Arab American University

Faculty of Graduate Studies

Prediction of Compressive Strength of Concrete in Palestinian Governorates Using Machine Learning Techniques

By

Abedelkareem Riad Kamel Zidan

Supervisor

Prof. Mohammed Awad

This Thesis was Submitted in Partial Fulfillment of the Requirements for the Master's degree in Computer Science.

June / 2021

Prediction of Compressive Strength of Concrete in Palestinian Governorates Using Machine Learning Techniques

By

Abedelkareem Riad Kamel Zidan

This thesis was defended successfully on **01 /June/2021** and approved by:

Committee Members	Signature
1. Supervisor: **Prof. Dr. Mohammed Awad**	....
2. Internal Examiner: **Dr. Mujahed Eleyat**	
3. External Examiner: **Dr. Yousef Daraghmeh**	....

Declaration

I declare that this thesis entitled "Prediction of Compressive Strength of Concrete in Palestinian Governorates Using Machine Learning Techniques" is my work and has been composed solely by myself and does not contain work from others researcher and has not been submitted for any other degree or scientific except the reference is made.

Dedication

I dedicate this thesis to my family and friends for their unconditional love and support.

To my mother and my father, Mr. Riad Zidan, for their support, which has not left me throughout my life.

And for my brothers and sisters, whose support I have not always forgotten.

To my dear brother Kareem Khaleel, for his continuous support during my life.

.

Acknowledgments

I would like to take this opportunity to express my deep regards to Prof. Mohammed Awad for his advice, support, and time that he has spent reviewing my work. Prof. Mohammed provided me valuable suggestions that have had a significant impact and helped in overcoming many obstacles in writing this thesis in the best way.

I would also like to take the opportunity to express my sincere respect and greetings to all Palestinian concrete factories and laboratories for assisting me in the process of collecting data on the compressive strength of Palestinian concrete.

I would also like to thank everyone who has supported me throughout my career, even if with a kind word.

Abstract

In this thesis, Machine Learning (ML) techniques were used to predict the compressive strength of concrete in the Palestinian governorates. The datasets were collected from Palestinian laboratories and factories from seven Palestinian governorates, which consists of five subsets, and each sub dataset is related to a specific type of Palestinian concrete.

The thesis work is divided into three phases: In the first phase, the process is divided into two parts, firstly; the implementation of clustering algorithms to the whole data of the Palestinian governorates. Secondly; implementation of clustering algorithms to each sub dataset that presents data in each governorate. The factors determining results showed that the Expectation-Maximization (EM) algorithm is completely identical to the Kohonen Self-Organizing Maps (KSOM) algorithm. The results from these two algorithms are similar, thus these two algorithms were used to determine the main factors that affect the concrete compressive strength (PCCS). The results obtained by using K-mean clustering algorithms show that they are more accurate prediction for improving the concrete compressive strength.

The second part is the use of ML techniques to classify the compressive strength of concrete, where three methods were used: MLPNNs, Support Vector Machine (SVM), and Ensemble Algorithm. The accuracy results were 93.5%, 80.4% and 90.2% respectively for B200 concrete, and the classification results for B250 concrete were 90.0%, 66.5% and 75.5% respectively. For B300 concrete, the classification results were 93.3%, 68.3% and 79.2% respectively.

The classification results were 90.6%, 83.3% and 85.6% respectively for B350 concrete, and the classification results were 90.0%, 80.6% and 78.6% respectively for B400 concrete.

The results showed that the MLPNNs using Levenberg–Marquardt algorithm are the most accurate for each type of concrete.

The classification models were applied on the dataset which was collected from Palestinian governorates laboratories after it removes other parameters and remains only factors that affect Palestinian Concrete Compressive Strength (PCCS) obtained from clustering algorithms. The new dataset was implemented on the classification models like MLPNNs, linear support vector machine, and Ensemble algorithm show the results are close to those obtained previous experiences that were implemented on pervious datasets and the accuracy results for the new dataset were 92.5%, 75.4% and 88.0% respectively.

The final part depends on the use of machine learning techniques to predict the compressive strength of concrete using three different Artificial Neural Networks (ANNs) techniques; Multilayer Perceptron Neural Networks (MLPNNs), Radial Basis Function Neural Networks (RBFNNs), and Recurrent Neural Networks (RNNs). It is found that the ANNs Techniques are effective tools for predicting the Compressive Strength of concrete. The mean square error (MSE) results were obtained from these ANNs models were 0.0107, 0.0064, and 0.0012 respectively where the MLPNNs using Levenberg–Marquardt model produce the best prediction result.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

KM	K-Means Clustering Algorithm
KSOM	Kohonen self-organizing map
EM	Expectation Maximization
ANNs	Artificial Neural Networks
NNs	Neural Networks
SVM	Support Vector Machine
CNNs	Convolutional Neural Networks
TP	True Positive
FP	False-Positive
FN	False-Negative
TN	True Negative
CCS	Concrete Compressive Strength
PCCS	Palestinian Concrete Compressive Strength
ML	Machine Learning
MLT	Machine Learning Techniques
MFFNNs	Multilayer Feed Forward Neural Networks
BP	Back Propagation
GP	Genetic Programming
ANNs	Artificial Neural Networks
LM	Levenberg-Marquardt
ES	Ensemble
MSE	Mean Squared Error

1.1 Introduction

1.1.1 Concrete Introduction (CCS)

Concrete is the most widely used material in the construction industry due to its durability and resistance over time. Its manufacture is carried out by mixing basic components: water, cement, fine aggregates, and coarse aggregates. But the problem is not so simple since the proportions of these components, as well as the inclusion of additives and other factors, will determine the resistance of this material. So, the correct choice of the dosing method directly influences this property. The resistance of concrete depends on the pressure as a primary criterion for the success or failure of the concrete mix design [1].

Concrete construction includes many raw materials, the most important components are fine aggregates, coarse aggregates, cement, water, and other additives [2]. Concrete must be designed adequately proportioned to give the properties of strength, homogeneity, uniformity, impermeability, durability, and others. Concrete is subject to engineering standards to ensure its quality, as the original when the building is to examine samples of concrete mixtures that provide values about suitability and validity for the project. Because it is a designed material, it is necessary to qualify and measure these properties to know the behavior of the concrete. The most important property of concrete is compressive strength (CS), which is "The maximum load for a unit area supported by a sample before failing by compression [3]. Considering that the compressive strength is ten times its resistance to tensile strength [4], so the goal of using concrete in projects is to take advantage of its distinctive resistance to pressure, which is determined experimentally by testing cylindrical samples at the age of 3, 7, 28 days.

1.1.2 Concrete Compressive Strength (CCS) Introduction)

The Compressive Strength (CS) is the most important property of hardened concrete at all, it reflects the degree of its quality and suitability. Other properties of concrete such as tensile, bending, shearing, and cohesion with the rebar are improved and increased by increasing the Compressive Strength and vice versa. Therefore, a Compressive Strength test is conducted to control the quality of concrete production at the project site. This test is also used for structural design purposes in order to determine the characteristic resistance of concrete at a pressure that is taken as a percentage of the maximum Compressive Strength. The pressure test is also useful in determining the validity of the aggregates and mixing water to identify the effect of impurities that may be found in them on the Compressive Strength of concrete. Currently, the Compressive Strength of conventional concrete ranges from 200-450 kg / cm^2. As for special installations and prefabricated units, the Compressive Strength is more than that and reaches 500 kg / cm^2. The pre-stressed concrete units should have a Compressive Strength of more than 400 kg / cm^2. It may reach 600 kg / cm^2 [5].

When the sample is successful, it is worked on in large quantities but if it is unsuccessful, the original is modified and re-examined and if it does not work again, it is disposed of successful concrete is characterized by strength. Concrete is exposed to many factors that affect its strength and suitability for the long term, and these factors are pressure, and weather factors from heat and moisture. The quality of concrete affects its durability and its resistance to weather factors [6]. Concrete is examined several days after it has been manufactured, and it may spend 28 days to verify its strength before proceeding to make large quantities of the same concrete mixture. Concrete Compressive strength is examined in three stages. The initial examination begins after 3 days which should achieve a strength of 45%, the second examination begins after 7 days which should achieve a strength of 67%, and the final examination will be 28 days after its preparation must achieve strength that is between 90% - 100%.

The CS through the laboratory requires time to reach its maximum resistance, therefore, the results of its tests are far from being immediate, and so, it is important to have a compression resistance prediction model through AI methods. Nowadays, the prediction that depends on machine learning techniques affects a wide sector in our life. The strength of concrete can be measured only in special laboratories. The use of ML methods can predict the strength of concrete pressure by using the concrete mixture as input data for the ML methods to predict the compressive strength of concrete.

This study applied in the Palestinian governorates using Machine Learning, classifications are made for the Compressive Strength of concrete in Palestinian governorates. The results showed that the difference between concrete mixture parameters in the governorates by detecting the main parameters that affect Compressive Strength of concrete. The study presents an application of the Machine Learning Technique (MLT) for prediction concrete Compressive Strength, as training and testing samples of the network were taken from the archive of laboratory experiments that were previously. After that the results of the test samples resulting from the network were compared with the real laboratory results on one hand and between the calculated values in a theoretical manner on the other hand, and a match appeared. Among the results, the effectiveness of the artificial neural networks in estimating the compressive strength of the concrete was proven despite the complexity and incompleteness of the available samples information, meaning that the trained network can be used by the designer of concrete mixtures to estimate the compressive strength of the concrete and to improve it, if necessary, by adjusting the ratios of the materials included in the mixture.

1.2 Objective

In this thesis, the main direct parameters that affect the concrete compressive strength will be identified. It is expected that results will vary depending on the quantities that are mixed, and clustering algorithms will be used for selecting the most important parameters that affect the CCS. On the other hand, some machine learning techniques will be used for the prediction and classification of the Compressive Strength of concrete using datasets collected from CCS Labs. Various ML models such as MLPNNs, RBFNNs, SVM, RNNs, Ensemble will be used. The process of predicting and classifying concrete compressive strength using AI techniques saves time, cost, and effort, instead of waiting for 28 days to find out the Compressive Strength of concrete.

The research objectives can be summarized in the following points

- Applying different clustering models to select the most important features that affect the Compressive Strength of concrete. These techniques enable us to know the factors affecting the Compressive Strength of concrete.
- Appling different ML techniques to classify the CCS in each type and each region by governorates, depending on the measures are used in Palestine.
- Applying different ANNs models to estimate the Compressive Strength of concrete based on a certain number of laboratory data that uses the largest possible number of factors affecting the Compressive Strength of concrete which was designed and broken by the civil testers.

1.3 Contribution

In this thesis, datasets of the Compressive Strength of concrete from Palestinian Governorates have been collected. Different Clustering Algorithms are used and applied to select the most important parameters that affect the CCS, such Expectation Maximization (EM), Kohonen Self-Organizing Map (KSOM), and K- Means Clustering Algorithm (KM) were applied. Clustering algorithms selected 4 important parameters from the 8 inputs parameters achieving an accuracy value that is close to the accuracy produced by the 8 inputs. Also, different AI models for the classification and prediction of CCS were applied. This is the first research that depends on AI models to predict and classify the CCS in Palestine. Different ML models that classify the CCS in each Governorate as MLPNNs, SVM, and Ensemble models were applied. On other hand, different models of ANNs were used to predict the CCS including; MLPNNs with Levenberg – Marquardt (LM) learning algorithm, RBFNNs, and RNNs. The ANNs architecture consist of eight inputs: coarse aggregate, fine aggregate, cement, water, w/c ratio, age, location, and super plasticizer, hidden layers with an incremental number of neurons, and an output layer that present the CCS. Finally, in this thesis, AI models that can predict and classify the CCS depending on the input values were presented, which means that no need to wait for 28 days to know the real test from labs, it can be observed that there is a strong relationship between the experimental results and those proposed by the model.

1.4 Overview

The remainder of this thesis is arranged as the following. Chapter 2 present a background that includes the description of the Palestinian Governorates concrete Compressive Strength dataset for prediction and classification. It also has another dataset for detecting main factors that affect Palestinian Governorates concrete Compressive Strength for each of Palestinian Governorates. A literature review of the related work in concrete Compressive Strength applications, and the techniques used to prediction and classification of the concrete Compressive Strength. In Chapter 3, has a description of the preprocessing phases which includes the feature selection and data normalization. The clustering algorithms, EM, KSOM, and KM were explained in order to detect the main factors that affect the Palestinian Governorate's Concrete Compressive Strength. SVM and Ensemble classifiers have been presented to make classifications. RBFNNs, RNNs, and MLPNNs models were explained to predict the concrete Compressive Strength in Palestinian Governorates. Chapter 4 shows all results obtained from all models in clustering, classification, and prediction phases. In Chapter 5, the conclusion and future work are presented.

