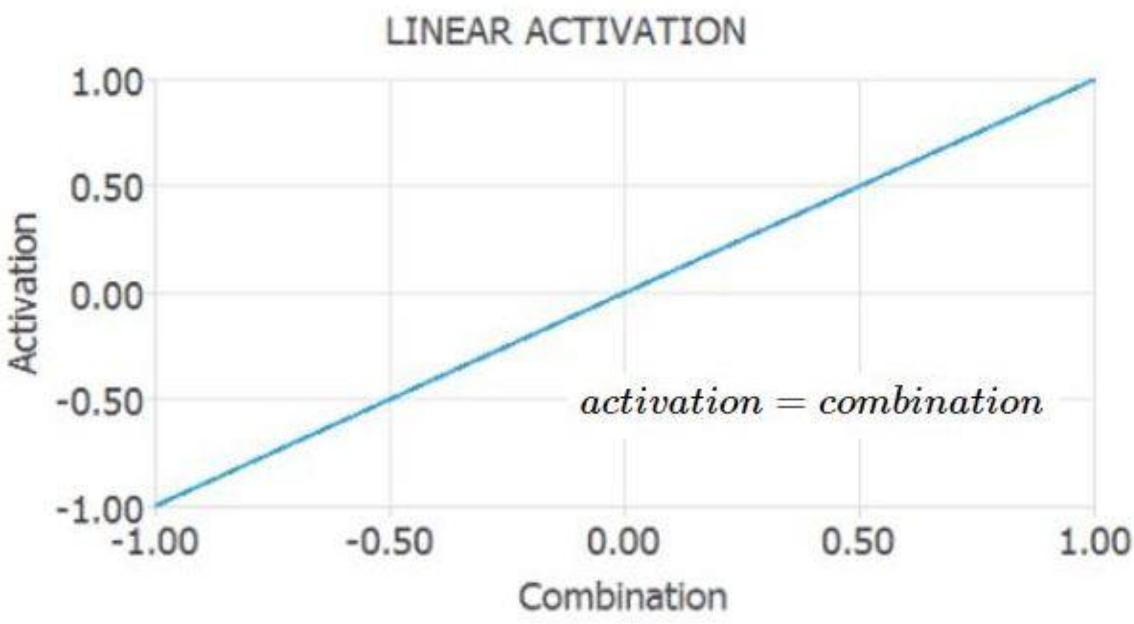

Figure 23. Linear function (Neural Designer, 2020)

4.5. Training strategy

The training technique is considered the approach used for the process of learning. The training technique is used to get the least possible loss to the model. This is achieved when several parameters are checked and match the model into the data set. It is divided into two concepts loss index and optimization algorithm (Neural Designer, 2020).

Network training is a mechanism used to change the weight or the strength of links by one or more learning methods. All test models tested in this study were trained using a backpropagation algorithm in a supervised manner. In order to minimize errors, between the actual output and the desired output are then determined and used to change the network weight (by the software automatically). As the workouts progress, the network weights are changed continuously until the error converges to an appropriate amount in the measured performance. In the backpropagation algorithm, the error between the model output and the target output is progressively decreased. Therefore, the output mapping data is generated to decrease the RMSE. The algorithm decreases the RMSE, and the training stops when the RMSE is constant. The RMSE is used to measure model performance or accuracy (Neural Designer, 2020).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(S_i - O_i\right)^2}$$

Equation 3. Root mean square error equation (Neil & Hashemi, 2018)

The above equation represents the RMSE equation where (Neil & Hashemi, 2018):

(S_i)= The Predicted output value from the ANN model.

(O_i)= The actual output value from the dataset.

(n)= Number of samples or projects.

4.6. Model selection

Model selection seeks to define a network architecture with the best overall results to minimize the error of the data set. The underfitting and overfitting processes are the two common problems for the design of the neural system. The best generalization is accomplished with the use of a model whose design is better suited to fit the data (Neural Designer, 2020).

4.7. Testing analysis

The research aim is to compare the model outputs to the projects left for the test instance. The test procedures depend on the form of the project.

If the test results area is accepted, then this model is selected. The testing analysis consists of testing errors, linear regression, confusion matrix, ROX curve, misclassified instances, and cumulative gain. In this research, RMSE and linear regression analysis are used to test the model (Neural Designer, 2020).

4.8. Creating data file

The Neural designer software train, validate and test the database. The database was prepared on an external file using Excel. It consisted of forty construction projects site OH cost percentage. The software automatically classifies the database to 60% of the projects for training the model, 20% of the projects for selection, and the software leaves 20% of the projects for testing the model. The following steps show how to create a file (Neural Designer, 2020).

- Start the Neural Designer software
- Select the type of application required for the project. In this research, an approximation application was selected as shown in Figure 24.
- Then the software will ask to save the file in a specific location.

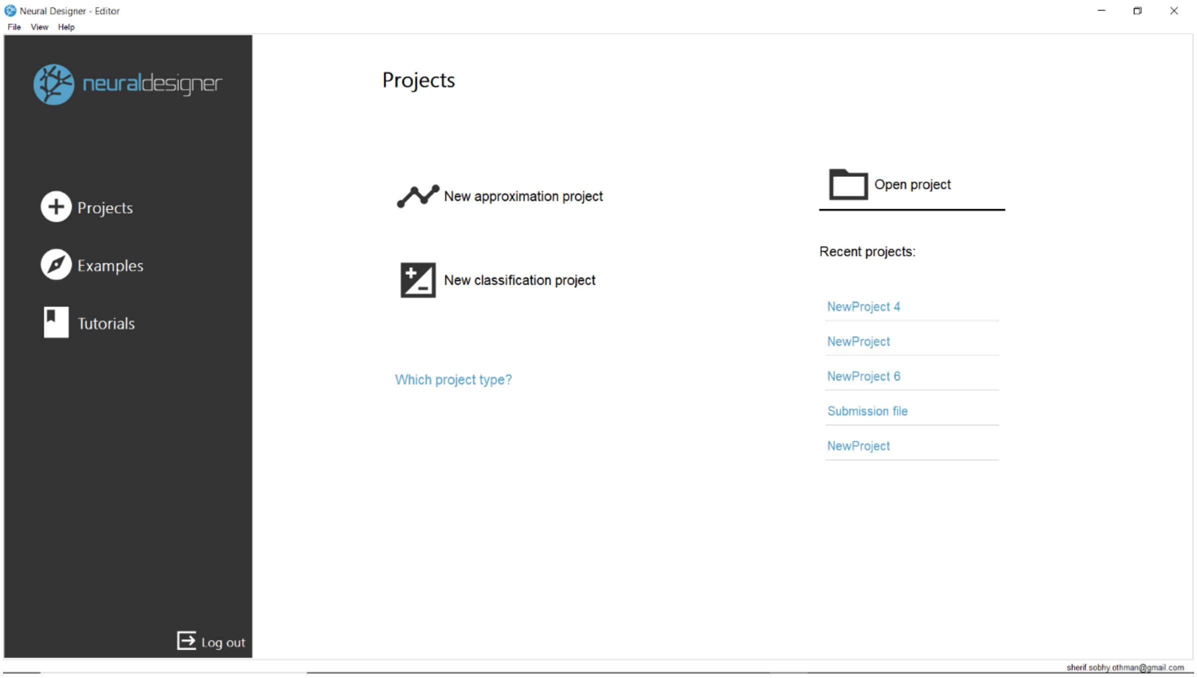

Figure 24. Model type selection (Author)

- When the software starts, it is time to import the dataset file as per the data source, as shown in Figure 25.
- Click on import data on the import data file since our data set is in Excel format. However, if the data set is in another format as SQL, Oracle, or another format, then the import database option should be selected.

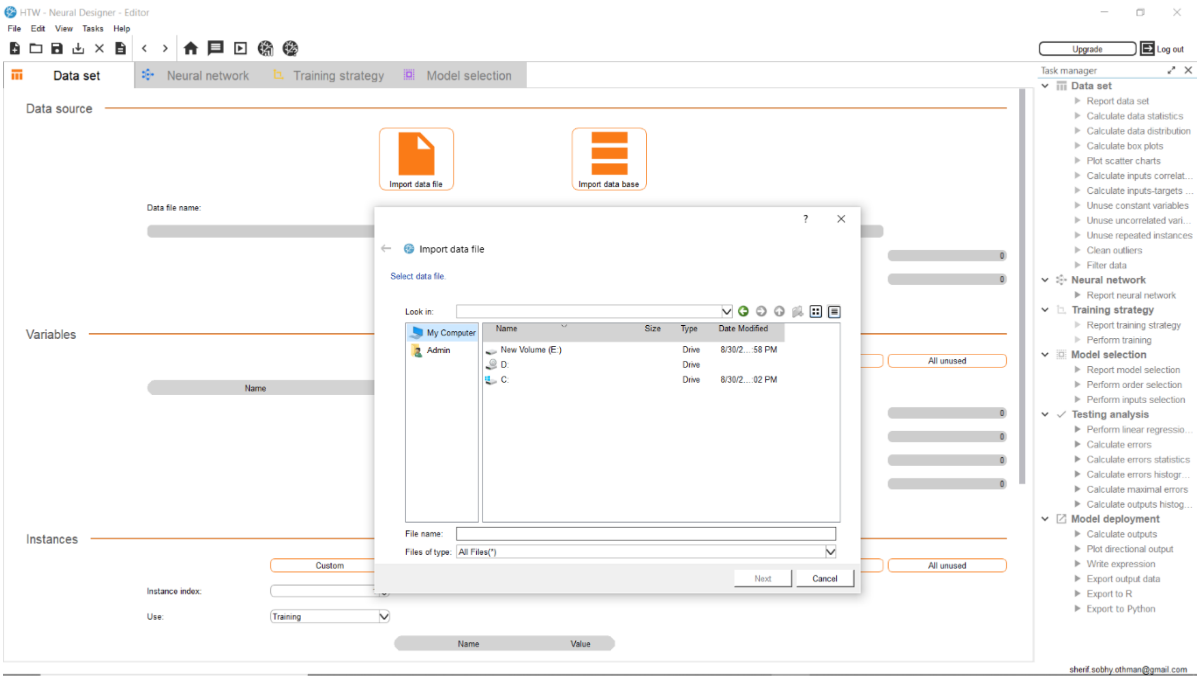

Figure 25. Import data file (Author)

- The data set file that will be imported to the software should already be created earlier, as shown Figure 26. This file should have specific criteria; otherwise, the software will not be able to identify the data.

- All the cells representing the data input should be in numeric format, not symbols or any other type, so the software can relate and connect the inputs.
- Each row is considered a single problem, and the columns are considered as variables of this problem. The last column represents the target output.
- The file should be saved in Excel format.

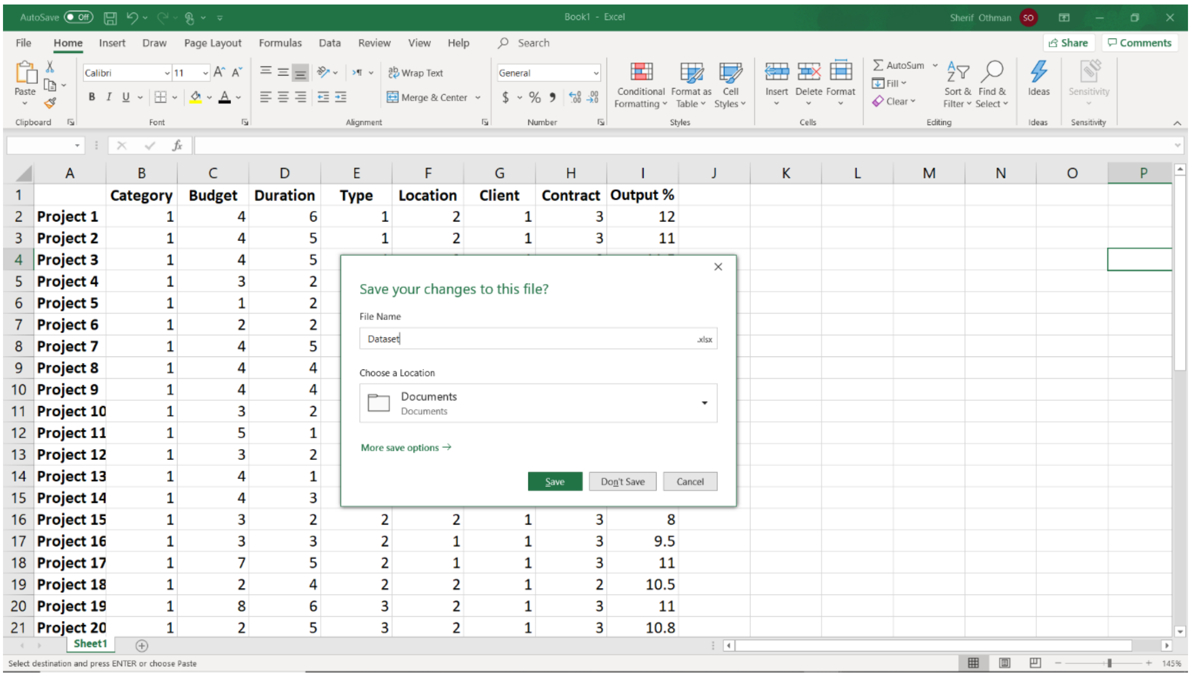

Figure 26.Excel data file (Author)

- Then back to the software, since the data file is ready to be imported
- The software will open a preview window for the data file, as shown in Figure 27, before it is imported to check that all the cells are in the right format and identified by the software.

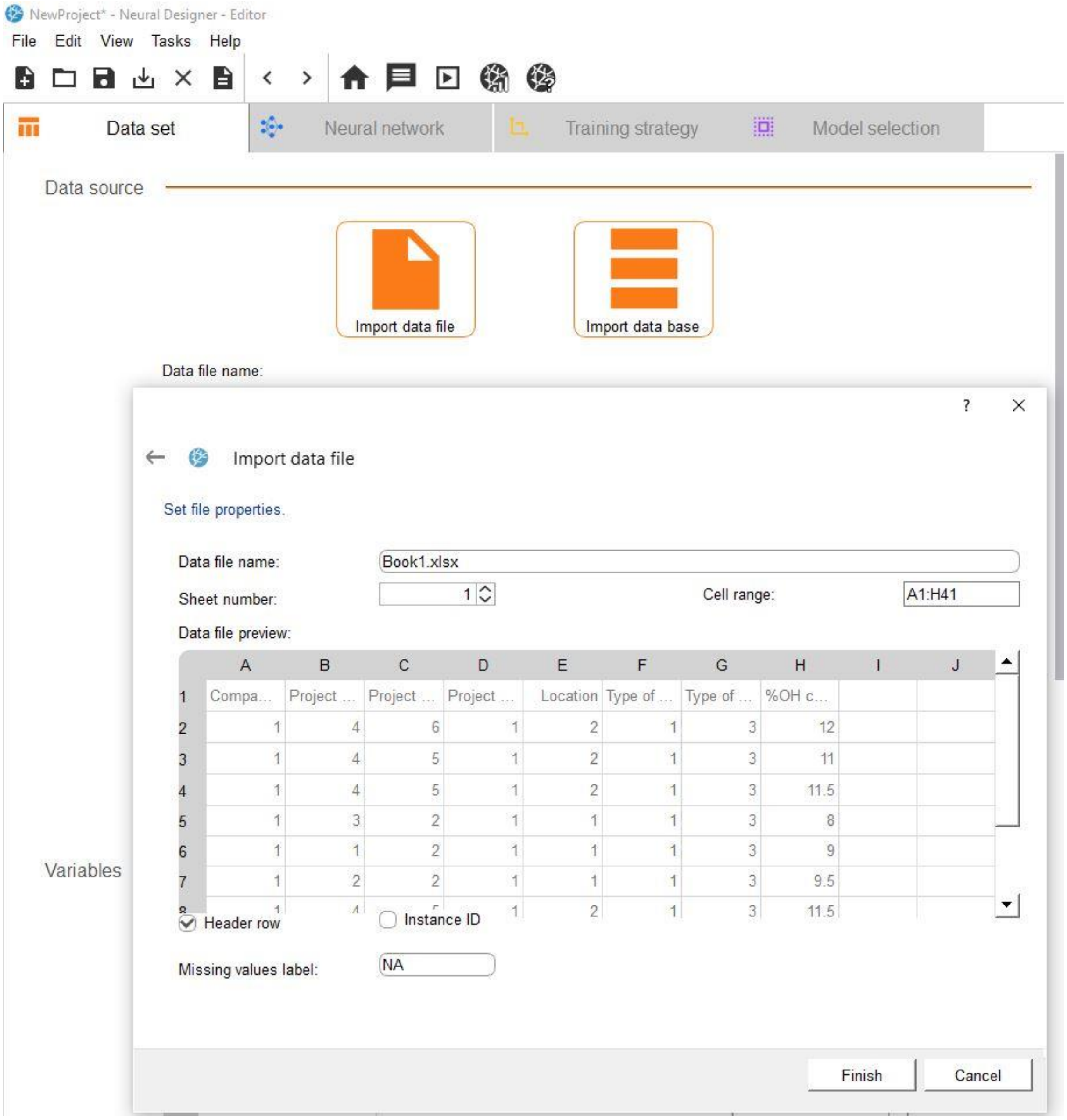

Figure 27. Data file - software view (Author)

- The software identified the variables as shown in Figure 28. From the variables option, identify what are the input variables and what is the output target.
- Neural designer software automatically indicates the input variables and the output target without adjusting anything.

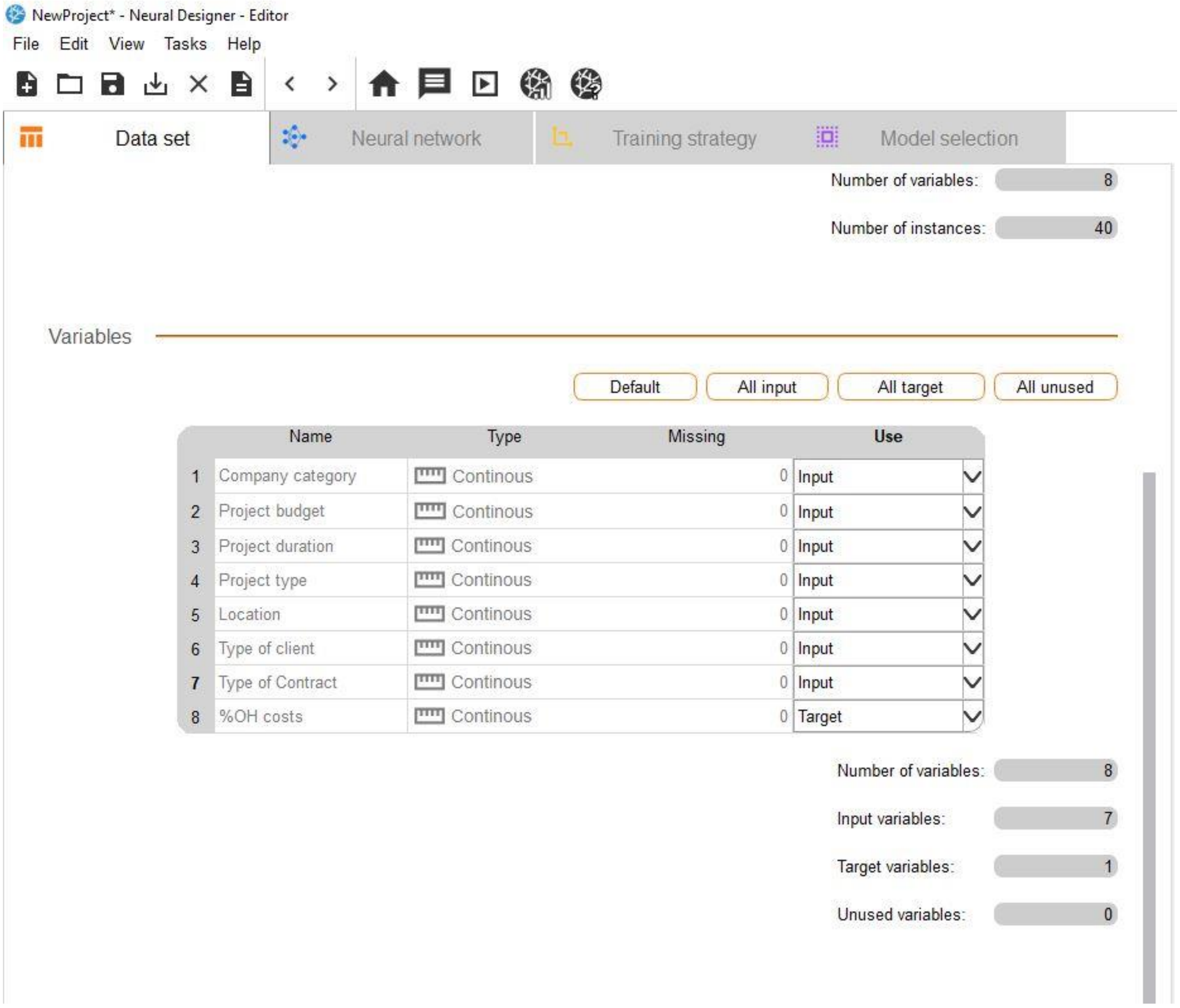

Figure 28. Variables Identification (Author)

- From the instance's option, the projects could be allocated to three types of instances, which are projects for training the model, selection, and testing the model.
- The software automatically separates the dataset to 60% of the projects for training, 20% for selection, and 20% for testing the model as shown in Figure 29.
- This could be changed depends on the objective of the user by pressing the custom button, or all the projects could be used for training by pressing the all training button.
- Now the dataset is ready, press on report dataset button, so the program reads the dataset parameters
- The next step is setting the parameters for the network architecture.

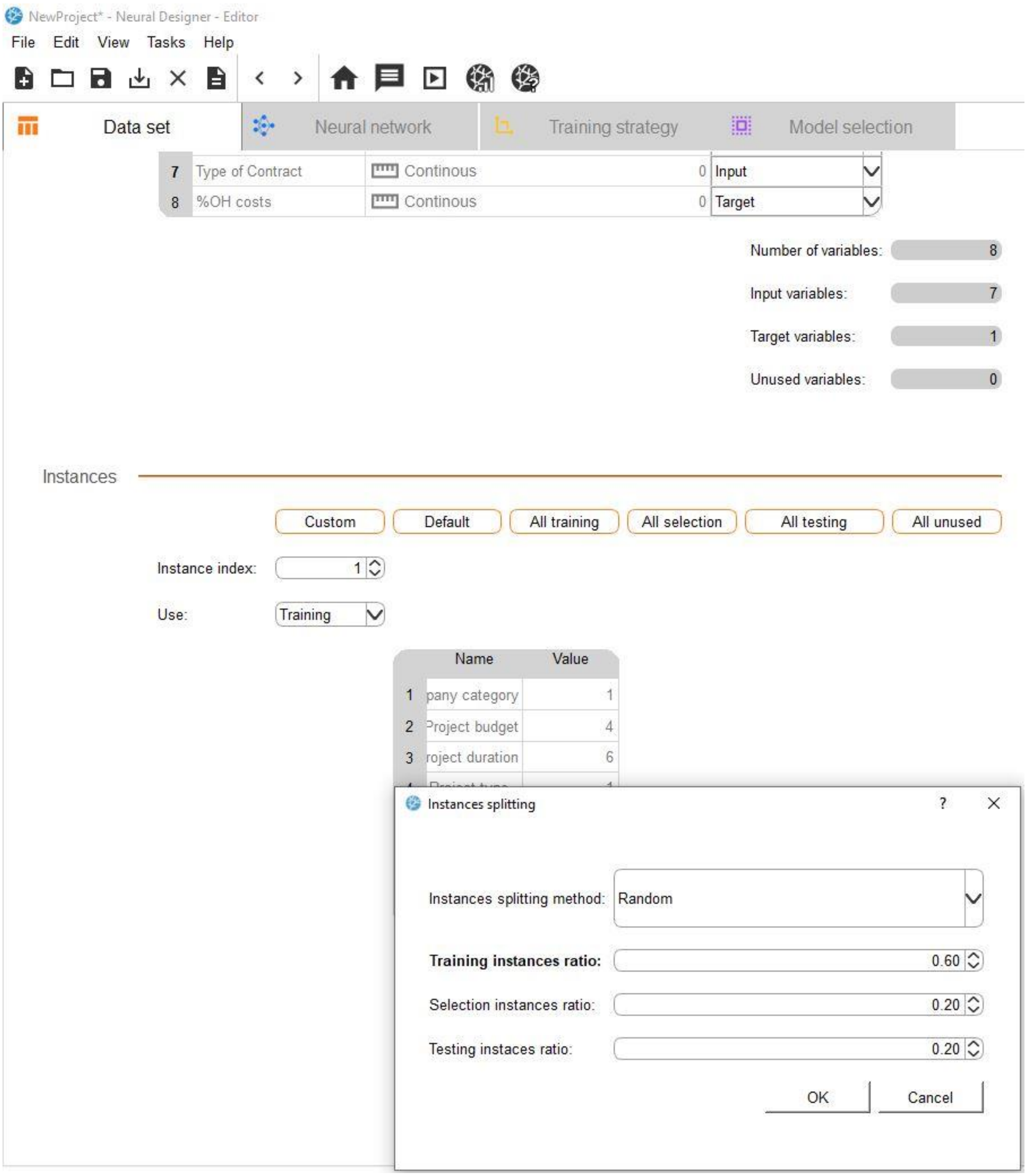

Figure 29. Instances splitting (Author)

- In the Neural network windows, the perceptron layers parameters, which were explained earlier, should be adjusted by selecting the number of hidden layers, a number of neurons in every layer, and the activation function, which is hyperbolic tangent or logistic activation function. The output layer is set as a linear function.
- After adjusting all the parameters, press on report neural network to where the model starts to design the model based on the previous selections as shown

in Figure 30.

- The next step will be the training strategy.

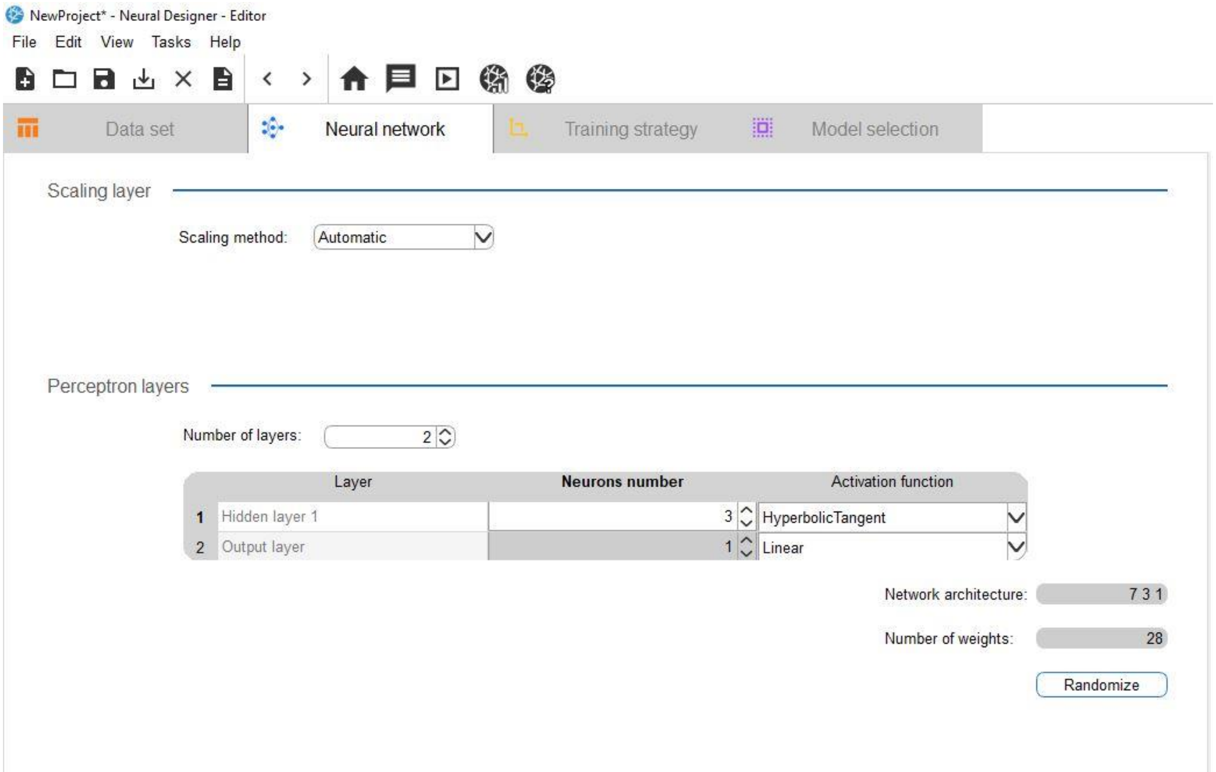

Figure 30. Neural Network architecture (Author)

- The training strategy is also called the learning process of the model. It is used to obtain a model with the least possible losses and is shown in Figure 31.
- The loss index was set by default to Normalized square error, and the optimization algorithm was also set by default to the Quasi-Newton method.
- The training accuracy can be set to low, medium, or high accuracy.
- The number of iterations is set by default to 1000 iterations and could be adjusted.
- When all the parameters are set, press on the report training strategy, then run the program by pressing perform training so the software can record the loss errors in each iteration.
- This process is used to train the network based on the model architecture that was selected earlier.
- This process should be repeated for every model in the trial and error section.
- The next step will be the model selection.

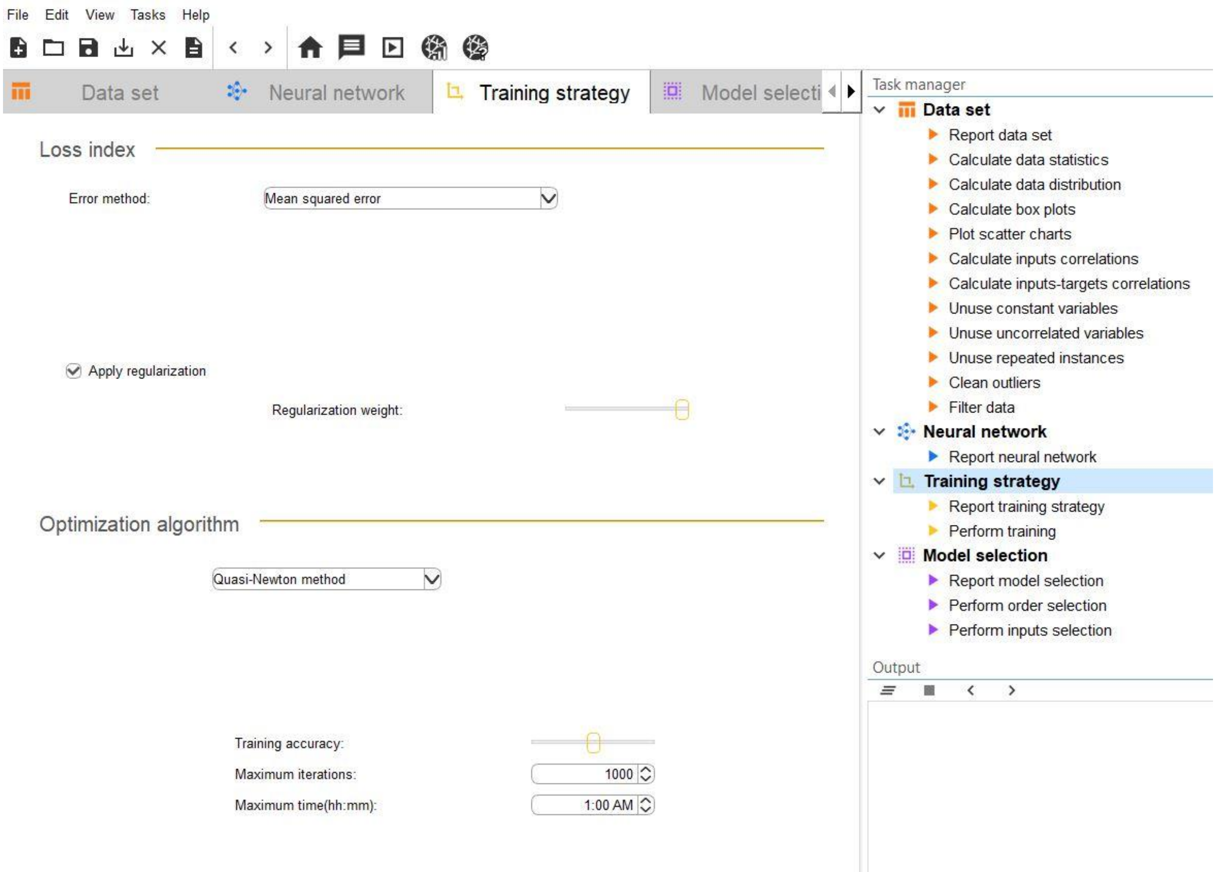

Figure 31. Training strategy (Author)

- The model selection is an excellent option in the Neural designer software, where the software automatically runs the model using different network architectures based on two algorithms, which are the order selection and inputs selection to select the best model with less possible losses as shown in Figure 32..
- The order selection algorithm is mainly responsible for identifying and selecting the optimal number of perceptron layers in the model. The default algorithm is an incremental order. This algorithm starts with a minimum number of perceptron layer, and this number increase in every iteration.
- The input selection algorithm is mainly responsible for selecting the optimum subset of the variables.
- After that, report the model by pressing on Report model selection.
- Run the model by performing order selection, then perform the selection of the input.
- The perceptron layers in the Neural Network tab should be adjusted to the new

architecture of the new model selected by the software

- Then repeat the process by running the training strategy.
- The next step will be testing of the model.