Chapter 2

Background

2 Background

Concrete is one of the materials that form the basic building block in Palestine, and concrete Compressive Strength is one of the factors affecting the success of construction in general and concrete in particular. In the construction industry, checking the quality of concrete used on site is a daily and obligatory task. To fulfill this task, different tests are carried out that allow quality assurance by determining some properties such as resistance compression. These tests occur both in the fresh state of the concrete and in its hardened state. It should be mentioned that the most important property of concrete is resistant to compression since structural designs are made with the value of this property. Therefore, the most important test is the breakage of specimens, which measures the compressive strength [8]. The process consists of taking standardized samples according to the ASTM standard which are test tubes of (diameter by height) 10 cm. x 20 cm. or 15 cm. x 30 cm, and it will be tested on 3, 7, and 28 days [9]. This process is long and sometimes there is a need to know the results as soon as possible in order to take action. Therefore, according to the international literature, attempts have been made to carry out more rapid tests. Thus, this work seeks to predict and classify the Compressive Strength of concrete with a level of confidence sufficient using AI models.

2.1 Datasets Description

Initially, Palestinian laboratories and concrete factories were contacted to collect the necessary data for the study in general in the Palestinian governorates and cooperation took place between them. The necessary data were collected from three years ago, that is, in the period from 2017 until 2020. The necessary data that were taken consists of some characteristics of concrete, and it has been filtered in order to take only the CCS. Samples were collected from concrete laboratories and factories in seven Palestinian governorates of Jenin, Ramallah, Tulkarm, Salfit, Hebron, Nablus, and Tubas.

The work is divided into three parts, determination of factors affecting CCS, prediction of CCS, and classification of CCS in each concert strength type. The datasets that were used in the process of predicting the Compressive Strength of concrete consist of 715 samples (mixture) taken from Palestinian laboratories with 100 samples from each governorate which consist of eight inputs: coarse aggregate, fine aggregate, cement, water, w/c ratio, age, location, and super plasticizer. Also, they consist of one output which is the Compressive Strength of concrete. Table 2.1 shows the consists of the ranges of each factor influencing concrete Compressive Strength and Concrete Compressive strength ranges.

Table 2. 1: The Palestinian concrete dataset and their valid ranges.

Feature name	Range
Coarse Aggregate (CA)	780-1411 (kg/m^3)
Fine Aggregate (FA)	420-1130 (kg/m^3)
Cement (C)	200-460 (kg/m^3)
Water (W)	109-270 (kg/m^3)
Super plasticizer (SP)	0 - 8.4 (kg/m^3)
W/c Ratio	0.4 - 0.66
Age	3, 7, 28 (days)
Location	1,2 ,3 ,4 ,5 ,6 ,7
Concrete Compressive Strength (CCS)	109 - 547.4 (MPa)

The datasets that were used in the classification process consisted of 200 samples (mixture) taken from Palestinian laboratories for each type of concrete. The types of concrete are B 200, B 250, B 300, B 350, and B 400 which are consisting of seven inputs: coarse aggregate, aggregate Fine, cement, water, water-to-cement ratio, age, and super plasticizer. It also consists of one outlet which is the Compressive Strength of concrete and the following tables 2.2, 2.3, 2.4, 2.5, and 2.6 consists of the ranges of each factor affecting the Compressive Strength of concrete, and the concrete Compressive Strength ranges for each type of concrete.

Table 2. 2: TheB200 Concrete dataset and their valid ranges.

factor name	Range
Coarse Aggregate (CA)	835 – 1375 (kg/m^3)
Fine Aggregate (FA)	525 – 1045 (kg/m^3)
Cement (C)	190 – 300 (kg/m^3)
Water (W)	120 – 180 (kg/m^3)
Super plasticizer (SP)	0 – 7.4 (kg/m^3)
W/c Ratio	0.4 – 0.6
Age	3, 7, 28 (days)
Concrete Compressive Strength (CCS)	0, 1

Table 2. 3: The B250 Concrete dataset and their valid ranges.

factor name	Range
Coarse Aggregate (CA)	820 – 1327 (kg/m^3)
Fine Aggregate (FA)	525 – 1130 (kg/m^3)
Cement (C)	250 – 275 (kg/m^3)
Water (W)	109 – 270 (kg/m^3)
Super plasticizer (SP)	0 – 5.4 (kg/m^3)
W/c Ratio	0.46 – 0.65
Age	3, 7, 28 (days)
Concrete Compressive Strength (CCS)	0, 1

Table 2. 4: The B300 Concrete dataset and their valid ranges.

factor name	Range
Coarse Aggregate (CA)	780 – 1375 (kg/m^3)
Fine Aggregate	595 – 780 (kg/m^3)
Cement	280 – 320 (kg/m^3)
Water	140 – 180 (kg/m^3)
Super plasticizer	0 – 8.4 (kg/m^3)
W/c Ratio	0.44 – 0.66
Age	3, 7, 28 (days)
Concrete Compressive Strength (CCS)	0 , 1

Table 2. 5: The B350 Concrete dataset and their valid ranges.

factor name	Range
Coarse Aggregate (CA)	830 – 1411 (kg/m^3)
Fine Aggregate	420 – 1030 (kg/m^3)
Cement	345 – 360 (kg/m^3)
Water	146 – 205 (kg/m^3)
Super plasticizer	0 – 7 (kg/m^3)
W/c Ratio	0.42 – 0.60
Age	3, 7, 28 (days)
Concrete Compressive Strength (CCS)	0 , 1

Table 2. 6: The B400 Concrete dataset and their valid ranges.

factor name	Range
Coarse Aggregate (CA)	540 – 835 (kg/m^3)
Fine Aggregate (FA)	525 – 975 (kg/m^3)
Cement (C)	380-403 (kg/m^3)
Water (W)	150 – 204 (kg/m^3)
Superplasticizer (SP)	0 – 7.4 (kg/m^3)
W/c Ratio	0.4 – 0.51
Age	3, 7, 28 (days)
Concrete Compressive Strength (CCS)	0 , 1

The following table 2.7 consists of the new datasets that were created after the process of determining the factors affecting the Compressive Strength of concrete in the Palestinian governorates using different clustering algorithms. It was applied to the classification process consisting of 715 samples that were entered in the prediction process, but here it consists only of 4 influencing factors, namely: super plasticizer, w/c ratio, age, and location, and also consists of one output which is the Compressive Strength of concrete and the following table consists the factors affecting the CCS, and its ranges values.

Table 2. 7: The New Dataset Concrete dataset and their valid ranges.

Factor Name	Range
Super plasticizer (SP)	0 - 8.4 (kg/m^3)
W/c Ratio	0.4 - 0.66
Age	3, 7, 28
Location	1,2 ,3 ,4 ,5 ,6 ,7
Concrete Compressive Strength (CCS)	0 , 1

The definitions and terms of the features are the following:

1- Fine Aggregates: it contains sand, crushed stones, gravel, or any other material with similar properties. The aggregates must be clean, hard and do not contain plankton from organic materials or any other impurities. All components of Fine Aggregates must pass through a sieve with an aperture of 6.35 mm (Sieve No. 4), and it can sometimes be overlooked so that what passes from the sieve is not less than 85% of the aggregate, and the components of this aggregate should be of acceptable dimensions, and there are no materials in it.

2- Coarse Aggregates: it contains crushed stone, gravel, or any other material with similar properties. It must be clean, free from impurities as is the case in Fine Aggregates, and the shape of its grains is as close as possible to regular, circular without sharp corners or flat surfaces. Granite or basalt rocks are one of the most important sources of aggregates, as well as limestone.

3- Water: In the preparation of concrete, clean water is used that is free of oils, acids, alkalis, organic materials, and other harmful impurities. Seawater must be avoided in the preparation of the concrete mixture, and the use of pure water from a source adjacent to granitic rocks that causes the dissolution of salts in the concrete should also be avoided.

4- Cement: One of the most important materials used in construction and the basic component for the manufacture of concrete, which is a fine, soft material in the form of a gray powder that is used as a soft binder and possesses cohesive and adhesive properties when adding water to it.

   This leads to harden and form strong building materials that resist the surrounding environmental influences and bind the concrete components together.

5- Super plasticizer: A Super plasticizer is a mixed-water mixture capable of producing significant water reduction or great flow ability without causing undue assembly delay or air leakage in the concrete. The main goal of using the super plasticizer is to avoid particles from separating; Super plasticizers are used to improve the quality of concrete mixtures.

The weak properties of concrete are improved by super plasticizers. Adding these compounds reduces the amount of water required for the concrete mix, which reduces the water-to-cement ratio, but it does not change the ductility of the concrete. The separation of different particles in a concrete mixture can also be avoided by adding super plasticizers.

6- W/C Ratio: The ratio of water to cement is defined as the ratio of the weight of water to the weight of the cement. The water/cement ratio according to the building codes for the concrete mixture ranges from 0.40 to 0.60. The ratio W / C = 0.50 indicates that for every 100 kg of cement, 50 liters of water are used. As the water/cement ratio increases, the total amount of water increases. Workability of concrete means being able to do operations such as mixing, pouring, compacting, and easily separating (granular separation). Moreover, portability to the ability to operate with ease. The m / s ratio is also an important factor for operability.

7- Location: The location is very important, as there are seven Palestinian governorates in this research, and data from laboratories in each governorate was brought separately. The governorates were, Ramallah, Jenin, Tulkarm, Salfit, Tubas, Hebron, and Nablus.

8- Age: Concrete is examined several days after its manufacture, and it may spend 28 days to verify its strength before proceeding to make large quantities of the same concrete mixture. Concrete Compressive strength after 28 days, the concrete Compressive Strength are examined in two stages, where the initial examination begins after 3 days should achieve a strength of 45%, the initial examination begins after 7 days should achieve a strength of 67%, and the final examination will be 28 days after its preparation, and it must achieve strength between 90% to 100%, and sometimes it may be achieved after samples that are at 7 days old as a result of 90%, and then the examination of other remaining samples is not required at 28 days old.

9- Type: In this characteristic of concrete, there are some types of concrete, and these types are B200, B250, B300, B350, and B400, and each of them has features different from others such as the quantity and quality of materials.

It is important to note that other factors affect concrete's Compressive Strength, such as temperature - frost - additives ... etc., but they could not be taken into account because there were not enough experiments to take training information.

2.2 Related Works

In recent years, many researchers have had research orientation in the use of AI in predicting and determining the Compressive Strength of concrete and have shown their success in this field [10-13].

2.2.1 Prediction Phase of Concrete compressive strength (CCS)

In [14], the authors developed a model that predicted the Compressive Strength at 28 days by replacing some parts of cement material with Nano-silica and the result shows that predicate and actual results were nearly the same. This model was trained and tested by neural network tools. The dataset was collected from literature, the result shows that the neural network technique is an effective tool to predicate the Compressive Strength of concrete, and this model predicates the Compressive Strength of concrete on 28th day. In [15], the authors developed a model that represents a neural network regression that predicates the Compressive Strength of concrete. Many tools of freely benchmark were used in this paper and the dataset was collected by UCI machine, and it consists of 1030 records with many of parameters of (Fine Aggregate, Coarse Aggregate, Cement, Fly ash, Super plasticizer, Water and Water/cement ratio). The results show that the actual and prediction result was also the same and the error between them was nearly 1% of predication with actual. These results show that this type of machine learning is one of the most prediction tools in this field.

In [16], the authors developed a methodology that uses Multilayer Feed Forward Neural Networks (MFFNNs) to predicate twenty-eight (28) days of concrete Compressive Strength based on some parameters. The dataset was collected by the UCI machine, and it consists of 1030 records with many parameters. The results show high prediction accuracy, and the authors' results generate some concrete mix proportion rules. In [17], the Compressive Strength is a major criterion in concrete production, but the test on it is complicated and must be saved in special circumstances and an examination should be conducted on it after 28 days in order to extract results that must be reasonable. When the tests fail, large residues and waste of time, Researchers have developed a method for predicting the Compressive Strength of concrete by using Back-Propagation (BP) and this technique was tested and trained in the past by using concrete strength data. The results showed that they can be reliable in predicting the Compressive Strength of concrete because the resulting value in Prediction is approximately equal to the true value of the Compressive Strength of concrete.

According to [18], the authors used Support Vector Machine (SVM) in order to predicate a Compressive Strength of concrete, and this technique uses Artificial Neural Networks in order to suggest a non-destructive technique and a correlation between core strengths and ultrasonic velocities using 6 panels with a zone of strength 24-60Mpa. The results show that ANN and SVM can be reliable in predicting the Compressive Strength of concrete more than that using traditional linear regression. Prediction of the Compressive Strength of concrete by using three types of ultrasonic generates small errors more than using one or two types of ultrasonic. In [19], the authors used two types of data mining techniques represented by Genetic Programming (GP) and Artificial Neural Networks (ANNs) to produce a method to predicate a Compressive Strength of concrete. The authors collected data from the laboratory which has been done in the past and the age was 28, 56, 91 days for every mixture of concrete mixes which were done under particular conditions and standards. The results showed that the prediction results were extracted from both techniques. Comparing results that have been extracted can be trusted by using Artificial Neural Networks with using training function called Levenberg-Marquardt (LM) for Compressive Strength of concrete, and this tool is reliable.

According to [20], the authors worked on Lightweight Aggregate that uses concrete with low-density for building projects (LWAC). The model aims to use a Support Vector Machine (SVM) in order to produce Compressive Strength prediction, Lightweight Aggregate consists of dry density and content; the inputs were water, cement, etc. 241 samples of data collected were done into the laboratory by experimental tests. This model was developed to predicate the Compressive Strength of concrete by using SVM. The results show that the Compressive Strength range that the model does not provide has an accurate result due to the different characteristics of the LWAC.

2.2.2 Main factors affecting the concrete Compressive Strength (CCS)

In [21], the authors used Data Mining for Concrete Compressive strength prediction, and the main aim of the research was to find the main factors affecting the concrete Compressive Strength. Researchers have used Waikato Knowledge Analysis Environment (WEKA) algorithms, and these algorithms are specialized in the classification and clustering process that applies to the dataset, and they used algorithms namely EM, KSOM, K-mean. The researchers made a comparison of the results that emerged from these algorithms on which the data were applied. It was found that the use of Data Mining is very effective and of a high level in predicting the strength of Concrete

Compression. The results showed that algorithms (K-mean, KSOM) are effective and have high accuracy in the process of predicting the Compressive Strength of concrete, and an algorithm (EM) is effective in determining the main factors affecting the compressive strength of concrete.

In [22], the authors designed a model to predict the strength of concrete pressure after 28 days, and the results showed that the Neural Networks can solve complex problems through training and examination. This leads to the emergence of relationships between the input and output, and the results showed that the Neural Networks in the process of predicting the strength of concrete pressure are very accurate after 28 days. This model saved the examiner's cost and time instead of running experiments that had been designed and failed. Using MATLAB, the data were divided into three groups, and they were also divided into training, testing, and validation that percentages were 70%, 15%, 15%. According to [23], the researchers designed a Neural Network based on a back-propagation process that was trained and tested using real datasets collected from laboratories and sources from previous studies. The authors designed a method for predicting the Compressive Strength of concrete by using Back-Propagation (BP). This technique was tested and trained in the past by using concrete strength data and examined dataset that was not used within the limits of the data which had been previously trained. The results showed that the largest error was 20% and it was also noticed that 90% of the results had an error of less than 10% which indicates that this tool is an effective one for predicting concrete Compressive Strength. Moreover, the results showed that the W/C factor is the largest factor that effecting in predicting concrete compressive strength. According to [24], the authors used the waste of solid which is the most important factor of environmental concerns in the world, and Tehran produces more than 20 million tons of construction waste each year. It contains a large amount of Recycled Aggregate Concrete (RAC) that can be obtained from recycled materials. The main aim of this paper was to make a model for the prediction of the Compressive Strength of Recycled Aggregate Concrete (RAC) by using Artificial Neural Networks. The authors divided the dataset for training and testing and the number of mixing was 139 obtained from 14 laborites and factories. The prediction model consists of an input layer with six factors represent fine aggregate, natural coarse aggregate, w/c ratio, water absorption, recycled coarse aggregate, and water-total material ratio, and the RAC Compressive Strength was obtained in the output layer. The results show that the ANNs model is an efficient technique to be utilized to the predicted RAC compressive strength.

According to [25], the author obtained high concrete Compressive Strength from matrix mixture. in this paper, the authors developed a model for the prediction of the Compressive Strength of concrete that contains different matrix mixture by statically model. The different matrix mixture was with different ages like 1,3,7,28,56,90,180 and 365 days. The prediction model examined matrix mixtures with eight factors represent as super plasticizer (SP), water, cement, lime, aggregate, silica fume (SF), sand (S), and meta-kaolin (MK). The matrix mixtures affect Compressive Strength was addressed by authors, and this led to improvements in the prediction of the concrete strength. The results which were obtained from this model have a high correlation for the concrete Compressive Strength of experimental results. In [26], the authors invented a data-driven model that predicts concrete Compressive Strength at 28-day of age. In this model, the authors used two data-driven models represented as Artificial Neural Networks (ANNs) and Multiple Linear Regression (MLR) models aimed to predict concrete Compressive Strength at the 28-day age of different concrete mix design. Also, the input layer considered as concrete components, and the output layer was concrete Compressive Strength. The results show that ANNs data-driven model is an efficient technique to be utilized to the predicted concrete Compressive Strength and MLR is not an efficient technique to be utilized to the predicted concrete Compressive

Strength. According to [27], the authors invented the model of a Neural Network that predicates of Compressive Strength of Light Weight Concrete (LWC) after 3, 7, 14, and 28-days of curing. In the feed-forward Back -Propagation (BP), the prediction model, examined the Compressive Strength of concrete (CCS) with eight factors represented as curing period, super plasticizer, lightweight fine aggregate, lightweight coarse aggregate and, w/c ratio silica fume used in solution and silica fume used in addition to cement. The results show that CC Neural Network is an efficient technique to be utilized to the predicted concrete Compressive Strength compared with BP with accuracy and speed. Also, it shows that the ANNs model is an efficient technique to be utilized in order to predict Light Weight Concrete (LWC) Compressive Strength. According to [28], the authors have invented an Evolutionary Artificial Neural Networks (EANNs) model that predicts concrete Compressive Strength (CCS), in this model, namely EANNS, the authors have used two techniques represented as Artificial Neural Networks (ANNs) and evolutionary search algorithms, like genetic algorithms (GA). The authors divided the dataset for training and testing and the number of mixing was 173 with different characteristics, the prediction model examined the Compressive Strength of concrete with seven factors represented as a large amount of sand, cement, w/c ratio, 3/4 sand, and 3/8 sand, coefficient of soft sand parameters and maximum size of coarse aggregates. The number of layers, nodes, and weights in ANNs models are optimized by using GA; the results show that optimized Neural Network is an efficient technique to be used to predict CCS compared with MLR in accuracy, capability, and flexibility.

In [29], the authors obtained High Concrete Compressive Strength from Artificial Neural Networks (ANNs). In this paper, the authors developed a model for the prediction of the Compressive Strength of concrete by ANNs technique at 28 days because this age is most often used for quality control. The prediction model was examined against Compressive Strength with eight factors represented as w/c ratio, water, cement, fine and coarse aggregates. The results showed that the ANNs model is an efficient technique to be utilized in order to predict high Concrete Compressive strength when obtaining results with higher reliability, capability, and flexibility than knowledge of the relationships between the parameters involved in the design. Furthermore, the results showed an obtained correlation of the order of 0.94 by ANN technique that predicts the Compressive Strength of concrete based on their manufacturing parameters. In [30], the authors invented the model of a neural network that predicates the Compressive Strength of ground granulated blast furnace slag concrete at 3, 7, 28, 90, and 360 days. The authors divided the dataset for training and testing, the number of mixing was 45 with different characteristics like were three different w/c ratios (0.3, 0.4, and 0.5), three different cement types (350, 400, and 450kg/m^3), and replacement ratios with four partial slag (20%, 40%, 60%, and 80%). The prediction model examined the Compressive Strength of ground granulated blast furnace slag concrete with six factors represented as ground granulated blast furnace slag, cement, water, super plasticizer, fine and coarse aggregate, and age of samples. The results showed that the ANNs model is an efficient technique to be used in order to predict ground granulated blast furnace slag concrete using concrete ingredients as input factors.

2.2.3 Classification phase for Concrete Compressive Strength (CCS)

In [31], the authors invented a new model to explore the capability of the Artificial Neural Networks (ANNs) model contains the Elastic modulus (Ec) of recycled aggregate concrete. The authors divided the work into two parts; the first part called ANNs-I and this part uses 324 datasets collected from 21 international published literature and the second part called ANNs-II which uses 16 more datasets adding to previous datasets and these additive datasets from authors, were randomly shared into three groups as the training, testing and validation sets, respectively. The results show that Artificial Neural Networks (ANNs) is an efficient technique to be utilized in order to predict concrete Compressive Strength of Elastic modulus (Ec) of recycled aggregate concrete. According to [32], the authors invented a new model of Artificial Neural Networks (ANNs) model that predicts the Compressive Strength of the concrete based on non-destructively determined parameters. The datasets were divided into three groups; training, validation, and testing sets, respectively, and they contain several inputs represented with concrete components one output represented as concrete Compressive Strength ranging from 24 to 105 MP a. The results showed that the Artificial Neural Networks (ANNs) model is an efficient technique to be used in order to predict concrete Compressive Strength of concrete.

In [33], the authors invented a new model form combination from techniques in order to explore the ability of Artificial Neural Networks (ANNs) as a first technique; the second technique is adaptive A Neuro-Fuzzy Inference System (ANFIS), the third technique is Multivariate Adaptive Regression Splines (MARS) and M5 Model Tree (M5Tree) technique in order to predict the Compressive Strength of ultimate conditions of Fiber-Reinforced Polymer (FRP)-confined concrete, by using datasets more than 1000 axial compression tests results of FRP-confined concrete mixtures. Datasets were randomly shared into three groups as the training, testing, and validation sets, respectively. Results showed that the combining models represented as ANN, ANFIS, MARS, and M5Tree models are suitable with the experimental test data. Also, they show that the proposed model represented as ANN, ANFIS, MARS, and M5Tree techniques is an efficient model to be utilized to predict Compressive Strength of ultimate conditions of Fiber-Reinforced Polymer (FRP)-confined concrete compared with those of the existing conventional and evolutionary algorithm models in accuracy and estimation. According to [34], the authors have invented a new model from a combination of techniques for exploring the ability of Artificial

Neural Networks (ANNs) as a first technique. The second technique is Adaptive Network-based Fuzzy Inference System (ANFIS); the third technique is parametric regression to predict the Compressive Strength of expanded polystyrene beads (EPSs), note that EPS concrete is a type which is very sensitive and of lightweight concrete made by partial replacement of concrete coarse aggregates with lightweight expanded polystyrene beads (EPSs). The results showed that ANNs are an efficient technique to be utilized in order to predict Concrete Compressive strength. It also shows that ANFIS elite model is an efficient technique to be utilized to predict the Compressive Strength of this concrete, comparable with the last model is not efficient and showed a weakness point.

According to [35], the authors used a data-driven model that predicts the Compressive Strength of no-slump concrete at 28-day of age, No-Slump Concrete (NSC) which means that the concrete has either very low or zero slumps. In this model, the authors used two data-driven models represented as Artificial Neural Networks (ANNs) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models which aimed to predict Compressive Strength at 28-day age with no slump mixtures. Datasets were divided into training and testing datasets and the input layer is considered a concrete component. The output layer was concrete Compressive Strength. The results showed that ANNs and AFIS data-driven models are an efficient technique to be utilized to predict the Compressive Strength of no-slump concrete and a traditional regression model is not an efficient technique to be utilized to predict the Compressive Strength of no-slump concrete. In [36], the authors developed a model for the prediction of Compressive Strength of high-performance concrete using a static model for a dataset obtained from authors, and which was with different ages like 3, 7, 14, 28, and 91 days. Multiple non-linear regression has an excellent correlation coefficient for the prediction of the Compressive Strength of high-performance concrete at the period of days mentioned previously. Thus, this led to improve the prediction of the concrete strength. The results were obtained from this model have a high correlation for the concrete Compressive Strength of experimental results, the coefficient of correlation was 99.99% for each Compressive Strength of high-performance concrete at each age. In [37], the author was obtained High-Performance Concrete Compressive Strength from highly complex material and this made the behavior of modeling a very difficult task. the authors developed a method for modeling the concrete slump flow using Artificial Neural Networks (ANNs) and second-order regression. The slump flow of High-Performance Concrete (HPS) is determined by the maximum size of coarse aggregate and water content. The result shows that the concrete slump flow model based on ANNs is more comparable than based on regression analysis which means that ANNs are much more accurate than regression analysis.