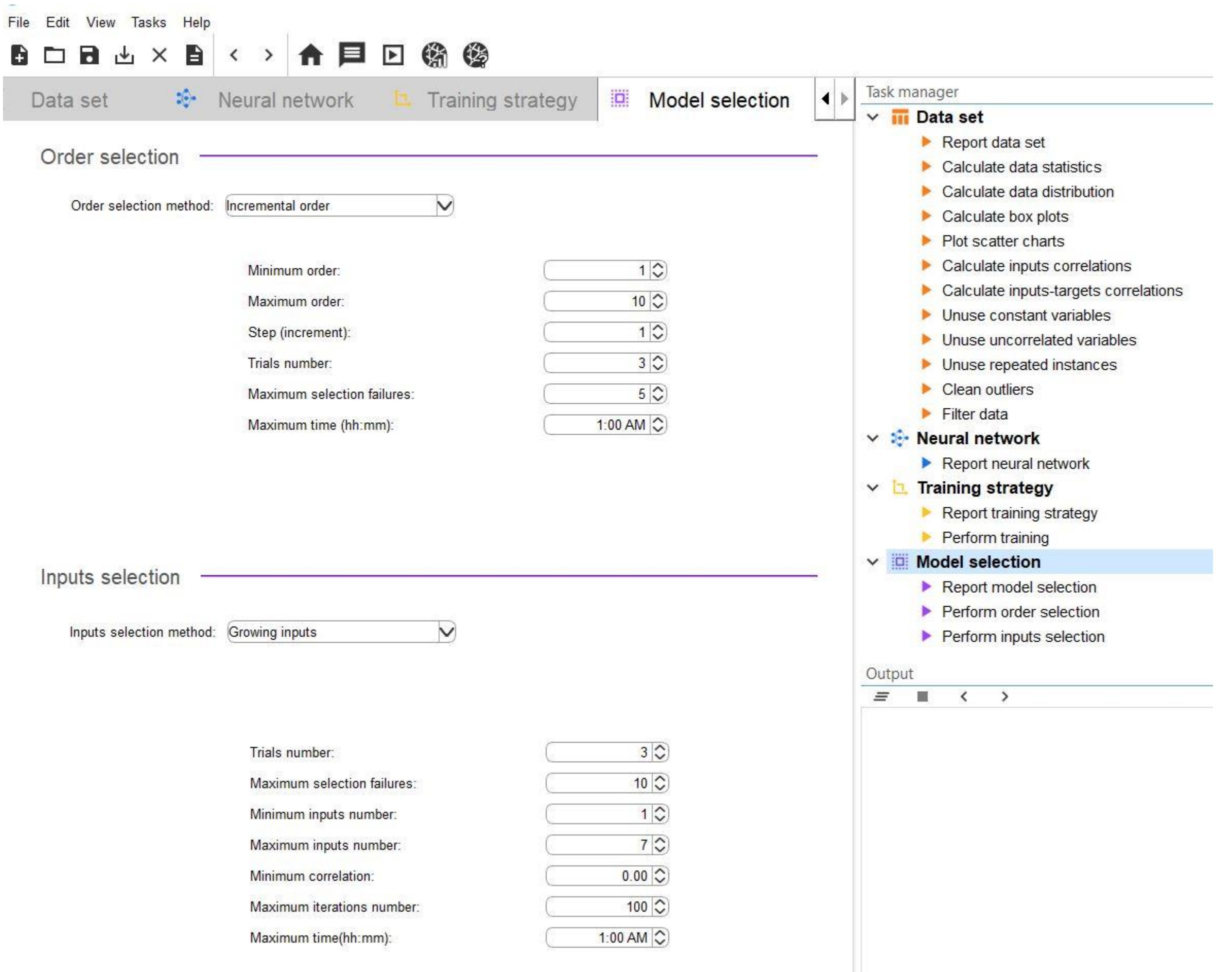

Figure 32. Model selection (Author)

- The testing analysis is used to measure the errors of the selected model, and all these errors should be recorded.
- The linear regression analysis represented in Figure 33 shows the relation between the predicted values and the actual values on the linear regression chart.

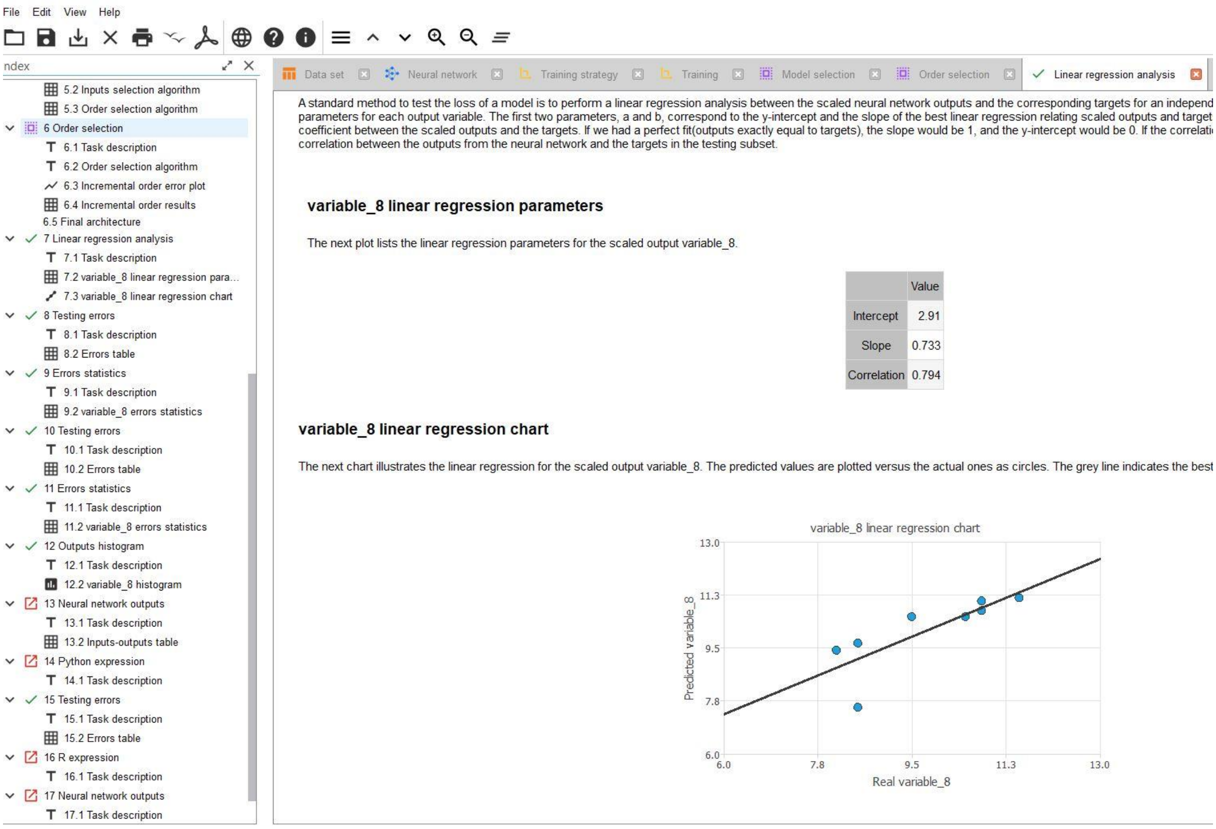

Figure 33. Linear regression chart (Author)

- The calculation error option is very important to calculate the different types of error of the model as sum squared error, MSE, RMSE, normalized squared error, and Minkowski error.
- One of these errors shown in Figure 34 should be selected to measure and compare the errors in training, selection, and testing between different models
- Record the errors of the RMSE

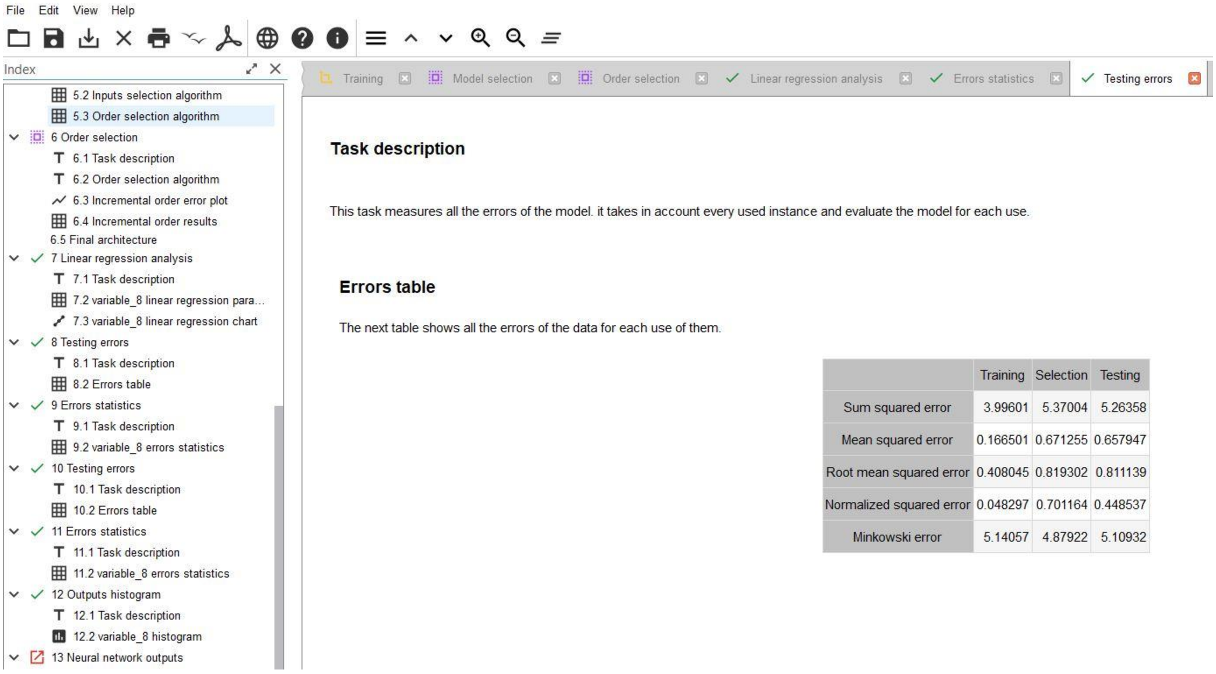

Figure 34. Training errors (Author)

- The error statistics options are used to identify the absolute and the percentage errors, which based on it, the percentage of error allowed when comparing the predicted output value to the actual one.
- Record the percentage value for each model as shown in Figure 35.
- The next step is the Model deployment

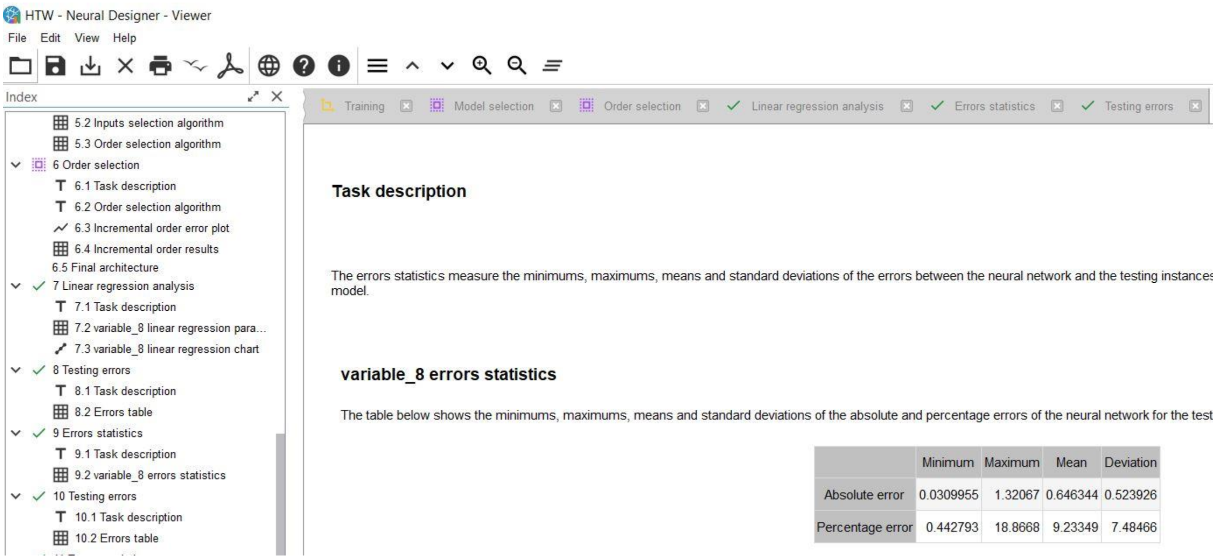

Figure 35. Error statistics (Author)

- The model deployment is used to put the selected model in function and estimate a new project.
- Press the calculate outputs button to estimate the percentage of site OH costs

for the new project, as shown in Figure 36.

- A new window will open to put the input factors based on the data encoding scheme explained earlier.
- The seven variables are identified, then press ok to estimate the output target.
- The model could be exported to Python or R languages.

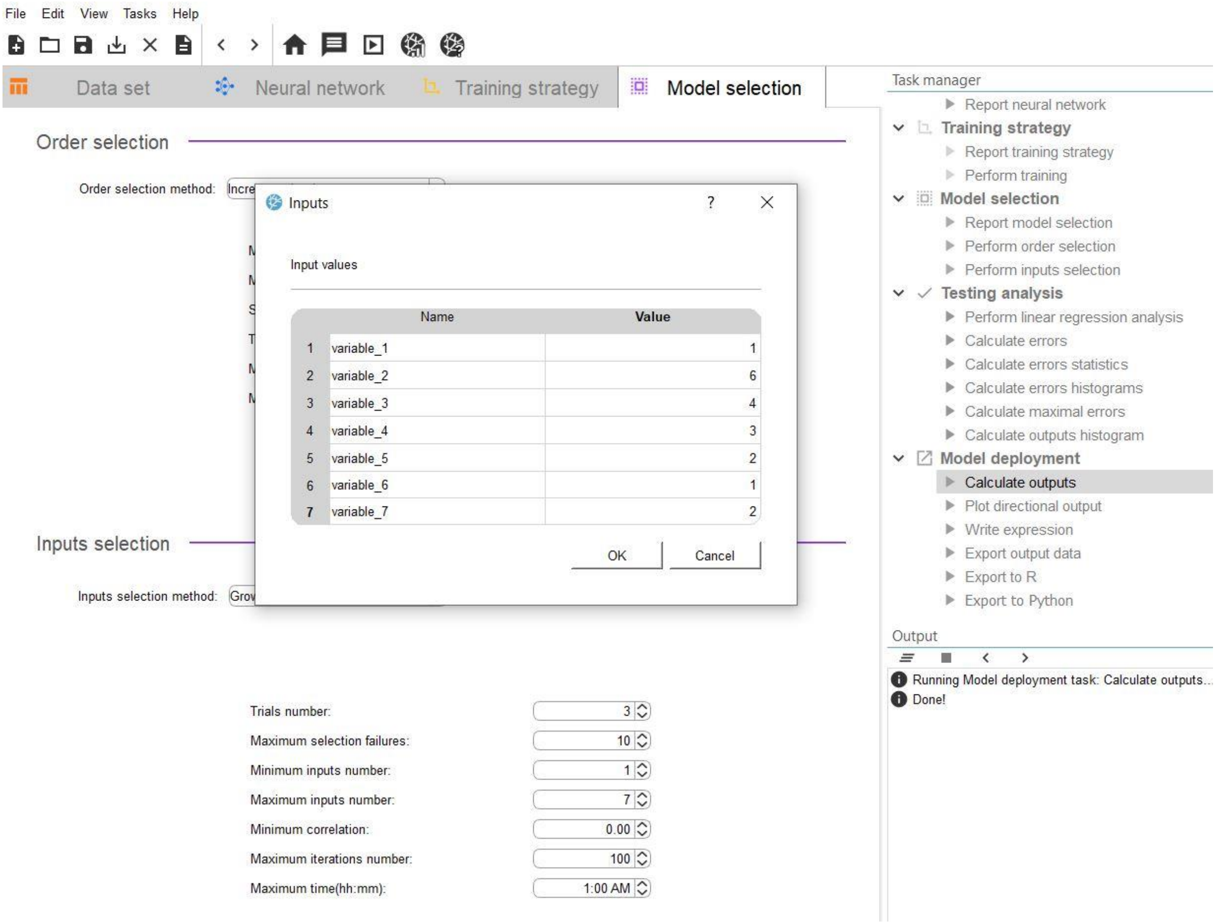

Figure 36. Estimate the percentage of site OH costs (Author)

5. Developing the model

5.1. Introduction

In the previous chapter, the guideline of the software was explained in detail, and the steps for training, selection, and testing the model. The concept of dataset preparation was explained briefly. This chapter will discuss the development of the actual model to forecast the percentage of site OH costs for construction projects in Egypt and show the results. The guideline from the previous chapter was used to achieve the below results. Figure 37 shows the steps for developing the ANN model.

Neural Designer software was used for training, selection, and testing the ANN model structure based on a feed-forward learning algorithm technology. Several alternative ANN model architectures were tested to achieve the minimum error percentage.

ANN is a supervised learning algorithm to approximate the function and predict the output based on the data input. This happens through the training process by minimizing the error (Brownlee, 2020).

"It is best to think of feedforward networks as function approximation machines that are designed to achieve statistical generalization, occasionally drawing some insights from what we know about the brain, rather than as models of brain function." (Goodfellow, et al., 2016)

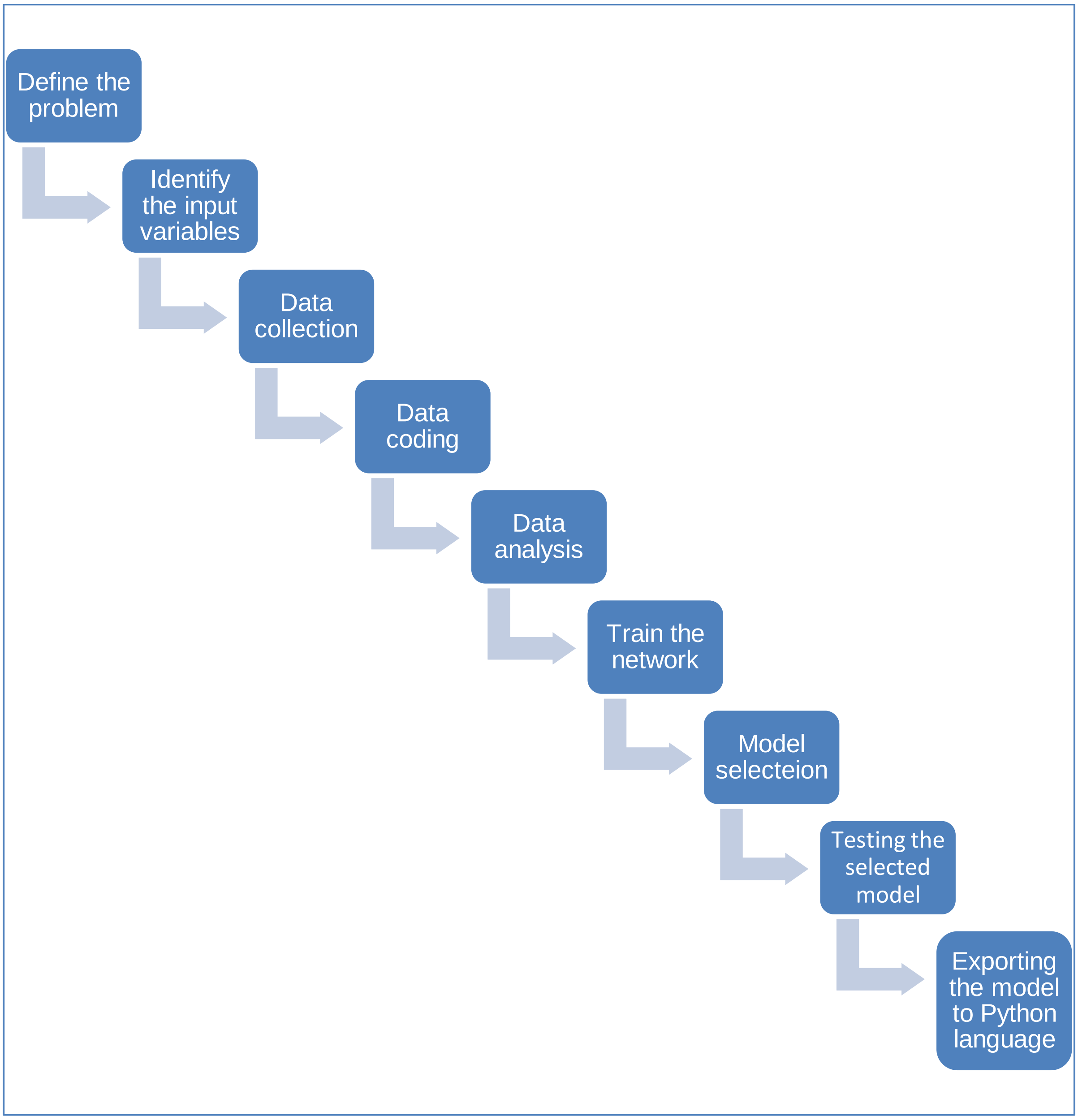

Figure 37. Steps of developing ANN model (Author)

5.2. Data encoding scheme

After data collection, this data needed to be formatted as ANN only deals with numeric form (Kshirsagar & Rathod, 2012). Table 3 explains the technique used to code the inputs to be readable by the software. All the project data representing the input variables were classified and coded as per this format.

Company category	
Category A	1
Category B	2
Project Budget	
≤ 30 (M.EGP.)	1
≤ 50 (M.EGP.)	2
≤ 100 (M.EGP.)	3
≤ 200 (M.EGP.)	4
≤ 350 (M.EGP.)	5
≤500 (M.EGP)	6
≤ One Billion EGP.	7
> One Billion E.G.P.	8
Project duration	
≤ 18	1
≤ 24	2
≤ 30	3
≤ 36	4
≤ 48	5
> 48	6
Project type	
Banks	1
Office buildings	2
Schools and Universities	3
Residential buildings	4
Hotels	5
Location	
City	1
Rural area	2
Nature of client	
Private	1
public identities	2
Type of contract	
Lumpsum	1
Cost plus	2
Unit rate	3

Table 3. Data Encoding scheme (Author)

5.3. Data set analysis using the software

The database that feeds into the excel document consisted of forty projects. Five projects were kept for testing. The columns in the table below represent the variables that are eight, and the rows represent the projects or instances.

	Company category	Project budget	Project duration	Project type	Location	Type of client	Type of Contract	%OH costs
1	1	4	6	1	2	1	3	12
2	1	4	5	1	2	1	3	11
...	...	...	...	...	...	...	...	...
40	1	8	6	5	1	1	3	11.5

Table 4. Data preview table (Author)

5.3.1. Variables table

Table 5 shows the variables are divided into seven independent input variables, and the output is the dependent variable. The variables are identified according to the data collection and the usage of each variable as shown in Table 6, so for variable 1 it represents the company category, variable 2 represents the project budget, variable 3 represents the project duration, variable 4 represents project type, variable 5 represents the location of the project if it is outside or inside the city, variable 6 represents the nature of the client if it private or public client, variable 7 represents the type of contract if it is lump sum, cost plus or unit rate and variable 8 represent the percentage of site OH costs.

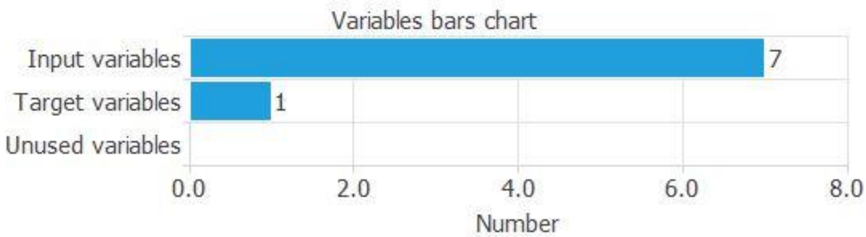

Table 5. Variables bar chart (Author)

	Name	Use
1	Company category	Input
2	Project budget	Input
3	Project duration	Input
4	Project type	Input
5	Location	Input
6	Type of client	Input
7	Type of Contract	Input
8	%OH costs	Target

Table 6. Variables table (Author)

The Neural Designer program automatically divides the dataset, as explained in chapter four. However, since the collected data is not much, so the number of training projects was increased, and the number of testing projects was minimized, as shown in Figure 38:

- Training 27 projects (67.5%)
- Selection 8 projects (20%)
- Testing 5 projects (12.5%)

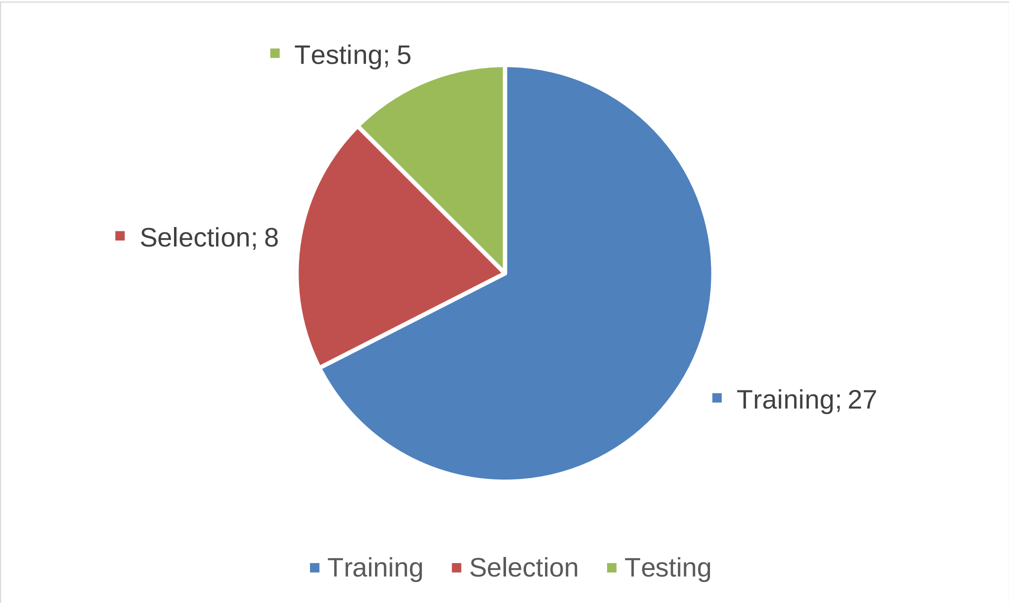

Figure 38. Data set allocation (Author)

5.3.2. Data set statistics

Basic statistics is highly needed to be checked before starting to design the model as it shows if there are any wrong input or information to be checked or corrected for every factor. Table 7 shows the basic statistics for each factor, which are minimum, maximum, mean, and standard deviations.

	Minimum	Maximum	Mean	Deviation
Company category	1.00	2.00	1.18	0.38
Project budget	1.00	8.00	4.03	2.13
Project duration	1.00	6.00	3.73	1.72
Project type	1.00	5.00	2.90	1.45
Location	1.00	2.00	1.55	0.50
Type of client	1.00	2.00	1.07	0.27
Type of Contract	1.00	3.00	2.70	0.61
%OH costs	6.00	13.00	10.03	1.66

Table 7. Data statistics (Author)

5.3.2.1. Company category

This factor was one of the main factors in the questionnaire as only the first and second categories were selected to proceed with their projects. The analysis shows a mean of 1.18, a standard deviation of 0.38, and it is classified into two categories, only first grade or second-grade construction company.

5.3.2.2. Project budget

The total project budget, which indicates the total contract amount, shows a mean of 4.03, a standard deviation of 2.13, and it is classified into eight different categories, which was explained earlier.

5.3.2.3. Project duration

The project duration shows a mean of 3.73 and a standard deviation of 1.72. The minimum is one, and the maximum is six, which indicates that there are six categories of the project duration, and the project will be allocated based on it is duration to one of these categories.

5.3.2.4. Project type

The project type factor, which could be residential, commercial, banks, office buildings, schools, or hotels, shows a mean of 2.90 and a standard deviation of 1.45. The minimum is one, and the maximum is five, which means that there are five different types of projects in the dataset.

5.3.2.5. Project location

This factor indicates if the project is in the city or a rural area, and it shows a mean of 1.55 and a standard deviation of 0.50. The minimum is one, and the maximum is two, which means there are only two categories of the location to be in the city or rural areas.

5.3.2.6. Nature of the client

This factor indicates if the client is private or public, and it shows a mean of 1.07 and a standard deviation of 0.27. The minimum is one, and the maximum is two, which means there are only two categories of clients, public or private.

5.3.2.7. Type of contract

The type of contract indicates a lump sum, cost-plus, or unit rate contract, and it shows a mean of 2.70 and a standard deviation of 0.61. The minimum and maximum indicate the number of categories of the types of contracts, which are three types.

5.3. Inputs correlation

Table 8 shows the inputs correlation's absolute values between the different variables or inputs. These absolute values are between 0 and 1, which indicates the relationship and strength between two variables.

- A correlation value near 1 indicates a strong relationship
- A correlation value near 0 indicates weak or no relationship

	Company category	Project budget	Project duration	Project type	Location	Type of client	Type of Contract
Company category	1	-0.55	-0.35	0.43	-0.11	-0.13	-0.28
Project budget		1	0.6	0.24	0.28	-0.045	0.12
Project duration			1	0.26	0.35	-0.14	-0.2
Project type				1	-0.17	-0.27	-0.3
Location					1	0.26	-0.23
Type of client						1	-0.017
Type of Contract							1

Table 8. Inputs correlation (Author)

5.4. Inputs-output correlation

Checking the correlations between a single input and the output can be useful. This function measures the coefficient of correlation values between all inputs and the output to show the dependencies and inputs strength on the output as shown in Figure 39 . However, in general, the output depends simultaneously on several inputs.

- A correlation value near 1 indicates a strong correlation.
- A correlation value near 0 indicates weak or no relationship.