In this Study, from the previous references, then the conclusion that neural networks are very effective in the phase of classification and prediction. While determining the factors affecting the compressive strength of Palestinian concrete has been identified for the first time in this field in the Palestinian governorates, while there are many techniques used in the process of predicting the compressive strength of concrete, and the results showed that some previous studies did not get the least square error compared to the results deduced from prediction phase.

In the classification phase, there are many references that have concluded that there are some effective techniques in all fields of engineering and science, and this is what we have reached, while more than one technique has been used, so that we have chosen the best three by consensus of previous studies.

The methodology divided into three phases, in the first phase, the aim was to determine the most important factors that affect CCS in the Palestinian governorates, the second phase was a process of classifications for all types of concrete, and the last phase was to predict Compressive Strength of concrete in the Palestinian governorates. Three algorithms were used for each phase and the results show that machine learning techniques are effective in prediction and classification processes. In the first phase, KM, EM, and KSOM algorithm were used for detecting the most important factors that affect CCS in Palestinian governorates, these techniques were applied to whole datasets of Palestinian governorates and applied on each dataset of each governorate of Palestinian governorates. In the second phase, MLPNNs, SVM, and Ensemble algorithms were used. The results show that MLPNNs are more accurate than others in each type of concrete type. The last phase, MLPNNs, RNNs and RBFNNs were used.

Chapter 3

Methodology

3.1Methodology

In this thesis, different ML algorithms were used to determine the factors that affect CCS, predict CCS and classify CCS for each type of concrete. For classification and prediction of CCS, MATLAB software was used, while the WEKA tool was used to determine the important factors that affect the CCS.

In the first phase, the work was divided into two parts, the first part was the implementation of algorithms to the entire combined data of the Palestinian governorates, and the second part was the implementation of algorithms to all Palestinian governorates separately showing the influencing factors in the governorates as a whole and showing each governorate separately.

The second part is the use of machine learning methods for the classification process. MLPNNs, SVM, and Ensemble Algorithm were used to classify the CCS with all types of strength. The classification results from these models were very efficient.

In the final part of this study, different ANNs models were used to predict the Compressive Strength; MLPNNs, RBFNNs, and RNNs techniques produce very good prediction results with an appropriate number of neurons.

The data collected from Palestinian factories and laboratories consists of five sets of data sets; each data set is related to a specific type of Palestinian concrete. Three methods were used: Figure 3.1 represents the output of proposed models for prediction, classification, and detecting main factors that affect the Compressive Strength of concrete in Palestinian governorates using different techniques for each output.

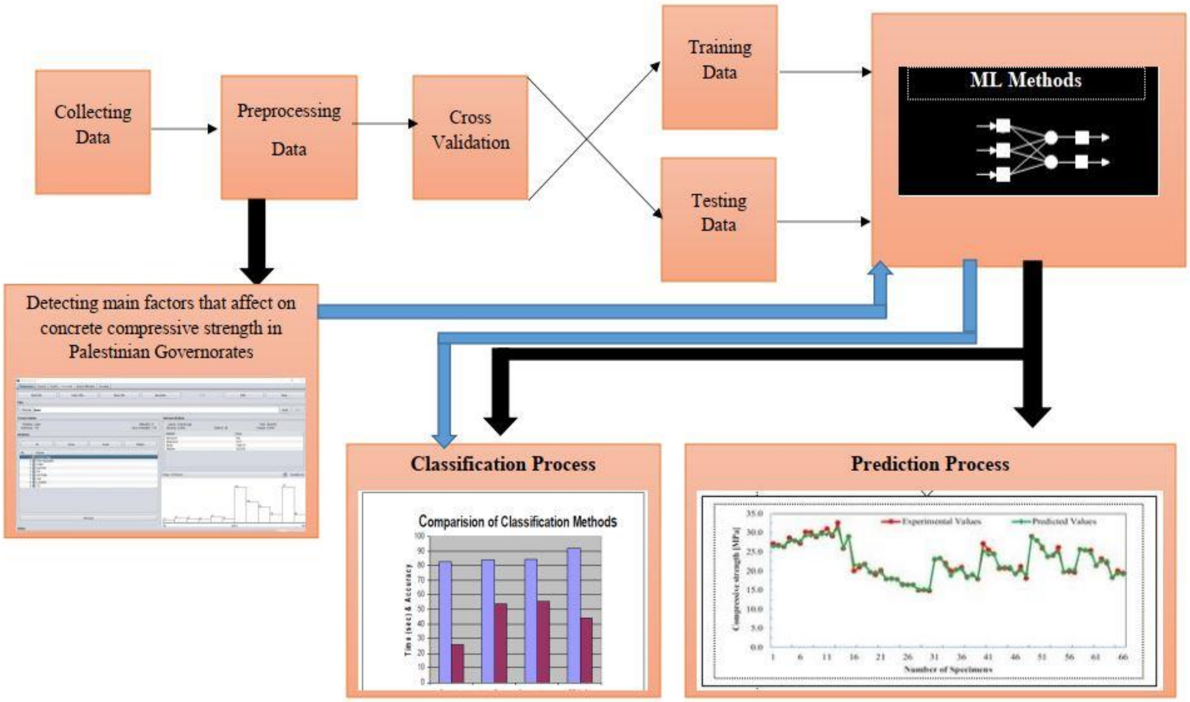


Figure 3. 1: Represents the architecture of the main work for all processes.

3.2 Preprocessing Phase

Data preprocessing is an important step in machine learning datasets application. Different preprocessing steps may be used depending on the nature of the dataset [38]. Output selection and data normalization were used in this thesis. This section will describe these steps in detail.

3.2.1 Output Selection

Many factors can be produced from the concrete mix, such as moisture, temperature. In this work, the concentration is on the Compressive Strength of concrete

3.2.2 Data Normalization

Normalizing data is an important preprocessing step in machine learning technique to deny one feature and to dominate the other features; normalization aims to make data points of all features have the same scale to have the same importance. Min-Max normalization is one of many data normalization methods [38]. In this research, this type is used.

1. Min-Max Normalization: it performs a linear transformation on the data and scales the attribute to a fixed range. In this work, the range between [0,1] is used, Min-Max normalization is calculated using the following equation:

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad 3.1$$

Where y is the normalized value obtained from this equation, x is the original value of the feature, x_{min} is representing as the minimum value of the feature, and x_{max} is represented as the maximum value of the feature.

3.3Building Models Phase

In the beginning, the collected data were analyzed and classified according to the data and the main governorates in Palestine, then three clustering algorithms as EM, KM, and KSOM, were used based on clustering algorithms. The main parameters affecting the compressive strength of the concrete will be determined, and then the results based on the output and determine the appropriate algorithm will be compared to determine the main factors. The number of clusters was determined and accordingly and the results will be recorded in a specific table so that each governorate will have different results from the others based on the information inferred through them and the principle of these algorithms depends on the standard deviation of all the clusters that available. So, the original standard deviation of each governorate will be compared with a standard deviation of the results from the clusters that available, and then it will be decided whether the influencing factor is of great influence or importance and it will be recorded based on our readings. The results show that each governorate has different influencing factors from the others because there is a difference in the quantities and systems in Palestine.

In the second step, classifications of concrete Compressive Strength were made in the Palestinian governorates for each type of concrete. This process was carried out using three algorithms which are MLPNNs, SVM, and Ensemble. Each one gives accurate results. In the third step, classifications of the most important factors affecting the Concrete Compressive Strength (CCS) in the Palestinian governorates were made resulted from the use of the clustering process, which are 4 input. The algorithms that were used in the previous classifications process were applied, and the results showed that the accuracy is close between the two phases.

In the last stage, a prediction for concrete Compressive Strength (CCS) in the Palestinian governorates was made using eight factors as input and only one output, which is the concrete pressure strength in the Palestinian governorates. The dataset consisted of 715 samples using three algorithms which are MLPNNs, RBFNNs, and RNNs.

3.3.1 K-Means Clustering Algorithm (KM)

Clustering is perhaps the most widely recognized exploratory data analysis strategy used to get an instinct about the construction of the data. It is very well and may be characterized as the errand of recognizing subgroups in the data with the end goal that information focuses in a similar subgroup (cluster) which is very much like while data focuses in various groups that are different. In general, it is attempted to discover homogeneous subgroups inside the data with the end goal that data focuses in each cluster are as comparative as conceivable as per a comparability measure. For example, Euclidean-based distance or connection-based distance. The choice of which likeness measure to utilize is application-specific [39].

Clustering analysis should be possible based on highlights where we attempt to discover subgroups of tests dependent on highlights or based on examples. We'll cover here clustering dependent on highlights. Clustering is utilized in the market division; where we attempt to discover clients that are like each other whether as far as behaviors or attributes, where we attempt to amass comparative districts, record grouping dependent on themes, and so on [40]. Clustering is viewed as an unsupervised learning technique since we don't have the ground truth to look at the yield of the grouping calculation to the true labels to assess its presentation. We just need to attempt to research the design of the information by gathering the information focuses into unmistakable subgroups, while KM is one of example that used clustering and it is the most used clustering algorithm. KM is an iterative algorithm that attempts to parcel the dataset into Kpre-characterized particular non-covering subgroups (clusters) where every data point has a place with just one gathering. It attempts to make the intra-group data focuses as comparative as could be expected under the circumstances while likewise keeping the clusters as various (far) as could reasonably be expected. It allocates points of data focuses to a group with the end goal that the amount of the squared distance between the points of data focuses and the cluster’s centroid is at the minimum [41]. The less variety available inside clusters, the more homogeneous (similar) points of data focuses are inside a similar cluster.

K-means Clustering Algorithm Work Steps:

1. Determining the number of clusters K.
2. Initializing centroids by first rearranging the dataset and afterward randomly choosing K point of data focuses for the centroids without replacement.
3. Continuing to emphasize until there is no change to the centroids .i.e the assignment of data points to clusters isn't changing.

- Computing the sum of the squared distance between all centroids and points of data using the following equation (3.2):

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\|x_i - v_i\|)^2 \qquad 3.2$$

Where $\|x_i - v_i\|$ is the Euclidean distance between the point x_i and a centroid v_i, iterated overall k points in the i_{ih} cluster, for all n clusters. c_i is the number of data points in i_{ih} cluster, C is the number of cluster centers.

- Appropriating each data point to the nearest cluster (centroid), where, c_i represents the number of data points in i_{ih} cluster using this following equation (3.3):

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i \qquad 3.3$$

- Taking the average of all data points that belong to each cluster and then compute the centroids for the clusters.

KM solved Expectation-Maximization Problem. The E-step appropriate the points of data to the nearest cluster. The M-step calculates the centroid of each cluster [42].

3.3.2 Kohonen Self-Organizing Map (KSOM)

Kohonen Self-Organizing Map (KSOM) is one of the most popular unsupervised learning techniques [43]. Invented by Teuvo Kohonen in 1982, and this algorithm performs vector quantization depending on similarities of patterns. It is a neural network-based cleavage clustering approach as it maps genes into a series of sections in the neuron layer resulting in the similarity of their expression vectors to the reference vectors or weights specified for each section.

KSOM clustering result might be influenced by a few parameters like learning parameters and topology map sizes, therefore, the dataset is frequently trained with different map sizes [44]. To find the most suitable map size that can accurately represent a clustering of the datasets, this algorithm is also able to process high-dimensional datasets because it is implemented to group data into clusters that show some similarities. Each cluster with similar features is displayed on the same node on the map. Otherwise, the difference is increased with the distance separating the two objects on the map, therefore, the space of the cluster is determined on the map so that the object enables the visualization and simultaneous observation of the cluster [45]. KSOM has many steps to group data that have similar features into clusters. This algorithm begins with measuring the distance between cluster centers or cluster nodes in the topographic map by using Euclidean Distance [46]. The Euclidean Distance is used to calculate the distance in the plane using equation (3.4) [47], [48] and these steps are:

Let Input: = {x1, x2, x3…xn}, w_{ij} be the weight vector associated with unit positioned between i and j.

1- All nodes weights must be initialized.
2- A vector is chosen at random from the set of training data and presented into a grid.
3- The distance between all inputs and output nodes is calculated using Euclidean Distance using the following equation (3.4):

$$d(i,j)=\sqrt{\sum_{j=1}^{N}(w_{ij}-x_i)^2} \qquad 3.4$$

4- The winning unit is selected by the minimum $d(i,j)$.

d_{ij} : Is the distance between weight vector w_{ij} and given pattern x_i.