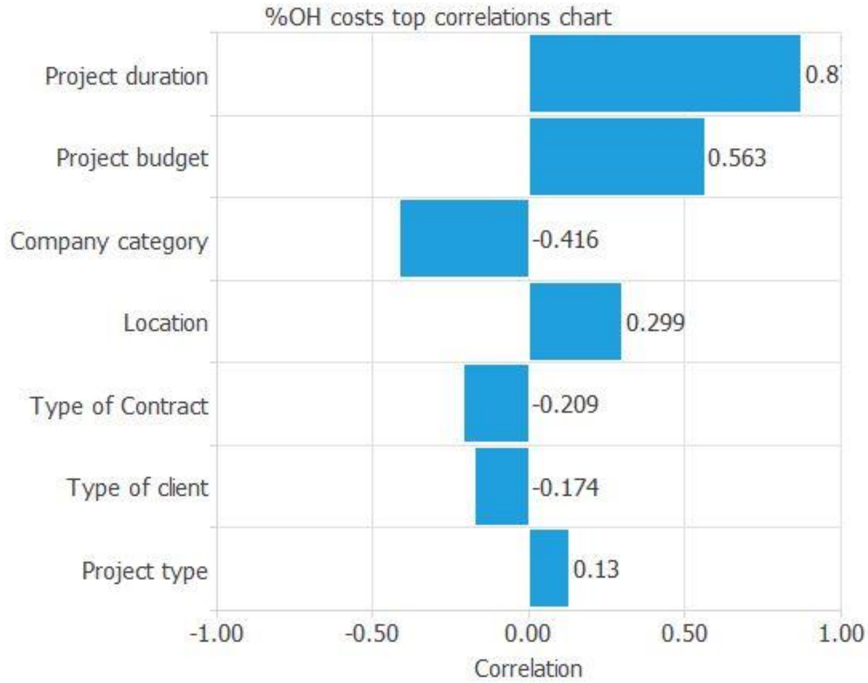

Figure 39. inputs-output correlation chart (Author)

Also, the software generates a table to summarize the impact of each factor on the output and identify the type of relationship between them, as shown in Table 9.

	type	%OH costs
Project duration	Linear	0.873329
Project budget	Linear	0.563073
Company category	Logistic	-0.415620
Location	Logistic	0.298518
Type of Contract	Linear	-0.208926
Type of client	Logistic	-0.173551
Project type	Linear	0.129509

Table 9. Inputs-output correlation (Author)

5.4.1. Project duration

The project duration was ranked the highest factor that influences the site OH costs with a correlation of 0.87, which proves a relationship directly proportional between the project duration and the site OH costs.

5.4.2. Project budget

The project budget was ranked the second-highest factor that influences the site OH costs with a correlation of 0.563, which shows that there is a relationship between the project budget or actual contract value and the percentage of site OH costs.

5.4.3. Company category

The company category was ranked the third-highest factor that influences the site OH costs with a correlation of -0.416, which shows a relationship between the company category and the percentage of site OH costs. The company category was limited to first and second-grade companies only.

5.4.4. Project Location

The project location influences the site OH costs with a correlation of 0.284, which shows a weak relationship between the project location and the percentage of site OH costs.

5.4.5. Contract type

The contract type influences the site OH costs with a correlation of -0.217, which shows that there is a weak relationship between the contract type and the percentage of site OH costs.

5.4.6. Client type

The client type influences the site OH costs with a correlation of -0.174, which shows that there is a weak relationship between the client nature and the percentage of site OH costs.

5.4.7. Project type

As the projects were classified into banks, office buildings, schools, residential, and hotels. The project type was ranked the lowest factors among the factors that influence the site OH costs with a correlation of 0.13, which is almost near zero. This indicates that the relationship between the project type and the percentage of site OH costs.

In conclusion, Based on the software analyzation to the inputs measure their influence on the output, it was clear that the project duration, project budget, and the company category are the main three factors that affect the site OH cost percentage in the construction projects in Egypt. However, the other factors, which are project location and contract, have a little influence on the site OH costs percentage. The type of client and project type have the lowest influence on the output and did not affect the site OH percentage.

5.5. Model selection

In this part, trial and error are used to train the network and select the best model. In each trial, the below needs to be adjusted in the perceptron layer:

- Number of layers
- Neurons number in each layer
- Determine the activation function

The best model was selected based on the lowest testing RMSE value and relative percentage error. The trial and error method included forty trails with the below parameters of the perceptron layer, as shown below. In each trial, the number of neurons changes.

- One hidden layer hyperbolic tangent activation function
- One hidden layer logistic activation
- Two hidden layers Logistic activation function for each
- Two hidden layers hyperbolic tangent activation function for each

5.5.1. One hidden layer hyperbolic tangent activation function

The first ten models with one hidden layer hyperbolic tangent activation function showed that RMSE (Figure 40) and Relative percentage error (Figure 41) change in a non-linear relationship with the change in the number of neurons. The model with the lowest errors in this trail had an RMSE value of 0.4766 and a Relative percentage error of 4.7358%, while the model with the highest errors had an RMSE value of 0.9438 and Relative percentage error of 7.7930%.

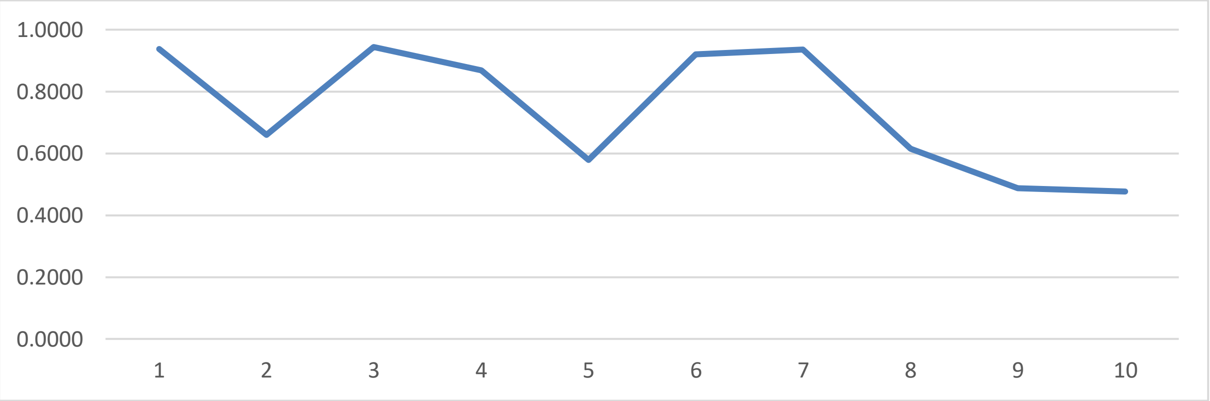

Figure 40. RMSE value for one hidden layer hyperbolic tangent (Author)

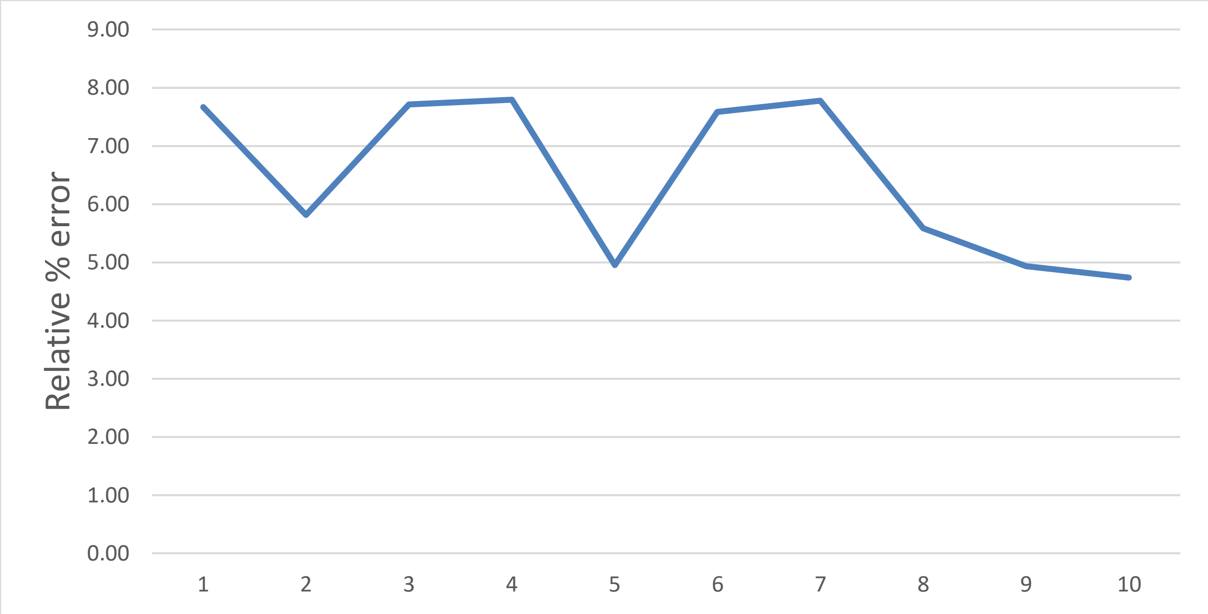

Figure 41. Relative percentage error for one hidden layer hyperbolic tangent (Author)

5.5.2. One hidden layer logistic activation

The models from 11 to 20 with one hidden layer logistic activation function showed that RMSE (Figure 42) and Relative percentage error (Figure 43) change in a non-linear relationship with the change in the number of neurons. The model with the lowest errors in this trail had an RMSE value of 0.5850 and a Relative percentage error of 4.8746%, while the model with the highest errors had an RMSE value of 1.1539 and Relative percentage error of 9.7397%.

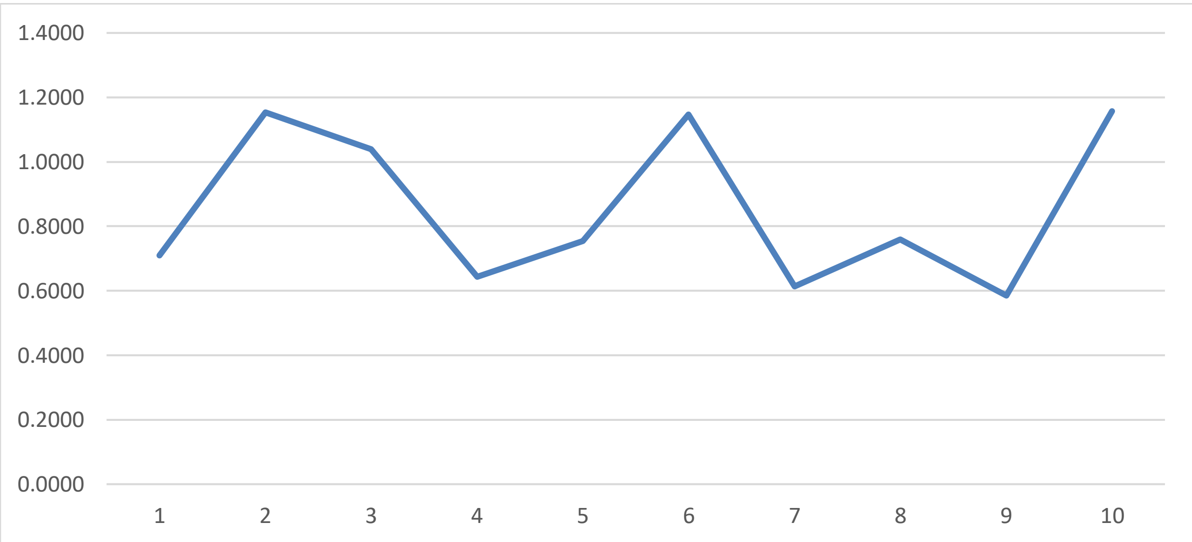

Figure 42. RMSE value One hidden layer logistic activation (Author)

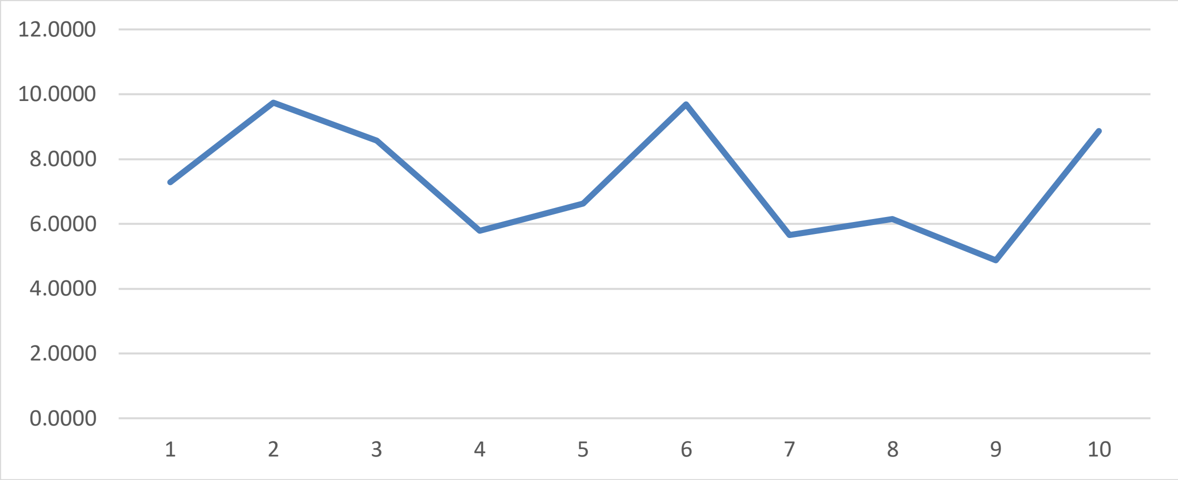

Figure 43. Relative Percentage error for One hidden layer logistic activation (Author)

5.5.3. Two hidden layers Logistic activation function for each

The models from 21 to 30 with two hidden layers logistic activation function showed that RMSE (Figure 44) and Relative percentage error (Figure 45) change in a non-linear relationship with the change in the number of neurons. The model with the lowest errors in this trail had an RMSE value of 0.2920 and a Relative percentage error of 2.9297%, while the model with the highest errors had an RMSE value of 1.4636 and a Relative percentage error of 9.8403%.

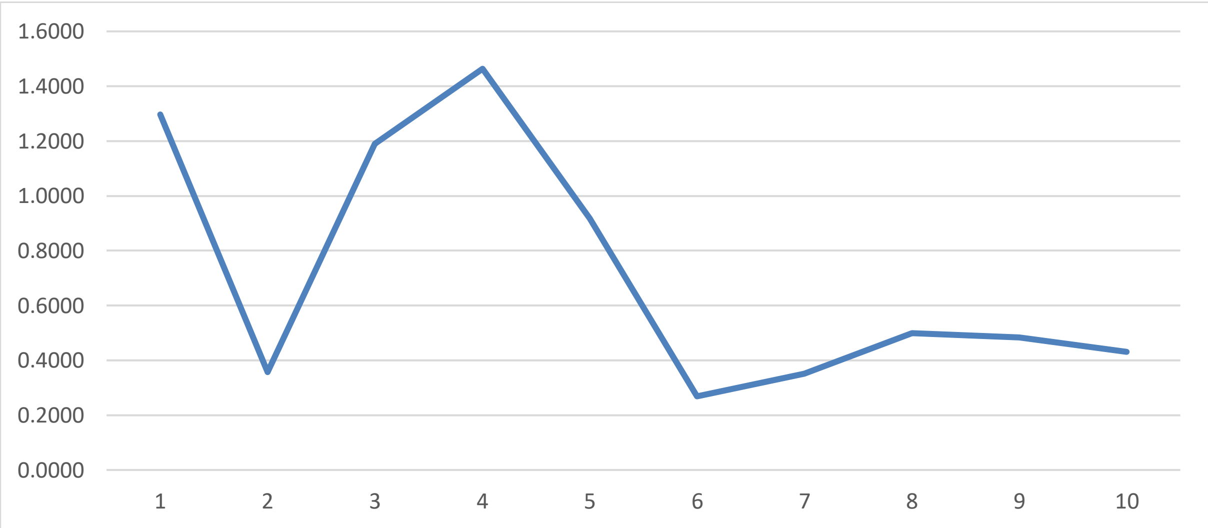

Figure 44. RMSE value for two hidden layers Logistic activation function (Author)

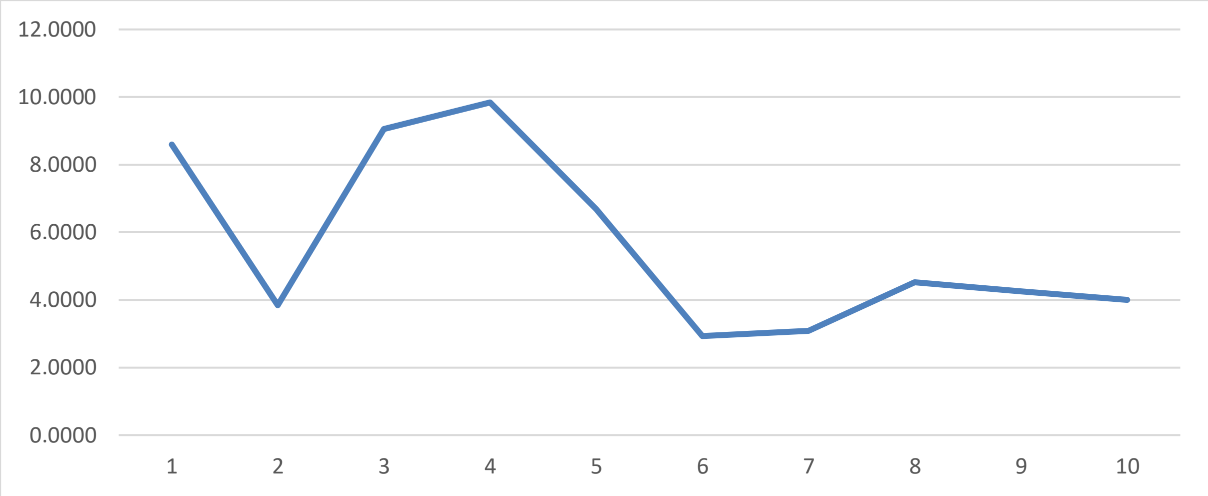

Figure 45. Relative percentage error for two hidden layers Logistic activation function (Author)

5.5.4. Two hidden layers hyperbolic tangent activation function for each

The models from 31 to 40 with two hidden layers hyperbolic tangent activation function for each showed that RMSE (Figure 46) and Relative percentage error (Figure 47) change in a non-linear relationship with the change in the number of neurons. The model with the lowest errors in this trail had an RMSE value of 0.4732 and a Relative percentage error of 4.1252%, while the model with the highest errors had an RMSE value of 1.2974 and a Relative percentage error of 10.720%.