5- Calculate all weights for neighbor nodes using Gaussian Function.

6- Nodes weight is adjusted to make it more likely the input vector by Each neighboring using this equation (3.5):

$$w_{ij(t+1)}=w_{ij(t)}+\alpha\left(\left(\frac{\|rc-rk\|^2}{\delta}\right)-w_{ij(t)}\right) \qquad 3.5$$

Where α is the learning rate, $\|rc-rk\|^2$ define the distance between neuron's 2D matrix positions (√N x √N matrix), delta are monotonically decreasing functions of time-varying.

7- The learning rate must be updated at a certain time.

8- Repeat Step 2 for N iterations.

3.3.3 Expectation-Maximization (EM)

The expectation-maximization (EM) algorithm utilizes when some of the data missing is used to obtain maximum probability estimates for parameters. Although this algorithm can also be used when there is hidden data, that is, unclear data which was never supposed to be monitored in the first place, It can be done in this case. It is simply assumed that the hidden data is lost and going ahead and applying the EM algorithm. This algorithm is very unique in statistics and mathematics, it is used extensively in machine learning, AI applications, data elicitation, and Bayesian statistics where they are used in background marginal distributions of parameters [49]. Assuming that the complete data-set consists of $z=(x,y)$, but that only χ is observed.

The complete-data log-likelihood is then denoted by $l(\theta; x, y)$ where θ is the unknown parameter vector.

Expectation Step (E-Step): EM algorithm in E- Step calculates the expected value of $l(\theta; x, y)$ given the observed data, x the current parameter estimates, θ_{old} say. In particular, by using the following equation (3.6):

$$Q(\theta;\theta_{old}) := E[l(\theta;x,y) \mid x;\theta_{old}] \quad = \int l(\theta;x,y)\, p(y \mid x,\theta_{old})\, dy \qquad 3.6$$

Where $p(y \mid x, \theta_{old})$ is the conditional density of Y given the observed data, X, and assuming $\theta = \theta_{old}$

M-Step: The M-step consists of maximizing over θ the expectation calculated in (3.7). That is, we set

$$\theta_{new} := \max_{\theta} Q(\theta;\theta_{old}) \qquad 3.7$$

We then set $\theta_{old} = \theta_{new}$

These steps are persistent as necessary until the sequence of θ_{new}'s converges. In fact, under very general conditions convergence can be guaranteed with a local maximum, and why this is explained below. If there is a suspicion that the log-likelihood function has multiple local maximums, the EM algorithm should be running multiple times, using a different starting value θ_{old} on each occasion. The ML estimate of θ is then taken to be the best of the set of local maximums obtained from the various operations of the EM algorithm.

Algorithms for Clustering: K Mean, KSOM, and EM.

Input: Train dataset, Test Dataset, the dataset for CCS from each Palestine governorates;
Output: detect main factors that affect CCS in Palestine governorates result;

Data Preprocessing Phase:

Step1: choosing the number of clusters

Step2: initializing centroid for each cluster

K Mean Algorithms Training Phase:

Get number of clusters

Get Train dataset

K Mean Testing Phase:

Get parameters from the training phase

Get Test dataset

Calculate final cluster centroids result for each factor

Output final cluster centroids result

Determine the cluster that has the maximum CCS value.

Choose this cluster.

Check final cluster centroids for each factor.

if final cluster centroids for any factor is greater than FULL data.

Choose this factor.

Save network parameters.

KSOM Algorithms Training Phase:

Get number of clusters

Get Train dataset

KSOM Testing Phase:

Get parameters from the training phase

Get Test dataset

Calculate standard deviation result for each factor

Output standard deviation result

Determine the cluster that has the maximum CCS value.

Choose this cluster.

Check standard deviation for each factor.

if the standard deviation for any factor is nearly equal to 0

Choose this factor.

Save network parameters.

EM Algorithms Training Phase:

Get number of clusters

Get Train dataset

EM Testing Phase:

Get parameters from the training phase

Get Test dataset

Calculate standard deviation result for each factor

Output standard deviation result

Determine the cluster that has the maximum CCS value.

Choose this cluster.

Check standard deviation for each factor.

if the standard deviation for any factor is nearly equal to 0

Choose this factor.

Save network parameters.

Main factor detection Results:

Get optimal parameters that affect CCS in Palestine for each cluster from these algorithms.

Get the common factors among the results obtained from these algorithms for each number of clusters.

Output Main factor results.

3.3.4 Multi-Layer Perceptron Neural Networks (MLPNNs)

Artificial Neural Network technology is a simulation of the work of the biological Neural Networks in the human brain [50]. It has been successfully used on multiple models and with many applications in industry, medicine, financial forecasting, and civil engineering [51]. It has been used to estimate the cost and productivity of water projects. It must be noted that Artificial Neural Networks is not a program but rather it teaches [52]. Neural Networks, as shown in Figure (3.2), consists of simple processing units called neurons with many connections between them distributed in several layers [53].

Basic principles of MLPNNs were explained in [54]:

1- **Input Layer:** it contains several neurons equal to the number of factors studied, and they do not perform any processing process, so it is where the network feeds the data and it is the one that transmits the information to the hidden layer.

2- **Output Layer:** It contains several neurons equal to the number of values that will be obtained or specified.

3- **Hidden Layer:** it is the one that feeds the output layer and is present between the first and last layers. Try and error until you get an optimized network.

   It is distinguished between two types of Artificial Neural Networks according to the number of hidden layers within them:

   1- Single-layer Artificial Neural Networks: which consists of the input layer and the output layer only.

   2- Multilayer Neural Networks: which are composed of the input layer, output layer, and one or more hidden layers.

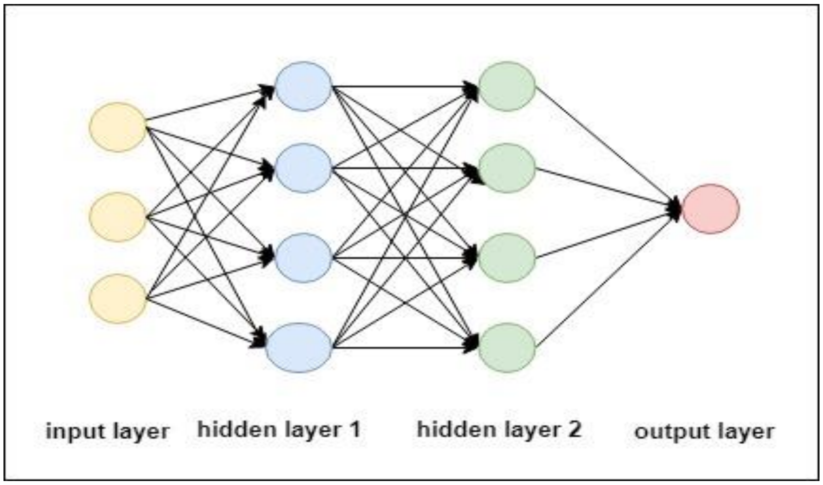


Figure 3. 2: General model of multilayer Artificial Neural Networks.

These factors that have been entered into the input layer are processed internally in the network and modified by multiplying them with numbers called the weights that the network changes during training [55], then they are transferred and processed again through the hidden layers to be transferred to an acceptable output layer using a similar output. MLPNNs have two phases: forward and backward propagation. In the forward phase, the output is predicted, the error is calculated and sent back to the backward prorogation phase. During the backward propagation, the calculated error is propagated back through the network to adjust the weights and reduce the error in the output layer.

The training process of the MLPNNs is mapping the input to the corresponding output. It begins with providing input and initial weights to the MLPNNs then adjusting the weights to minimize the error between the desired and actual output of the network. The output of the MLPNNs is the weighted sums of the inputs which calculated using the following equation 3.8:

$$Y_{ij} = w_{ij}.x_i \quad 3.8$$

Where w_{ij}: is the connection weight between the *i*th node in the input layer and the *j*th node in the hidden layer, and x_i: is the *i*th input. To stop the training process, there is a certain threshold θ is set depends on the error of the MLPNN [56] which represents the difference between the desired and actual output. The error is calculated using the following equation 3.9:

$$MSE = \frac{1}{2}\sum_{i}^{n}(y_d - y_i)^2 \qquad 3.9$$

y_d is a predictive output and y_i is the actual output. The training process continues to tune the weights and minimize the error to be small enough regarding θ. The weights updated using the following equation 3.10:

$$\Delta w_{i+1} = \alpha . \mathrm{E} . x_i \qquad 3.10$$

MLPNNs with feed-forward back-propagation algorithm consists of three layers represented by feed forward, multi-input multi-output as follows:

- Input layer X_i , i = 1, 2, …, n. Where n is the number of input nodes.
- Hidden layer j: Each node is a neuron, each neuron connected to the input layer by the processing unit called weights w_{ij}, where i is the input node and j is the hidden layer node.
- Output layer k: Contains the nodes that produce the output of the network represented by several neurons depends on several outputs, Y_k.

When the training phase fed to the input layer, the sum of weights from input to the j^{th} node in the hidden layer is given by:

$$y = \sum W_{ij} X_i + \theta_j \qquad 3.11$$

θ_j: called the bias node that always has a value of 1, calculate the gradients efficiently done by back propagation algorithm when using MLPNNs. The back-propagation algorithm always starts from the last layer (output layer) and propagates backward to update the weights of the network, it needs an activation function, typically used the sigmoid function. The actual output of the j^{th} node is:

$$Y_j = X_k = \frac{1}{1+e^{-y}} \qquad 3.12$$

In the output layer, the difference value between the actual and the target value is Δ_k, where the actual value of the node k is Y_k and the target value is t_k , while X_k is the input to the next layer' node.

$$\Delta_k = t_k - Y_k \qquad 3.13$$

δ_k: The error signal of the output layer is calculated by Δ_k and the derivative of the sigmoid function.

$$\delta_k = \Delta_k Y_k (1 - Y_k) \qquad 3.14$$

The change in the weight between node j and node k is done by multiplying the error at node k by the output of node j by using the delta rule.

$$\Delta w_{jk} = l\ \delta_k X_k \qquad 3.15$$

w_{jk}: The weight between node j and k, where l is the learning rate, so, to update it by the following formula:

$$w_{jk} = w_{jk} + \Delta w_{jk} \qquad 3.16$$

To calculate, the error signal δ_j for node j in the hidden layer, δ_j: The error signal for node j in the hidden layer is calculated by the following formula:

$$\delta_j = (t_k - Y_k) Y_k \sum w_{jk}\, \delta_k \qquad 3.17$$

w_{ij} is the weights between the input node i and the node j can be updated by using 14 and 15 so

$$\Delta w_{ij} = l\ \delta_j X_j \qquad 3.18$$

$$w_{ij} = w_{ij} + \Delta w_{ij} \qquad 3.19$$

The back-propagation algorithm repeats until the error on the output node is minimized.

3.3.5 Support Vector Machine (SVM)

The main aim of the Support vector machine technique is to search a hyper-plane in an N-dimensional space (where N is the number of features that classified the points of data).To separate the two categories of data points, several possible hyper-planes can be chosen. The main aim is to search for a plane that has the maximum distance between data points of both classes and maximizing the margin distance leads to provide some reinforcement so that future points of data can be categorized with more trust.

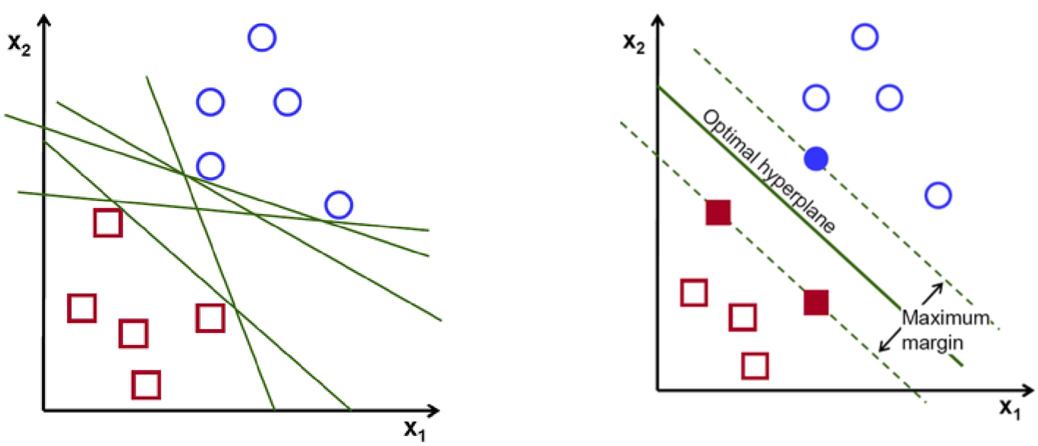


Figure 3.3: SVM Algorithm.

In the support vector machine, the look is to maximize the margin between hyper-plane and point of data, the function that helps maximize the margin is discontinue loss called loss function c(x,y,f(x)) given by training data (xi, yi) for i = 1 ...N, with xi ∈ Rd and yi ∈ {−1, 1}, learn a classifier f(x).

$$c(x, y, f(x)) = \begin{cases} 0 \quad , & if\ y * f(x) \geq 1 \\ 1 - y * f(x), & if\ y * f(x) < 1 \end{cases} \qquad 3.20$$

$$c\big(x, y, f(x)\big) = (1 - y * f(x))_+ \qquad 3.21$$

The cost equals zero. If the actual value and the predicted value are of the same sign, if the two values are not equal to zero, then it calculates the loss value by the cost function. When it adds a parameter called regularization parameters to the cost function, it looks as below, the regularization parameter aims to balance the margin maximization and loss by using (3.22), where Xi is the input, w is the weight and λ is regularization parameters:

$$\min_w \lambda \|w\|^2 + \sum_{i=1}^{n} (1 - y_i \langle x_i, w\rangle)_+ \qquad 3.22$$

By using the loss function, it takes partial derivatives concerning the weights of the data points to find the gradients δthat able to update the weights represented by (3.23).

$$\frac{\delta}{\delta w_k} \lambda \|w\|^2 = 2\lambda w_k$$

$$\frac{\delta}{\delta w_k} (1 - y_i \langle x_i, w\rangle)_+ = \begin{cases} 0 \quad , & if\ y_i \langle x_i, w\rangle \geq 1 \\ -y_i x_i, & if\ y_i \langle x_i, w\rangle < 1 \end{cases} \qquad 3.23$$

No misclassification is in our method when correctly predicts the class of data points, it has one solution that is represented in (3.24) by updating the gradients from the regularization parameter.

$$\omega = \omega - \alpha. (2\lambda\omega) \qquad 3.24$$

When our method makes an error when predicting the class of data points, this is called misclassification, it has one solution that is represented in (3.25) by updating the gradients including the loss along with the regularization parameter.

$$\omega = \omega + \alpha.(y_i.x_i - 2\lambda\omega) \qquad 3.25$$

3.3.6 Ensemble Algorithm (ES)

Ensemble algorithms are methods that lead to improving results accuracy by using multiple models combined instead of using a single model. The results accuracy increasing significantly when combing models, which is using in machine learning. Ensemble methods have two categories that are: called parallel ensemble techniques and sequential ensemble techniques. The sequential techniques have a base learner in a sequence. The dependence between the base learners is promoting by the sequential generation of base learners, assigning higher weights to previously misrepresented learners improved the performance of the models. The other type of ensemble algorithms is called parallel ensemble techniques that the base learners are working in a parallel format like the random forest, it used this method to encourage independence between base learners, the base learner's independence reduces significantly by the mistakes in the application of averages. Some ensemble algorithms methods apply in base learning with a single algorithm only this leads to make a result in all base learners as a homogeneity, but inhomogeneous with similar quality, the base learner back to base learners of the same type, but in distinct types the base learner is heterogeneous. Ensemble algorithms using multiple models combined instead of using a single model and these models are:

Bagging Ensemble Algorithm: The idea is simply to collect several different expectations about our data and find the best results after collecting it. The following image simply illustrates the idea.

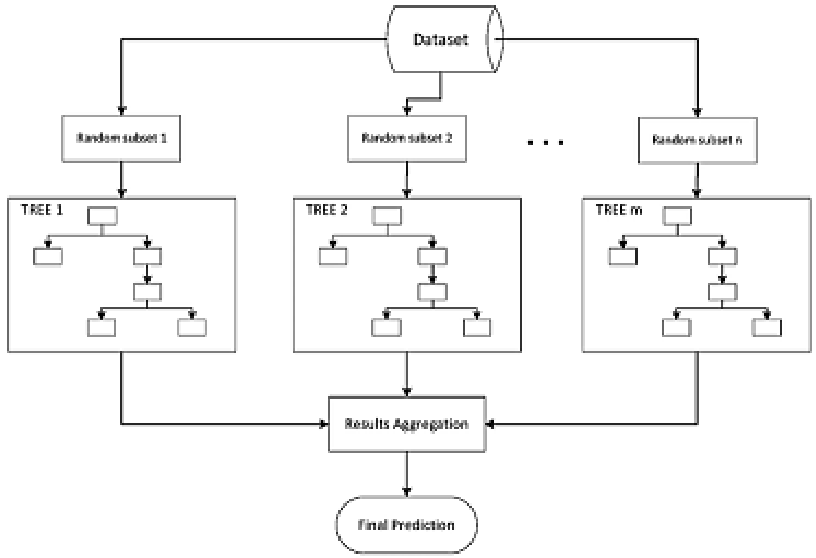


Figure 3. 4: Bagging Ensemble architecture.

Boosting Ensemble Algorithm: In the same way as before, but in it, the wrong results are taken and retrained until arrived at a suitable model.

Random Forests Ensemble Algorithm: This algorithm works with the same idea as the decision tree algorithm, and its operation will be as follows:

1- It takes several features and creates a decision tree out of it.

2- The same process is repeated with different variables (a variable can be repeated in more than one tree).

3- After completion, the test is done, the test is done on each tree, which is created in the first two steps and its results are shown.

4- And from those results, the best is chosen.

Algorithms for classifications: MLPNNS, SVM, and Ensemble.

Input: Train dataset, Test Dataset, the optimal value for CCS of concrete type;
Output: Train and test accuracy result;
Data Preprocessing Phase:
Step1: Normalize the dataset samples 0 or 1

```
for subset in dataset do
            for an item in subset do
               if age =3
                      if CCS>= CCS (28 day *0.45)
                        CCS=1
                      Else
                        CCS=0
               if age =7
                      if CCS>= CCS (28 day *0.67)
                        CCS=1
                      Else
                        CCS=0
               if age =28
                      if CCS>= optimal value for CCS of concrete type
                        CCS=1
                      Else
                        CCS=0
```

MLPNNs Training Phase:

```
Get optimal Parameters
        Get Train dataset
While    MSE less than the threshold
           While   Termination condition not satisfied
                    Calculate classification result
                    Calculate MSE
                    Update weights
        Number of neurons= Number of neurons+2
        Output MSE
        Output classification result
        Save network parameters
```

MLPNNs Testing Phase:

```
       Get network parameters from the training phase
       Get Test dataset
       Calculate test classifications result
       Output MSE
       Output test classification result
```

SVM Training Phase:

```
Get optimal Parameters
        Get Train dataset
While    MSE less than the threshold
           While   Termination condition not satisfied
                    Calculate classification result
                    Calculate MSE
                    Update weights
        Number of neurons= Number of neurons+2
        Output MSE
        Output classification result
        Save network parameters
```

SVM Testing Phase:

```
       Get network parameters from the training phase
```

Get Test dataset
Calculate test classification result
Output MSE
Output test prediction result

Ensemble Training Phase:

Get optimal Parameters
Get Train dataset
While MSE less than the threshold
While Termination condition not satisfied
Calculate classification result
Calculate MSE
Update weights
Number of neurons= Number of neurons+2
Output MSE
Output classification result
Save network parameters

Ensemble Testing Phase:

Get network parameters from the training phase
Get Test dataset
Calculate test classification result
Output MSE
Output test classification result.

Comparison Test Results:

Get the optimal number of neurons that have maximum accuracy from MLPNNs.
Get Accuracy results for each algorithm.
While any algorithm has maximum accuracy.
Output accuracy results.
Choose this algorithm.

Main factor Test Results:

Get the optimal number of neurons that have maximum accuracy.
Get optimal parameters that affect CCS in Palestine.
Get Accuracy results from this algorithm.
If classification accuracy nearly equal main affect classification accuracy.
Output accuracy results.
Choose this algorithm.

3.3.7 Radial Basis Function Neural Networks (RBFNNs)

RBF Neural Networks are also a type of feed-forward neural network trained using a supervised training algorithm. The main point of this type has only one hidden layer, it uses an activation function called radial basis function, and this function is very strong in approximation and calculation., These types of Neural Networks are implemented in different problems and successful implementation could be achieved by a lot of researchers, the RBF network algorithms trains much faster than back-propagation networks. The general equation for the output of RBFNNs network [57] can be represented as follows) by using the Gaussian function as the basis function.

$$y(x) = \sum_{i-1}^{M} w_i e^{(\frac{-(\|x-c_i\|)^2}{2\sigma^2})} \qquad 3.26$$

These parameters $x, y(x), c_i, \sigma, M$ represent input, output, center, width, and several basis functions centered at ci, Similarly, w_i represents weights. Denoting this algorithm, it can show the basic architecture of RBFNN (Fig 3.5) by constructed an RBFNNs by taking the Gaussian function as the basis function and considering randomized centers and width. The data transform from input neurons to hidden neurons that have a radial basis function as an activation function calculate the distance between the input layer and hidden layer centers by this function.

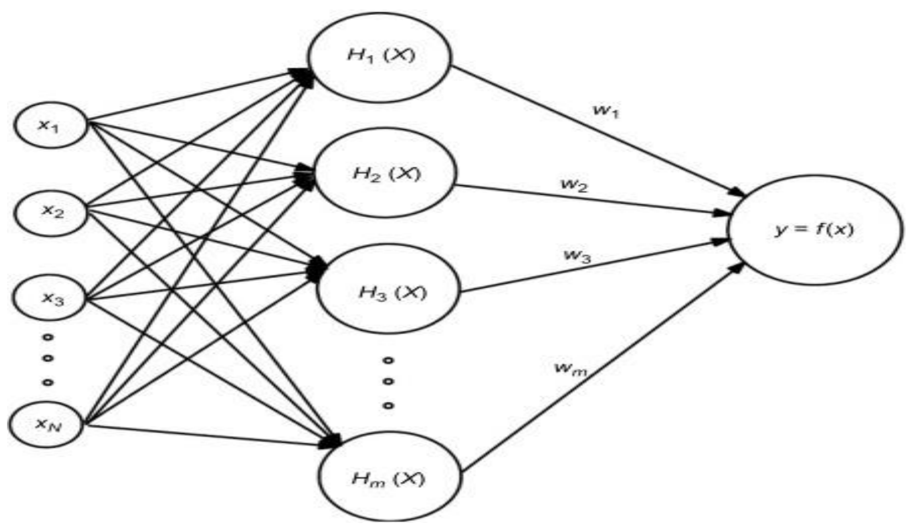


Figure 3. 5: Architecture of RBFNNs

The output summation of hidden layers with some weight of these neurons in the hidden layer is provided as the output layer and value of Radial Basis Function Neural Networks (RBFNNs). The detection of neurons number in the hidden layer is one of the most problematic tasks because under-fitting and over-fitting problems may occur due to neuron numbers in the hidden layer. The under-fitting problem means that the network is not able to proper pattern recognition. On the other hand, another problem called the over-fitting problem means that the network leads to poor generalization. Eq. (3.27) gives the formula for radial basis functions by using Gaussian basis functions.

$$H_j(x) = e^{\frac{\|x-c_j\|^2}{\sigma^2}} \quad 3.27$$

Gaussian Function:

Radial Basis Function Neural Networks (RBFNN) can be trained in several ways, the gradient distance approach is the most important method for training and describing Back-Prorogation (BP) algorithm, but the speed of the training in RBFNN is very high as compared to Multilayer Perceptron Neural Network (MLPNN) with back-propagation.

3.3.8 Recurrent Neural Networks (RNNs)

A recurrent neural networks (RNNs) is a type of Artificial Neural Networks (ANNs) that uses time-series data or sequential data. RNNs use training data for learning, Recurrent Neural Networks (RNNs) output based on the prior parameters in order, while other deep learning neural networks assume that outputs and inputs are not dependable and share parameters across each layer of the network is a feature from recurrent networks features, Recurrent Neural Networks share the same weight elements within each layer of the network, while feed forward networks have different weights across each node. These weights are still adjusted through the processes of back propagation and gradient descent to facilitate reinforcement learning [58].