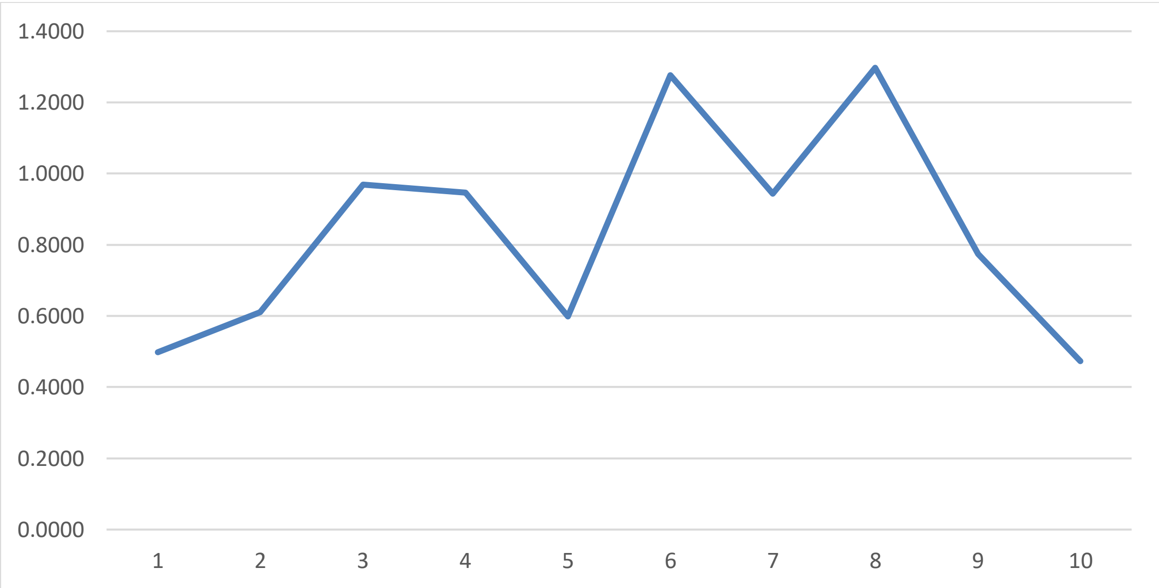

Figure 46. RMSE value for Two hidden layers hyperbolic tangent activation function (Author)

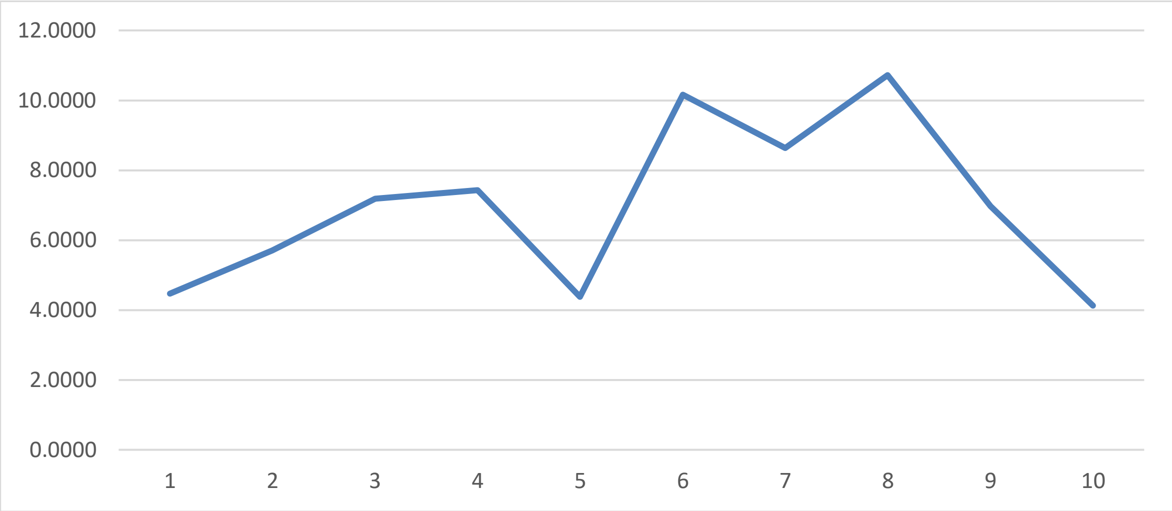

Figure 47. Relative percentage error for Two hidden layers hyperbolic tangent activation function (Author)

The best model that was selected based on the lowest RMSE value which is model 26 with two hidden layers logistic activation function, the first hidden layer has five neurons and the second hidden layer has three neurons as shown in Figure 48, which has the lowest RMSE for testing 0.2920 and a Relative percentage error of 2.9297%. The figure below shows the structure of the selected model. It consists of a scaling layer represented by yellow circles, perceptron layers represented in blue circles, and the unscaling layer represented in red circles.

- Input variables: 7 variables
- First hidden layer: 5 neurons
- Second hidden layer: 3 neurons
- Output layer: 1 neuron
- Activation function: Logistic sigmoid activation function
- RMSE: 0.2920
- Relative percentage error: 2.9297%

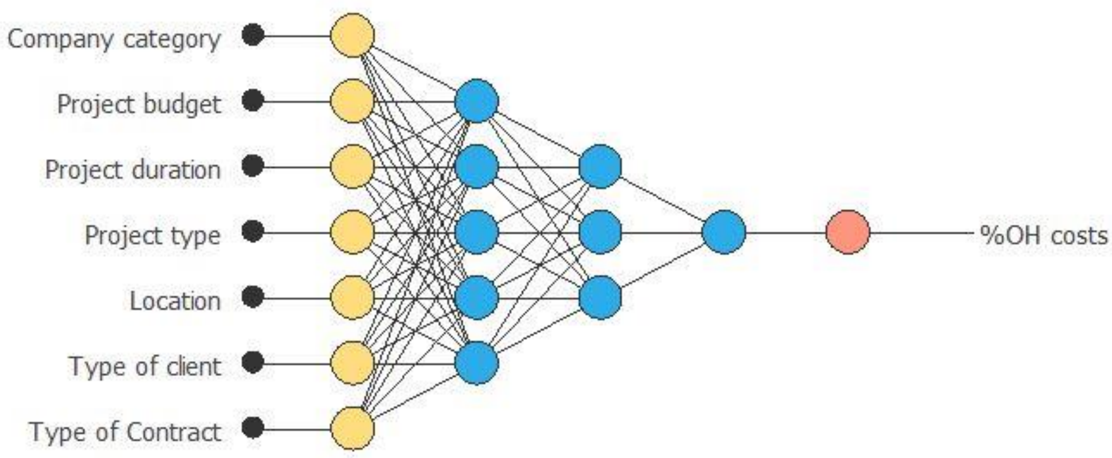

Figure 48. The selected model design (Author)

The linear regression analysis for the output, as shown in Figure 49, which is the forecasted of the percentage of site OH costs for the selected model 26 using the five projects that were classified for testing instances. From the linear regression analysis, the chart showed a strong correlation with R^2= 0.888.

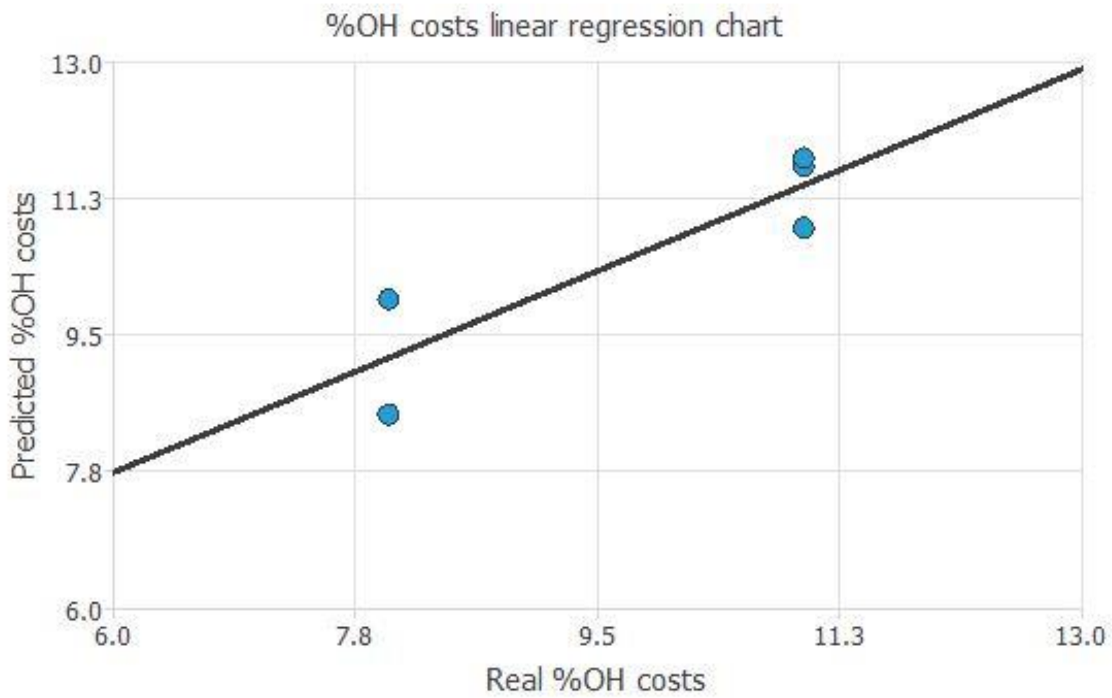

Figure 49. The selected model linear regression chart (Author)

5.6. Model deployment

The principle of machine learning implementation relates to the application of a prediction model for new data. It is necessary to organize and present the knowledge acquired during the development of a predictive model in a way that customers can use it. The Neural designer software can export the model into different expressions for further studies and understandings. The expression is as follows (Neural Designer, 2020):

- Python expression
- R expression
- Mathematical expression

In this research, the model will only be exported to Python language because it is the most famous and used language in machine learning.

5.6.1. Export to python

Python is one of today's most popular languages in programming and is considered the best AI and neural network programming language. It is an interpreted language, which means that the machine language instruction does not need to be compiled, and the developer can use the program directly to run it. This makes them sufficiently complete to allow the language to be controlled in the native machine learning language which is understood by the hardware (Cuelogic Technologies Pvt. Ltd., 2016).

It is a high-level language of programming that can be used for complex scenarios. Python deals with variables, objects, arrays, complex arithmetic, and other computer science concepts that increase its usability exponentially. Python is also a language of general programming that can be used in different domains and technologies. Python also has an automated memory management scheme of diverse styles supporting several programming paradigms, including imperative, functional, and object-oriented. Python has an open-source version called C-python, which is also very popular with all operating systems (Cuelogic Technologies Pvt. Ltd., 2016).

Deep learning plays a vital role in machine learning as the primary developmental feature of the neural network. It is the foundation of several technologies in various industries. Python can be used to build ANN for machine learning. Two libraries are available with Python for deep learning, which are Theano and Caffe (Magnimind , 2019).

In this research, the selected model was exported to python language and was viewed using Visual Studio 2019 software to view the file, as shown in Figure 50. The model in Python language was written as below:

```
#!/usr/bin/python

from math import exp
def Logistic(x) :
   return(1/(1+exp(-x)))

def expression(inputs) :

    if type(inputs) != list:
       print('Argument must be a list')
       exit()
    if len(inputs) != 7:
       print('Incorrect number of inputs')
       exit()
    Companycategory=inputs[0]
    Projectbudget=inputs[1]
    Projectduration=inputs[2]
    Projecttype=inputs[3]
    Location=inputs[4]
    Typeofclient=inputs[5]
    TypeofContract=inputs[6]
    scaled_Companycategory = (Companycategory-1.175)/0.384808
    scaled_Projectbudget = (Projectbudget-4.025)/2.13022
    scaled_Projectduration = 2*(Projectduration-1)/(6-1)-1
    scaled_Projecttype = (Projecttype-2.9)/1.44648
    scaled_Location = 2*(Location-1)/(2-1)-1
    scaled_Typeofclient = (Typeofclient-1.075)/0.266747
    scaled_TypeofContract = 2*(TypeofContract-1)/(3-1)-1
    y_1_1 = Logistic (0.648813+ (scaled_Companycategory*-0.098206)+ (scaled_Project-
budget*0.475306)+ (scaled_Projectduration*0.56265)+ (scaled_Projecttype*-1.42937)+
(scaled_Location*0.31473)+ (scaled_Typeofclient*1.76293)+ (scaled_TypeofCon-
tract*0.573361))
    y_1_2 = Logistic (0.421492+ (scaled_Companycategory*1.41238)+ (scaled_Project-
budget*-0.796516)+ (scaled_Projectduration*-0.417334)+ (scaled_Projecttype*-0.969011)+
(scaled_Location*2.023)+ (scaled_Typeofclient*-0.221975)+ (scaled_TypeofCon-
tract*2.01511))
    y_1_3 = Logistic (-0.60516+ (scaled_Companycategory*-1.74375)+ (scaled_Project-
budget*0.625127)+ (scaled_Projectduration*-1.41318)+ (scaled_Projecttype*-0.397128)+
(scaled_Location*1.50221)+ (scaled_Typeofclient*-1.67949)+ (scaled_TypeofCon-
tract*1.40095))
    y_1_4 = Logistic (-1.22115+ (scaled_Companycategory*-0.325518)+ (scaled_Project-
budget*-1.28677)+ (scaled_Projectduration*0.594703)+ (scaled_Projecttype*-0.246565)+
(scaled_Location*-1.07789)+ (scaled_Typeofclient*1.13003)+ (scaled_TypeofCon-
tract*0.21472))
    y_1_5 = Logistic (1.98621+ (scaled_Companycategory*-0.975683)+ (scaled_Project-
budget*0.881856)+ (scaled_Projectduration*2.49552)+ (scaled_Projecttype*1.67006)+
(scaled_Location*0.402731)+ (scaled_Typeofclient*0.888699)+ (scaled_TypeofCon-
tract*0.24032))
    y_2_1 = Logistic (-1.95225+ (y_1_1*-0.0336423)+ (y_1_2*0.243734)+
(y_1_3*0.0370916)+ (y_1_4*-1.25758)+ (y_1_5*0.390768))
    y_2_2 = Logistic (-0.233325+ (y_1_1*0.32628)+ (y_1_2*1.13534)+ (y_1_3*-0.36904)+
(y_1_4*1.28065)+ (y_1_5*-1.86298))
    y_2_3 = Logistic (1.25214+ (y_1_1*1.2102)+ (y_1_2*-0.517766)+ (y_1_3*0.316069)+
(y_1_4*0.0943698)+ (y_1_5*1.56534))
```

```
    scaled_%OHcosts =  (-0.928712+ (y_2_1*0.560853)+ (y_2_2*-1.26797)+
(y_2_3*1.73943))
    %OHcosts = (0.5*(scaled_%OHcosts+1.0)*(13-6)+6)

    return %OHcosts
```

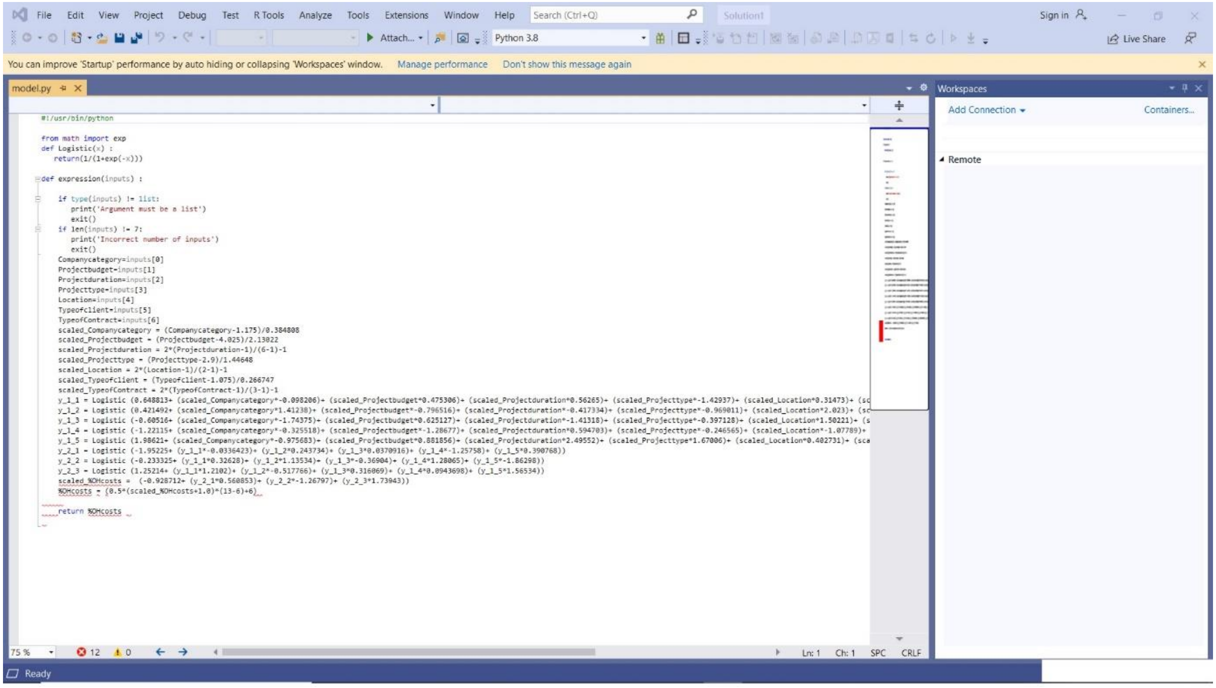

Figure 50. The selected model in Python language opened with Visual Studio 2019 software.

6. Conclusion and Recommendation

This research listed all the factors affecting the site OH costs and developed an ANN model to estimate the percentage of site OH costs for the construction projects in Egypt. This chapter will summarize the work done in the summary section. The conclusion section will provide answers to all the research questions and will list the outputs and results of the study. The recommendation section will present the author's point of view for future work.

6.1. Summary

The research included six chapters which covered all the information related to this topic. Chapter one introduced the topic, the importance of studying site OH costs, the importance of integrating AI in the construction industry, research objective, research methodology, scope, and limitations.

Chapter two discussed the site OH costs and ANN in literature, providing different research studies, articles, masters, and Ph.D. dissertations, also books that are related to the same topic. This chapter was divided into eight sections. The definitions of cost estimation, direct cost, indirect costs and general OH costs, types of general OH costs, factors affect the site OH costs, ANN, and finally, previous works related to the integration of ANN in solving the challenges in the construction industry.

Chapter three is for data collection and analysis. It explains the questionnaire design and the steps used for data collection. Forty projects were collected, and these projects were analyzed according to the factors listed in chapter two. This analysis illustrated the impact of each factor on the percentage of site OH costs.

Chapter four included a brief explanation of the content of ANN, which is application type, data set, network architecture, training strategy, model selection, and testing analysis, in addition to a guideline for the Neural designer software through listing all the steps which will be followed to develop the model.

Chapter five explains how the collected data was coded to be readable for the software using the data encoding scheme method, data analysis done by the Neural Designer software showing the correlation between the different factors, and the influence of each factor on the percentage of site OH costs. The trial and error method was used to select the best ANN model. Five projects were kept for testing the model and was exported it into python language and viewed by Visual Studio software.

6.2. Conclusion

The following conclusions and answers for the questions can be drawn from this research:

- Through the literature review, the direct costs are directly attributable to the implementation of specific project activity as labour costs, equipment, supplies, and materials. Indirect costs are often called OH costs, which are expenses that support the project and cannot be assigned to an activity.
- The site OH costs are the expenses related to site work and site accommodation or, in other words, costs of administration and managing a specific project. It includes site office rent, site office expenses, utilities, insurance, site trash removal, safety supplies, telephone bills, and other site expenses. The site OH costs are very important in the tendering phase as it mainly differs from a company from another when bidding for a project. Since it is affected by several factors, understating these factors reduces the risk and enhances the bid accuracy.
- Ten factors that affect the site OH costs in the construction projects were identified from previous studies. These ten factors were simplified into seven factors to ease the data collection process because of COVID-19, as not all the project's information was accessible. These factors are project duration, project budget, company category, project location, contract type, client type, and project type.
- The data collection result was forty construction projects in Egypt. The data analysis illustrated the impact of each factor on the percentage of site OH costs in Egypt's construction industry. The project duration and the project budget are the highest factors that affect the percentage of site OH costs, while the client type and project type are the lowest factors that affect the percentage of site OH costs in the construction projects.
- Data analysis was done again by the Neural designer software to prove the ANN model's efficiency in identifying the weight of each factor (inputs) on the

percentage of site OH costs (output). The highest two factors were project duration and project budget with correlations 0.86 and 0.56, respectively, which validate the results from the manual data analysis done earlier.

- The data analysis showed that there is no ideal percentage for the site OH costs in Egypt. Therefore, It was found that the ANN model is a useful tool to minimize the amount of work need to estimate the percentage of site OH costs with higher accuracy, so this research included an explanation of ANN components generally and a guideline for the usage of Neural Designer software.
- The forty projects were coded and used as a dataset to build the ANN model to predict the percentage of site OH costs.
- The dataset was trained, selected, and tested with different activation functions and different perception layers. Trial and error methods were done to select the best model with the minimum RMSE value using a backpropagation algorithm. The selected model is a logistic activation function with RMSE value 0.2920 and a Relative percentage error of 2.9297%. The model architecture consisted of seven input variables, which represent the factors that affect site OH cost, two hidden layers the first layer has five neurons and the second layer has three neurons, one output variable representing the percentage of site OH costs. The model used seven projects for testing, and the result of the linear regression analysis showed a high correlation coefficient $R^2 = 0.888$.
- The ANN model was written into the python language and viewed using Visual studio 2019 software.

6.3. Recommendations

According to the research findings and the conclusion discussed earlier, the following should be considered for further researches.

1. The model should be modified to include other types of construction projects as malls, sporting clubs, industrial and infrastructure projects.
2. The ANN models or any other type of model require a strong database of previous projects, which is almost not available in the construction companies in Egypt. Most of the companies in Egypt do not have a database to support researchers with the needed information. It is recommended to develop a robust

database by all the construction companies with information about previous projects as a first step in adopting new technologies.

3. ANN, therefore, has essential advantages that make it an effective tool to solve a large number of CM problems and challenges. It is necessary to experiment on several training algorithms, network architectures, and hybrid models for better performance of the model. There is still massive scope that still exists to be explored.
4. The usage of ANN models in construction management can be improved if specific guidelines are established through a deep evaluation of all the past works to determine the network control parameters, optimum learning algorithm, and network architecture, instead of using the trial and error traditional method.
5. In the future, It is recommended that these types of models should be done on a bigger scale with a big database to access the potential knowledge and capability of the ANN in solving the problems in the construction industry and will also increase the awareness among the professionals of the benefits of ANN.
6. Since it takes a lot of time and effort as detailed design to estimate a task in the construction project, it is necessary to implement the ANN model to ongoing projects to ease data analysis and provide accurate results with less effort and time.
7. The ANN models should be integrated with other models such as case-Based Reasoning, fuzzy logic, Ant colony optimization, particle swarm optimization, genetic algorithms, and Artificial Bee Colony for better results and enhance the accuracy of the output.

Declaration of Authorship

I hereby declare that the attached master's thesis was completed independently and without the prohibited assistance of third parties, and that no sources or assistance were used other than those listed. All passages whose content or wording originates from another publication have been marked as such. Neither this thesis nor any variant of it has previously been submitted to an examining authority or published.

Berlin, 29.10.2020

Location, Date

Sherif. Othman

Signature of the student

Appendix

Appendix A

Metropolia
University of Applied Sciences

htw
Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences

Questionnaire form Used for research only	Click or tap to enter a date. Form no.
1. Personal information	
Name: Click or tap here to enter text.	Position: Click or tap here to enter text.
Telephone No: Click or tap here to enter text.	Email: Click or tap here to enter text.
2. Organization Data	
The company name: Click or tap here to enter text.	
Category of the company: ☐ A ☐ B ☐ C ☐ D	
How many years of experience does the company have in building projects? ☐ >Less than 5 years ☐5-10 years ☐10-20 years ☐20-40 years ☐> 40 years	
3. Project data	
Project name: Click or tap here to enter text.	
Project location: Click or tap here to enter text.	

What is the project contract value in EGP?		
☐≤ 30 (M.EGP.)	☐≤ 50 (M.EGP.)	☐≤ 100 (M.EGP.)
☐≤ 200 (M.EGP.)	☐≤ 350 (M.EGP.)	☐≤ 500 (M.EGP.)
☐≤ One Billion EGP	☐> One Billion E.G.P.	

Percentage of Site Overhead Costs (%): Click or tap here to enter text.

Project duration in months

☐ <18 ☐18-24 ☐ 24-30 ☐ 30-36 ☐ 36-48

☐> 48

What is the project type?

☐Residential ☐Educational ☐Commercial ☐Hotels

☐Banks ☐Sporting clubs ☐Office buildings ☐Others,

What type of contract?

☐Lump sum ☐Cost plus ☐Unit rate ☐Others, ...

What is the owner's type?

☐Private ☐Public ☐Others, ...

Appendix B

#	Company category	Project budget	Project duration	Project type	Loca-tion	Type of client	Type of Contract	%OH costs
Project 1	1	4	6	1	2	1	3	12
Project 2	1	4	5	1	2	1	3	11
Project 3	1	4	5	1	2	1	3	11.5
Project 4	1	3	2	1	1	1	3	8
Project 5	1	1	2	1	1	1	3	9
Project 6	1	2	2	1	1	1	3	9.5
Project 7	1	4	5	1	2	1	3	11.5
Project 8	1	4	4	1	2	1	3	10
Project 9	1	4	4	1	2	2	3	10

Project 10	1	3	2	2	2	2	3	10
Project 11	1	5	1	2	2	1	3	9
Project 12	1	3	2	2	1	1	3	8.5
Project 13	1	4	1	2	1	1	3	8
Project 14	1	4	3	2	2	2	2	8.5
Project 15	1	3	2	2	2	1	3	8
Project 16	1	3	3	2	1	1	3	9.5
Project 17	1	7	5	2	1	1	3	11
Project 18	1	2	4	2	2	1	2	10.5
Project 19	1	8	6	3	2	1	3	11
Project 20	1	2	5	3	2	1	3	11
Project 21	1	6	4	3	2	1	2	10.5
Project 22	1	8	6	3	2	1	3	11.5
Project 23	1	4	1	3	1	1	3	10
Project 24	2	3	2	3	1	1	3	8.5
Project 25	2	4	4	4	2	1	1	10.5
Project 26	2	1	1	4	1	1	3	6
Project 27	2	1	1	4	1	1	3	6.5
Project 28	2	1	2	4	2	1	2	7
Project 29	2	1	2	4	1	1	3	7.5
Project 30	1	8	5	4	2	1	3	12
Project 31	1	3	6	4	1	1	1	13
Project 32	1	4	5	4	1	1	3	12
Project 33	2	2	5	4	2	1	1	11.5
Project 34	1	7	5	5	2	1	3	11
Project 35	1	6	5	5	1	1	2	11
Project 36	1	5	5	5	2	1	2	11
Project 37	1	4	4	5	1	1	3	10.5

Project 38	1	8	6	5	2	1	3	11
Project 39	1	3	5	5	1	1	3	11
Project 40	1	8	6	5	1	1	3	11.5

References

AACE International Recommended Practice, 2019. *Cost Engineering Terminology.* RP 10S-90 ed. London: AACE International.

Abdul-Malak, M. A., El-Saadi, M. M. H. & Abou-Zeid, M. G., 2002. Process Model for Administrating Construction Claims. *Journal of Management In Engineering,* 1 April, 18(2), pp. 84-94.

Aibinu, A. A., Dassanayake, D. & CHAU, T. V., 2011. *Use of Artifical Intelligence to predict the accuracy of pre-tender building cost estimate.* Amsterdam, The Netherlands, Conference: Management and Innovation for a Sustainable Built Environment.

Alaloul, W. S. et al., 2018. An Artificial neural networks (ANN) model for evaluating construction project performance based on coordination factors. *Cogent Engineering,* 5(1), pp. 1-18.

Alqahtani, A. & Whyte, A., 2013. Artificial neural networks incorporating cost significant Items towards enhancing estimation for (life-cycle) costing of construction projects. *Australasian Journal of Construction Economics and Building,* Sept., 13(3), pp. 51-64.

AL-Zwainy, F. M. S., Rasheed, H. A. & Ibraheem, H. F., 2012. Development of the construction productivity estimation model using artificial neural network for finishing works for floors with marble. *ARPN Journal of Engineering and Applied Sciences,* 7(6), pp. 714-722.

Arafa, M. & Alqedra, M., 2011. Early Stage Cost Estimation of Buildings Construction Projects using Artificial Neural Networks. *Journal of Artificial Intelligence,* 4(1), p. 63–75.

Assaf, S. A., Bubshait, A. A., Atiyah, S. & Al-Shahri, M., 2001. The management of construction company overhead costs. *International Journal of Project Management,* 19(5), pp. 295-303.

Aswed, G., 2016. Productivity Estimating Model for Bricklayer In Construction Projects Using Artificial Neural Networks. *Al-qadisiya Journal of Engineering Sciences,* 9(2), pp. 183-199.

Attal, A., 2010. *Development of Neural Network Models for Prediction of Highway Construction Cost and Project Duration,* Athens: The Russ College of Engineering and Technology of Ohio University.

Bhosale, A. & Konnur, B. A., 2019. Use of Artificial Neural Network in Construction Management. *International Research Journal of Engineering and Technology (IRJET),* 06(05), pp. 5905-5908.

Bluman, A. G., 2012. *Elementary statistics : a step by step approach.* 8th ed. New York: Mcgraw-Hill.

Brownlee, J., 2020. *Neural Networks are Function Approximation Algorithms.* [Online] Available at: https://machinelearningmastery.com/neural-networks-are-function-approximators/#:~:text=Neural%20networks%20are%20an%20example%20of%20a%20supervised%20learning%20algorithm,error%20during%20the%20training%20process.
[Accessed 14 August 2020].