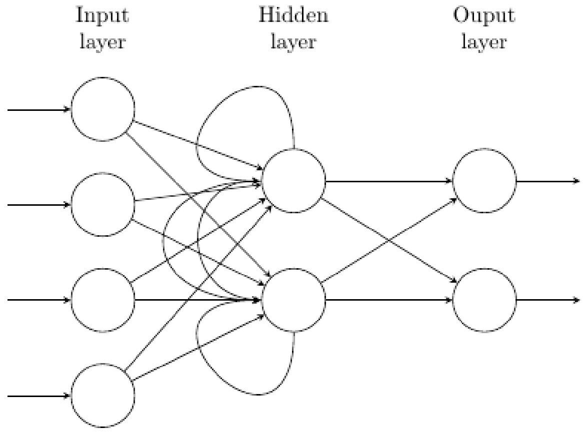


Figure 3. 6: Recurrent Layer Algorithm.

Order is defined as a table of (x_i, y_i) pairs, these parameters represent the following sequence: x_i is the input at a time i, while y_i is the Desired output. In the dataset, each time step has additional input: the previous time step in the hidden state h_{i-1}, knowing that at the time i the Recurrent Network has; vector of input is x_i, while vector of output is y_i and $\hat{y}_i$ is the vector of predicted output and hidden layer is h_i. The Recurrent Network being a simple one-hidden-layer feed-forward network at a single time step, knowing that at the time i the Recurrent Network has; the vector of input is x_i , while vector of output is y_i and $\hat{y}_i$ is the vector of predicted output and hidden layer is h_i. Therefore, it has three types of spate matrices of weights: Input-to-hidden weights w_{hx} , Hidden-to-hidden weights w_{hh} , Hidden-to-output weights w_{yh} .the forward propagation equations(3.28, 3.29)for this network are:

$$h_i = \sigma(W_{hh}h_{i-1} + W_{hx}x_i + b_h) \tag{3.28}$$

$$\hat{y}_i = W_{yh}h_i \tag{3.29}$$

Algorithm for Prediction: MLPNNS, RNNs, and RBF

Input: Train dataset, Test Dataset, Levenberg-Marquardt algorithm;
Output: Train and test result;
Data Preprocessing Phase:
Step1: Normalize the dataset samples between 0 and 1
for subset **in** dataset **do**
for an item **in** subset **do**

$$\text{subset[item]} \leftarrow \frac{\text{subset[item]} - \min(\text{subset})}{\max(subset) - \min(subset)}$$

MLPNNs Training Phase:
Get optimal Parameters
Get Train dataset
While MSE less than the threshold
While Termination condition not satisfied
Calculate prediction result
Calculate MSE
Update weights
Number of neurons= Number of neurons+2
Output MSE
Output prediction result
Save network parameters
MLPNNs Testing Phase:
Get network parameters from the training phase
Get Test dataset
Calculate test prediction result
Output MSE
Output test prediction result
RBFNNs Training Phase:
Get optimal Parameters
Get Train dataset
While MSE less than the threshold
While Termination condition not satisfied
Calculate prediction result
Calculate MSE
Update weights
Number of neurons= Number of neurons+2
Output MSE
Output prediction result
Save network parameters
RBFNNs Testing Phase:
Get network parameters from the training phase
Get Test dataset
Calculate test prediction result
Output MSE
Output test prediction result
RNNs Training Phase:
Get optimal Parameters
Get Train dataset

While MSE less than the threshold
While Termination condition not satisfied
Calculate prediction result
Calculate MSE
Update weights
Number of neurons= Number of neurons+2
Output MSE
Output prediction result
Save network parameters

RNNs Testing Phase:
Get network parameters from the training phase
Get Test dataset
Calculate test prediction result
Output MSE
Output test prediction result.

Comparison MSE Results:
Get the optimal number of neurons has minimum MSE.
Get MSE for each algorithm.
While the MSE of each algorithm has the lowest MES.
Output MSE.
Choose this algorithm.

3.3.9 Levenberg Marquardt

The Levenberg-Marquardt (LM) algorithm [59] is one of the most efficient training algorithms, it is an iterative technique that solves the problems existing in both the Gauss-Newton method for neural-networks training and gradient descent method done by the combination of those two algorithms, and can be thought of as a combination of steepest descent and the Gauss-Newton method [60]. LM algorithm has its flaws, it has many problems, one of these problems is the Hessian matrix inversion, the HM needs to be calculated each time to update the weights, and may have several updates is each time. For networks that have a small size, the computation is efficient, otherwise for the large size is not efficient, like image processing problems, this calculation of inversion is going to be damaged and the speed gained by second-order approximation may be wasted, in this case, LM maybe even slower than the steepest descent algorithm. Jacobian matrix is another problem that has to be saved for computation, and its size is P × M × N, where: P: is the number of patterns, M: is the number of outputs, N: is the number of weights.

When training patterns have a large size, the cost of memory may be too much to be practical for Jacobian matrix, also the LM algorithm was implemented for multilayer perceptron neural networks (MLPNNs) and Recurrent Neural Networks (RNNs). When making a combination between the Gauss-Newton algorithm and the steepest descent algorithm, during the training process, the LM algorithm switches between the two algorithms. Gauss-Newton algorithm is used when the combination coefficient μ is very small, Equation 3.30 is approaching Equation 3.31. Steepest descent algorithm is used when the combination coefficient μ is very large, Equation 3.30 approximates to Equation 3.32. When the combination coefficient μ in Equation 3.30 is very big, it can be interpreted as the learning coefficient in the steepest descent method (3.32):

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - (\boldsymbol{J}_K^T \boldsymbol{J}_k + \mu \boldsymbol{I})^{-1} \boldsymbol{J}_k \boldsymbol{e}_k \qquad 3.30$$

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - (\boldsymbol{J}_K^T \boldsymbol{J}_k)^{-1} \boldsymbol{J}_k \boldsymbol{e}_k \qquad 3.31$$

Where $\boldsymbol{J}$ is the Jacobian matrix and $\boldsymbol{e}$ is the error vector, $\boldsymbol{I}$ is the identity matrix and μ is always positive, called combination coefficient.

$$w_{k+1} = w_k - \alpha \boldsymbol{g}_k \qquad 3.32$$

Where α is learning rate, while $\boldsymbol{g} = \frac{\partial\, E(\boldsymbol{x},\boldsymbol{w})}{\partial\, \boldsymbol{w}}$ is the first-order derivative of total error.

3.3.10 Matlab Software

In this research, Matlab software has been used which has features that enable to make predictions and classifications using many algorithms available in this software, and also in recent versions of this software. The designers have added new tools that enable them to complete the work and application easily and extract results beautifully and terribly. This software is easy to deal with in terms of design, arrangement, entering the necessary data, extracting the necessary data, and showing the results. Many algorithms have been used in classification and prediction, and it has shown effective results and has been relied upon in building the system completely and extracting the important data necessary for classification or prediction. In the beginning, it experiments with important tools in the prediction process, we sed many tools were used like MLPNNs, RBFNNs, and RNNs. One of the three most important tools was devised for the prediction process. A comparison between the three tools and these tools has been made showing intense competition and showing close results, but the tool that got the least error was taken.

3.3.11 Weka Software

As for the Weka tool, WEKA is an open-source software package that contains a set of algorithms that aid in data mining. These algorithms can be easily applied to a set of data either directly through the WEKA program interface, or by invoking them (Java code) using their classes. By downloading the library of WEKA, WEKA software has been used, this software is very important and it is very easy to deal with so that a classification process can be made based on many algorithms. Three algorithms have been made and compared the results so that the results were shown with knowledge of the factors affecting the compressive strength of concrete (CCS) in the Palestinian governorates by making a comparison between results that were obtained from K-Mean, KSOM, and EM Algorithms. The first step in analyzing using WEKA is to use data with a format that can be understood by WEKA. One of these formulas is a formula. ARFF. The word “ARFF” stands for Attribute-Relation File Format. It is one of the file formats. WEKA deals with data analysis. It could be said that a file (ARFF) is like a table containing a set of data represented by several columns and rows. The columns represent attributes or attributes while the rows represent instances or models. One model consists of several characteristics.

3.4 Classification Metrics Selection

There are many parameters linked with technique “pattern recognition and classification” and mathematically measure its performance [60]. This research will focus on the following mercies: Accuracy, Sensitivity, Specificity, Confusion Matrix, True positive (TP), False positive (FP), False negative (FN), and True negative (TN). In the following paragraphs, the definitions of these terms according to the prediction of the problem are:

TP: The number of samples correctly classified as Actual.
FP: The number of samples incorrectly classified as Actual.
TN: The number of samples correctly classified as Predicted.
FN: The number of samples incorrectly classified as Predicted.

Confusion matrix: is a table that is used to show the results of the classification model. The table consists of two-dimensional, each column represents the actual values, and each row represents the predicted values. It is used to calculate most of the performance measures. The following table describes the confusion matrix for prediction models to diagnose the problem.

Table 3. 1: Confusion matrix description for Compressive strength of concrete.

		Predicted Classes		
		Actual value	Predicted value	Total value
Actual Classes	Actual value	TP	FP	TP+FP
	Predicted value	FN	TN	FN+TN

	Total value	TP+FN	FP+TN	TP+FP+FN+TN

Accuracy: The main metric that utilized to calculate the performance in pattern recognition and classification model, which represented by this equation:

$$Accuracy = \frac{\mathrm{TP} + TN}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \qquad 3.\ 33$$

Sensitivity: The percentage of records that classified correctly as Actual to all records that classified as Actual as represented by the following formula.

It is the percentage of records that are predicted to a certain class correctly to all records predicted in that class. It is calculated using the following equation:

$$Recall\ = \frac{T\mathrm{P}}{T\mathrm{P} + F\mathrm{N}} \qquad 3.\ 34$$

Specificity: The percentage of records correctly predicted as Predicted to all records predicted in the Predicted class.

$$Specificity = \frac{TN}{TN + \mathrm{FP}} \qquad 3.\ 35$$

Precision: The percentage of records correctly predicted as actual to all records predicted in Actual class.

$$Precision = \frac{TP}{TP + \mathrm{FP}} \qquad 3.\ 36$$

ROC curves [60] is short of Receiver-Operating Characteristic Analysis and this represented as logistic regression which is utilized for detecting the best values of the cutoff for predicting which has a new observation like 0 called failure and 1 called a success, and this is used for showing the classification model performance at different probability thresholds. The ROC curve will be more flexible when predicting the class label's probability instead of the class labels itself. By using class label's probabilities, it can calibrate thresholds values, by default in logistic regression, the value 0.5 is represented as a probability threshold, and any values between this range [0.0 – 0.49] is represented as a negative label and any values between this range [0.5-1.0] are represented as a positive label. It could be optimized this probability threshold and may have better results. This graph is represented by plotting False Positive Rate (FPR) plotted on the x-axis versus True Positive Rate (TPR) plotted on the y-axis for different values between [0.0-1.0] and these values called probability threshold values, while TPR is the ratio of correctly predicted positive labels from all the positive labels and FPR is the ratio of incorrectly predicted positive labels from all the negative labels. (AUC) is a short area under curve, It calculates the entire two-dimensional area underneath the entire ROC curve from [(0, 0) – (1, 1)]. The ROC curve is a representing sensitivity versus (1-specificity), sensitivity is the true positive rate, and (1-specificity) is the false positive rate, the network is very good when 100% sensitivity and 100% specificity. The classification results obtained True Positive(TP), False Positive(FP), True Negative (TN), and False Negative (FN). The percentage for testing accuracy obtained from these parameters. A perfect classifier when the curve has AUC = 1 and a completely random classifier has AUC = 0.5. This range [0, 1] of AUC values usually the model will score value between these ranges.

Chapter 4

Experiments and Results

Experiments and Results

4.1.1 Results of Extrapolating Factors Affecting CCS in Palestinian Governorates.

The factors determining work was divided into two parts, the first one was the implementation of algorithms to the entire combined data of the Palestinian governorates, and the second one was the implementation of algorithms to all Palestinian governorates separately showing the influencing factors in the governorates as a whole and showing each governorate separately, respectively. The results showed that the EM algorithm is completely identical to the KSOM algorithms it depends on the standard deviation of the input that was entered so that mathematically, according to a special analysis of the standard deviation in the algorithms, then account these factors are considered and considering them the factors affecting the strength of the concrete's compressive as shown in the tables below. All the algorithms were applied to the combined data collected from all governorates so that each algorithm shows its results, and it can be deduced that the EM and KSOM algorithms are completely similar, and thus it can be relied on these two algorithms to determine the main factors that affect Palestinian governorates Concrete Compressive Strength (PCCS).