Chan, C. T. W. & Pasquire, C., 2002. *Estimation of project overheads: a contractor's perspective.* University of Northumbria, Association of Researchers in Construction Management, pp. 53-62.

Chauhan, N. S., 2019. *Introduction to Artificial Neural Networks(ANN).* [Online] Available at: https://towardsdatascience.com/introduction-to-artificial-neural-networks-ann-1aea15775ef9
[Accessed 05 Aug. 2020].

Choon, T. T. & Ali, K. N., 2008. REVIEW OF POTENTIAL AREAS OF CONSTRUCTION COST ESTIMATING AND IDENTIFICATION OF RESEARCH GAPS. *Jurnal Alam Bina, Jilid,* 11(2), pp. 61-72.

Chou, J. S., 2012. Comparison of multilabel classification models to forecast project dispute resolutions. *Expert Systems with Applications,* 39(11), p. 10202–10211..

Consultants Estimating Manual Commonwealth of Massachusetts, 2006. *Consultants Estimating Maunal,* Massachusetts: Commonwealth of Massachusetts.

Cuelogic Technologies Pvt. Ltd., 2016. *Role of Python in Artificial Intelligence (AI).* [Online]

Available at: https://www.mybib.com/tools/harvard-referencing-generator [Accessed 26 August 2020].

Cunningham, T., 2013. *Factors Affecting The Cost of Building Work - An Overview,* Dublin: Dublin Institute of Technology.

Dagbui, D. A. & Smith, S., 2012. *Neural networks for modelling the final target cost of water projects.* Edinburgh, UK, Proceedings 28th Annual ARCOM Conference, pp. 307-316.

Dagostino, F. a. F. L., 2002. *Estimating In Building Construction.* 6th ed. Upper Saddle River, New Jersey: Prentice Hall.

Dormehl, L., 2019. *Digital trends.* [Online] Available at: https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/
[Accessed 02 September 2020].

Eksteen, B. & Rosenberg, D., 2002. *The management of overhead costs in Construction Companies.* University of Northumbria, Association of Researchers in Construction Management, pp. 13-22.

Elevado, K. J. T., Galupino, J. G. & Gallardo, R. S., 2018. Artificial Neural Network (ANN) modelling of concrete mixed with waste ceramic tiles and fly ash. *International Journal of GEOMATE,* 15(51), pp. 154-159.

Elhassan, S. M. M., Ghazali, Z. B. & Zawawi, N. A. W. A., 2012. Decision making framework for optimizing construction management objectives: A review. *Recent Advances in Engineering,* 1(1), pp. 169-173.

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

El-Sawalhi, N. I. & Shehatto, O., 2014. A Neural Network Model for Building Construction Projects Cost Estimating. *Korean Journal of Construction Engineering and Project Management,* 4(4), pp. 9-16.

El-Sawy, I. Y., Abdel Razek, M. E. & Mohammed, H. H., 2011. Assessment of Overhead Cost for Building Construction Projects. *International Journal of Computer Science Issues,* 8(3), pp. 273-283.

Emad, M. A., 2014. *Factors Affecting building Construction Project's Cost Estimating,* Cairo: Academia.edu.

Emerging professional‘s companion, 2013. *Doarch economics.* [Online] Available at: https://doarcheconomics2015.files.wordpress.com/2015/04/emerging-professionals-companion.pdf
[Accessed 07 September 2020].

Fatima, A., Prasad, B. S. R. K. & Sekhar, S. T., 2017. A Comprehensive Study of Dispute Resolution using Artificial Neural Network in Build Operate and Transfer (BOT) Project. *International Journal of Applied Engineering Research,* 12(24), pp. 14984-14988.

Glinskiy, V., Serga, L., Khvan, M. & Zaykov, K., 2016. *Fuzzy Neural Networks in the Assessment of Environmental Safety.* Novosibirsk, Procedia CIRP, pp. 615-619.

Goa, H., 2016. *Safety Evaluation on Building Construction Based on Hopfield Neural Network.* Jinan, 5th International Conference on Civil, Architectural and Hydraulic Engineering.

Goh, Y. M. & Chua, D., 2013. Neural network analysis of construction safety management systems: a case study in Singapore. *Construction Management and Economics,* 31(5), pp. 460-470.

Golizadeh, H., Sadeghifam, A. N., Aadal, H. & Majid, M. A., 2016. Automated tool for predicting duration of construction activities in tropical countries. *KSCE Journal of Civil Engineering,* 20(1), pp. 12-22.

Golnaraghi, S., Zangenehmadar, Z., Moselhi, O. & Alkass, S., 2019. Application of Artificial Neural Network(s) in Predicting Formwork Labour Productivity. *Advances in Civil Engineering,* pp. 1-11.

Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep learning. Cambridge.* Massachusetts: The Mit Press.

Guru99, 2020. *Back Propagation Neural Network: Explained With Simple Example.* [Online]
Available at: https://www.guru99.com/backpropogation-neural-network.html
[Accessed 02 Aug. 2020].

Heravi, G. & Eslamdoost, E., 2015. Applying Artificial Neural Networks for Measuring and Predicting Construction-Labor Productivity. *Journal of Construction Engineering and Management,* 141(10).

Hong, Y., Liao, H. & Jiang, Y., 2014. Construction Engineering Cost Evaluation Model and Application Based on RS-IPSO-BP Neural Network. *Journal of Computers,* 9(4), pp. 1020-1026.

Jain, M. & Pathak, K. K., 2014. Applications of Artificial Neural Network in Construction Engineering and Management-A Review. *International Journal of Engineering Technology, Management and Applied Sciences,* 2(3), pp. 134-142.

Janani, R., Rangarajan, P. T. & Yazhini, S., 2015. A Systematic Study On Site Overhead Costs In Construction Industry. *International Journal of Research in Engineering and Technology,* 04(10), pp. 149-151.

Jiang, Z. & Shi, H., 2016. The selection of firms based on productivity: Different roles of entry and overhead cost. *Economic Modelling,* Volume 54, pp. 537-544.

Kaushik, A., Soni, A. K. & Soni, R., 2013. A Simple Neural Network Approach to Software Cost Estimation. *Global Journal of Computer Science and Technology,* 13(1), pp. 23-30.

Kim, G. H., Shin, J. M., Kim, S. & Shin, Y., 2013. Comparison of School Building Construction Costs Estimation Methods Using Regression Analysis, Neural Network, and Support Vector Machine. *Journal of Building Construction and Planning Research,* 1(1), pp. 1-7.

Kshirsagar, A. P. & Rathod, M. N., 2012. Artificial Neural Network. *IJCA Proceedings on National Conference on Recent Trends in Computing,* NCRTC(2), pp. 12-16.

Kulkarni, P. S., Londhe, S. N. & Deo, M. C., 2017. Artificial Neural Networks for Construction Management. *Journal of Soft Computing in Civil Engineering,* 1(2), pp. 70-88.

Lazarevska, M., Knezevic, M., Cvetkovska, M. & Gavriloska, A. T., 2014. Application of Aritifical Neural Networks In Civil Engineering. *Tehnički vjesnik,* 21(6), pp. 1353-1359.

Lee, D., Kim, S. & Kim, S., 2016. Development of Hybrid Model for Estimating Construction Waste for Multifamily Residential Buildings Using Artificial Neural Networks and Ant Colony Optimization. *Sustainability,* 8(9), pp. 870-883.

Lesniak, A. & Juszczyk, M., 2018. Prediction of site overhead costs with the use of artificial neural network based model. *Archives of Civil and Mechanical Engineering,* 18(3), pp. 973-982.

Lhee, S. C., Flood, I. & Issa, R. R. A., 2014. Development of a two-step neural network-based model to predict construction cost contingency. *Electronic Journal of Information Technology in Construction,* 19(24), pp. 399-411.

Liu, J. & Guo, F., 2014. Construction quality risk management of projects on the basis of rough set and neural network. *Computer Modelling and New Technologies,* 18(11), pp. 791-797.

Li, W. & Shi, H., 2010. Integration of Fuzzy Logic, Particle Swarm Optimization and Neural Networks in Quality Assessment of Construction Project. *Journal of Software,* 5(7), pp. 737-744.

Lorman, 2014. *Current Issues to watch for in construction claims, part III: Overhead..* [Online]
Available at: https://www.lorman.com/resources/current-issues-to-watch-for-in-construction-claims-part-iii-overhead-claims-15877
[Accessed 07 Aug. 2020].

Lowe, S., Bielek, R. & Burnham, R., 2003. *Compensation for Contractors' Home Office Overhead,* Washington, D.C: National Cooperative Highway Research Program.

Luu, T. V. & Kim, S. Y., 2009. Neural Network Model for Construction Cost Prediction of Apartment Projects in Vietnam. *Korean Journal of Construction Engineering and Management,* 10(3), pp. 139-147.

Lyne, C. & Maximino, C., 2014. *An Artificial Neural Network Approach to Structural Cost Estimation of Building Projects in the Philippines.* Manila, DLSU Research Congress.

Maghrebi, M., Sammut, C. & Waller, T. S., 2014. Predicting the Duration of Concrete Operations Via Artificial Neural Network and by Focusing on Supply Chain Parameters. *Building Research Journal,* 61(1), pp. 1-14.

Magnimind , 2019. *Why Learning Python Is Important For Machine Learning Aspirants?.* [Online]
Available at: https://becominghuman.ai/why-learning-python-is-important-for-machine-learning-aspirants-a97c5ec0629a
[Accessed 26 August 2020].

Mark Saunders, P. L. A. T., 2009. *Research methods for business students.* 5 ed. Harlow: Pearson Education Limited.

Minli, Z. & Shanshan, Q., 2012. Research on the Application of Artificial Neural Networks in Tender Offer for Construction Projects. *Physics Procedia,* Volume 24, pp. 1781-1788.

Mirahadi, F. & Zayed, T., 2016. Simulation-based construction productivity forecast using Neural-Network-Driven Fuzzy Reasoning. *Automation in Construction,* Volume 65, pp. 102-115.

Muqeem, S. et al., 2011. Construction Labor Production Rates Modeling Using Artificial Neural Network. *Journal of Information Technology in Construction (ITcon),* Volume 16, pp. 713-723.

Mwiya, B., Muya, M., Kaliba, C. & Mukalula , P., 2015. Construction Unit Rate Factor Modelling Using Neural Networks. *International Journal of Civil and Environmental Engineering,* 9(1), pp. 29-34.

Nagappan, K., Kanmani, S. & Uthariaraj, V. R., 2014. ANN models optimized using swarm intelligence algorithms. *WSEAS Transactions on Computers,* WSEAS Transactions on Computers, Volume 501-519, p. 2014.

Nani, G., Mensah, I. & Kumi, T. A., 2017. Duration estimation model for bridge construction projects in Ghana. *Journal of Engineering, Design and Technology,* 15(6), pp. 754-777.

Neil, S. P. & Hashemi, M. R., 2018. Chapter 8 - Ocean Modelling for Resource Characterization. In: S. P. Neil & M. R. Hashemi, eds. *Fundamentals of Ocean Renewable Energy.* E-Business Solutions ed. s.l.:Elsevier Ltd., pp. 193-235.

Neural Designer, 2020. *What is Neural Designer?.* [Online]
Available at: https://www.neuraldesigner.com/learning/user-guide/what-is-neural-

designer
[Accessed 23 August 2020].

Oduyemi, O., Okoroh, M. & Dean, A., 2015. *Developing An Artificial Neural Network Model For Life Cycle Costing In Buildings.* Lincoln, UK, Procs 31st Annual ARCOM Conference, pp. 843-852.

Patel, D. A. & Jha, K. N., 2015. *Neural Network Approach for Safety Climate Prediction.* Atlanta, USA, International Conference on Construction Research Congress Georgia Institute of Technology, pp. 1861-1870.

Patil, S. S. & Bhangale, P. P., 2014. Overhead Cost In Construction Industry. *International Journal of IndustrialEngineering & Technology (IJIET),* 4(2), pp. 1-6.

Pengcheng, X. & Kun, L., 2012. The evaluation for the behavioural risk of participants in construction projects based on BP Neural Network. *Advances in information Sciences and Service Sciences(AISS),* Aug., 4(14), pp. 97-107.

Petroutsatou, K., Georgopoulos, E., Lambropoulos, S. & Pantouvakis, J. P., 2012. Early Cost Estimating of Road Tunnel Construction Using Neural Networks. *Journal of Construction Engineering and Management,* 138(6), p. 679–687.

Peurifoy, R. a. O. G., 2014. *Estimating construction costs.* 6th ed. New York ; London: Mcgraw-Hill.

Polat, G., 2012. ANN approach to determine cost contingency in international construction project. *Journal of Applied Management and Investments,* 1(2), pp. 195-201.

Pratt, D., 2011. *Fundamentals of construction estimating.* 3rd ed. Clifton Park, NY: Delmar Cengage Learning.

Raghed, A., Hanafy, H. & Tarek, A., 2018. Developing an ANN model to suggests the optimum typology for effectively managing construction projects. *International Journal of Advanced Research,* 6(7), pp. 93-97.

Ruf, H. B. & Ruf, S., 2007. *Documentation and presenting cost in underground construction claims.* San Diego, California, No-Dig Conference & Exhibition.

Sharmila, S., Lekha, G. & Kaushik, S., 2016. Cost and Time Effective Prediction of Soil Characteristics using ANN Model. *International Journal of Innovative Research in Science, Engineering and Technology,* 5(3), pp. 3829-3834.

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

Shrestha, K. K. & Shrestha, P. P., 2016. A Contingency Cost Estimation System for Road Maintenance Contracts. *Procedia Engineering,* Volume 145, pp. 128-135.

Stolz, J. M., 2010. *Overhead and Uncertainty in Cost Estimates: A Guide to Their Review,* San Francisco, CA: Jacobs Associates.

Tatari, O. & Kucukvar, M., 2011. Cost premium prediction of certified green buildings: A neural network approach. *Building and Environment,* 46(5), pp. 1801-1086.

Techopedia, 2019. *What is an Artificial Neural Network (ANN)?.* [Online] Available at: https://www.techopedia.com/definition/5967/artificial-neural-network-ann [Accessed 02 Aug. 2020].

The Chartered Institute of Building, 2009. *Code of estimating practice.* 7th ed. Chichester: Wiley-Blackwell.

The Institute of Cost Accountants of India, 2013. *Guidance Note,* New Delhi: The institute of Cost Accountants of India.

The project Management Institute, Inc., 2013. *A Guide to the Project Management Body of Knowledge (PMBOK Guide).* 5th ed. Pennsylvania: Project Management Institute, Inc..

Vahdani, B., Mousavi, S. M., Mousakhani, M. & Sharifid, M., 2012. A Neural Network Model Based on Support Vector Machine for Conceptual Cost Estimation in Construction Projects. *Journal of Optimization in Industrial Engineering,* Volume 10, pp. 11-18.

Wang, Y. R., Yu, C. Y. & Chan, H. H., 2012. Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models. *International Journal of Project Management,* Volume 30, pp. 470-478.

Waziri, B. S., Bala, K. & Bustani, S. A., 2017. Artificial Neural Networks in Construction Engineering and Management. *International Journal of Architecture, Engineering and Construction,* 6(1), pp. 50-60.

Yadav, R., Vyas, V., Vyas, M. & Agrawal, S., 2016. Cost Estimation Model (Cem) for Residential Building using Artificial Neural Network. *International Journal of Engineering Research & Technology (IJERT),* 5(1), pp. 430-432.

Yitmen, I. & Soujeri, E., 2010. *An artificial neural network model for estimating the influence ofchange orders on project performance and dispute resolution.* Nottingham, The International Conference on Computing in Civil and Building Engineering.

Zack, J. G., 2002. *Calculation and Recovery of Home/Head Office Overhead.* Melbourne, Fluor.

Zheng, W. X., Chen, D. X. & Yan, L. J., 2010. Application of Neural Network in the Cost Estimation of Highway Engineering. *JOURNAL OF COMPUTERS,* 5(11), pp. 1762-1766.