The Computing Environment in this work is Dell Latitude E5430 g: I5-3210M 2.50GHz, RAM: 12GB, HD: 256 SSD with windows 10 pro. Based on previous studies and special analysis in each algorithm, it is possible to rely on the standard deviation in determining the factors that affect concrete Compressive Strength, so that in the EM and KSOM algorithms, when the data is divided into a certain number of clusters, the largest value of the concrete Compressive Strength that was taken, and the clusters that are taken and completed so that the specific results that have been determined in the cluster of the largest value are taken into consideration, if the standard deviation is close to zero or turns to zero, then this factor is considered as a factor affecting the Compressive

Strength of the concrete. But in the K-Means algorithm, the data were divided on a certain number of clusters, then the largest value of the concrete compressive strength is taken, and the whole cluster is also taken so that it is looked at the specific results that have been determined in the cluster of the largest value if the recorded value is greater than the original value in the first column, then its factors influencing concrete Compressive Strength are considered. This analysis was based on previous studies as mentioned in the research paper [20].

The dataset is first selected, and then implement these algorithms on combined data of the Palestinian governorates, In general, the dataset consists of 8 input parameters (Fine Aggregates, Coarse Aggregates, Super plasticizer, Water, Cement, W/C ratio, Age, and Location) were examined against Concrete Compressive Strength (CCS) using these algorithms (EM, KSOM, K-Means). This section compares and evaluates these algorithms using combined datasets from the Palestinian governorate to investigate the most important factors that affect the concrete component mix. Table A.1 shows the main results of EM algorithms. To extract these results, different datasets were used with different numbers of clusters (k=3, 4, 5, 6, 7, 8, 9, 10, 20, and 50) as represented in table A. 1.

Table A. 1: ALL Governorates of Palestine “EM Results”:

Number of Clusters	**Main factors that affect Concrete Compressive Strength**
K = 3	W/C ratio, SP, Location, Age
K= 4	W/C ratio, SP, Location, Age
K = 5	W/C ratio, SP, Location, Age
K=6	W/C ratio, SP, Location, Age
K = 7	W/C ratio, SP, Location, Age
K = 8	W/C ratio, SP, Location, Age
K= 9	W/C ratio, SP, Location, Age
K=10	W/C ratio, SP, Location, Age
K = 11	W/C ratio, SP, Location, Age
K = 20	W/C ratio, SP, Location, Age
K=50	W/C ratio, SP, Location, Age

For each cluster as mentioned above, mathematical values are used, called standard deviation. In our method, this parameter is used to detect the main influence factors that affect the Compressive Strength of concrete. For example, when K=10, it has four factors (W/C ratio, Age, Location, Super plasticizer). After the implementation of the EM algorithm for detecting the main factors that affect Concrete Compressive Strength (CCS) and analyzing the results extracted from the EM algorithm, the results show that the EM model archives the optimal mix of concrete components. After applying EM on the combined dataset of the Palestinian governorate many times with several clusters, the main parameters were detected on their standard deviation values. Table A. 2 represents the main factors that affect Concrete Compressive Strength and their standard deviation values.

KSOM algorithm is also used for detecting the main factors that affect Concrete Compressive Strength after applying KSOM on the combined dataset of Palestinian governorates many times with several clusters. The main parameters are detected on their standard deviation values. Table A. 3 represents the main factors that affect Concrete Compressive Strength and these parameters (SP, W/c ratio, Age, Location) were obtained from this algorithm.

Table A. 2: List of the main factors with their standard deviations by EM.

Number of Clusters	**Standard Deviation**	**Main factors that affect Concrete Compressive Strength**
K = 3	1.3022 0.0483 0 1.9566	SP W/C ratio Age Location
K= 4	1.401 0.0383 0.0001 1.7236	SP W/C ratio Age Location
K = 5	1.2715 0.0516 0.001 0.4927	SP W/C ratio Age Location

K=6	1.382 0.0457 0.031 0.5921	SP W/C ratio Age Location
K = 7	1.2532 0.043 0.0896 0.8399	SP W/C ratio Age Location
K = 8	1.2790 0. 77 0.0196 0.7301	SP W/C ratio Age Location
K= 9	1.2837 0.0514 0.5156 0.4934	SP W/C ratio Age Location
K=10	1.3837 0.0414 0.3152 0.4358	SP W/C ratio Age Location
K = 11	1.4033 0.0224 0.1414 1.0054	SP W/C ratio Age Location
K = 20	1.5033 0.0234 0.2414 1.0054	SP W/C ratio Age Location
K=50	1.7544 0.0555 0.3196 0.7298	SP W/C ratio Age Location

Table A. 3: All Governorates of Palestine “ KSOM Results”.

Name	**KSOM**
Number of Clusters	**Main factors that affect Concrete Compressive Strength**
k=3	SP, W/c ratio, Age, Location
K=4	SP, W/c ratio, Age, Location
K = 5	SP, W/c ratio, Age, Location
K=6	SP, W/c ratio, Age, Location
K = 7	SP, W/c ratio, Age, Location
K=8	SP, W/c ratio, Age, Location
K= 9	SP, W/c ratio, Age, Location
K=10	SP, W/c ratio, Age, Location
K = 11	SP, W/c ratio, Age, Location
K=50	SP, W/c ratio, Age, Location

K-mean algorithm is also used for detecting the main factors that affect Concrete Compressive Strength, after applying K-Means on the combined dataset of Palestinian governorates many times with different numbers of clusters as shown in Table A. 4. The main parameters are detected on their values and Table A. 5 represents the main factors that affect Concrete Compressive Strength and these parameters (Coarse Aggregates, Fine Aggregates, SP, w/c ratio, Location, and Age) were obtained from this algorithm. Table A. 6 represents the comparison between the three algorithms to detect the main factors that affect the Compressive Strength of concrete. K-Mean algorithm produces these parameters (Coarse Aggregates, Fine Aggregates, SP, w/c ratio, Location, and Age) and according to the results which were obtained in EM and KSOM algorithms (SP, w/c ratio, Location, and Age), K-Mean shows distinguished factors which are the (Coarse Aggregates and Fine Aggregates).It is clear that the three algorithms show intersection and provide different results and the analysis concludes that these factors (SP, w/c ratio, Location, and Age) are common and they are the four primary components that affect Concrete Compressive Strength).

Table A. 4: Results for K-Means algorithm based on different numbers of clusters (K=3, 5, 7 and 9).

Number of clusters	Result for K=3, 5, 7 ,9				
K=3			Cluster#		
	Attribute	Full Data (500.0)	0 (228.0)	1 (185.0)	2 (87.0)
	Coarse agg	1189.8952	1139.8596	1189.3924	1322.092
	Fine Aggegate	643.7344	680.8561	647.7946	537.8161
	Water	158.172	158.4386	157.8757	158.1034
	Cement	322.646	321.7982	321.0973	328.1609
	SP	1.9357	2.0587	1.9365	1.6156
	w/c Ratio	0.4955	0.4955	0.4964	0.4933
	Age	13.594	5.2105	28	4.931
	Location	4.066	4.9868	4.1189	1.5402
	CS	226.5513	175.211	313.8732	175.4138

K=5

Attribute	Full Data (500.0)	Cluster# 0 (103.0)	1 (174.0)	2 (87.0)	3 (58.0)	4 (78.0)
Coarse agg	1189.8952	1181.833	1207.8299	1322.092	932.8655	1204.2077
Fine Aggegate	643.7344	606.4019	626.1908	537.8161	954.8517	618.9641
Water	158.172	167.301	157.9023	158.1034	159.431	145.859
Cement	322.646	371.5146	322.408	328.1609	291.8621	275.3846
SP	1.9357	1.4769	1.6963	1.6156	4.8069	1.3086
w/c Ratio	0.4955	0.4503	0.4965	0.4933	0.5359	0.5296
Age	13.594	4.9806	28	4.931	9.8793	5.2564
Location	4.066	5.3495	4.1437	1.5402	3.4138	5.5
CS	226.5513	199.8373	328.7086	175.4138	141.2831	158.1283

K=7

Attribute	Full Data (500.0)	Cluster# 0 (53.0)	1 (87.0)	2 (87.0)	3 (50.0)	4 (78.0)	5 (91.0)	6 (54.0)
Coarse agg	1189.8952	1202.7509	1182.4299	1322.092	1159.66	1204.2077	1221.4286	930.5037
Fine Aggegate	643.7344	600.7434	674.2092	537.8161	612.4	618.9641	594.1429	955.8407
Water	158.172	163.2075	144.7931	158.1034	171.64	145.859	170.7582	159
Cement	322.646	372	268.2184	328.1609	371	275.3846	371.8571	293.5741
SP	1.9357	2.7004	1.9128	1.6156	0.18	1.3086	1.9558	4.7407
w/c Ratio	0.4955	0.4379	0.5339	0.4933	0.4634	0.5296	0.4991	0.53
Age	13.594	5.0377	28	4.931	4.92	5.2564	28	8.537
Location	4.066	5.6038	4.5057	1.5402	5.08	5.5	4.8352	3.2963
CS	226.5513	202.426	256.1974	175.4138	197.0932	158.1283	386.1516	142.008

K=9

Attribute	Full Data (500.0)	Cluster# 0 (55.0)	1 (71.0)	2 (49.0)	3 (45.0)	4 (68.0)	5 (62.0)	6 (54.0)	7 (45.0)	8 (51.0)
Coarse agg	1189.8952	1189.1636	1155.0761	1322.0204	1162.4	1204.2647	1170	930.5037	1321.5778	1299.9608
Fine Aggegate	643.7344	605.9636	699.6085	537.9592	612.6667	610.5971	625.8548	955.8407	538.8444	563.7255
Water	158.172	162.3818	147.5493	174.449	173.6	148.4706	168.0806	159	160.8667	136.8039
Cement	322.646	371.7455	270.493	376.1224	373.3333	283.0882	372.7258	293.5741	330.2222	262.1569
SP	1.9357	2.592	2.0531	1.9347	0	1.4946	1.9719	4.7407	1.5036	1.0933
w/c Ratio	0.4955	0.4358	0.539	0.4749	0.466	0.5246	0.4991	0.53	0.4922	0.5241
Age	13.594	5.0364	28	4.9592	4.8667	5.2353	28	8.537	28	5.0392
Location	4.066	5.3818	5	1.1837	5.2	5.8529	5.1613	3.2963	1.4667	2.5098
CS	226.5513	213.0673	258.8911	196.9184	195.2369	153.0381	391.6887	142.008	328.0667	149.3784

Table A. 5: ALL Governorates of Palestine “K-Mean Results”.

Name	K-mean
Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Location , Age
K= 4	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Location , Age
K = 5	Coarse Aggregates, Fine Aggregates, SP, w/c ratio, Location , Age
K = 6	Coarse Aggregates, Fine Aggregates, SP ,w/c ratio, Location , Age
K = 7	Coarse Aggregates, Fine Aggregates, SP ,w/c ratio, Location , Age
K = 8	Coarse Aggregates, Fine Aggregates, SP ,w/c ratio, Location , Age
K = 9	Coarse Aggregates, Fine Aggregates, SP, w/c ratio, Location , Age
K = 10	Coarse Aggregates, Fine Aggregates, SP ,w/c ratio, Location , Age
K = 11	Coarse Aggregates, Fine Aggregates, SP, w/c ratio, Location , Age
K = 20	Coarse Aggregates, Fine Aggregates, SP ,w/c ratio, Location , Age
K = 50	Coarse Aggregates, Fine Aggregates, SP ,w/c ratio, Location , Age

Table A. 6: Summary of the main parameters that affect compressive strength of concrete using the three algorithms in All Governorates.

Name	EM	KSOM	K-Means	Intersection
Main parameters	SP, W/c ratio , Age , Location	SP, W/c ratio , Age , Location	Coarse Aggregates, Fine Aggregates, SP ,w/c ratio, Location , Age	SP, W/c ratio , Age , Location

Table A. 7: the relationship between the numbers of cluster, iteration, sum square error and CCS using K- Mean algorithm.

K mean	**# of iteration**	**SSE**	**CCS (Average Actual Data) = 280.84**
K = 3	4	156.3	313.8732
K= 4	12	127.5	327.0286
K = 5	12	116.146	328.7086
K = 6	15	102.409	374.1795
K = 7	11	98.3	386.1516
K = 8	9	86.622	391.6887
K = 9	8	80.55	391.6887
K = 10	16	69.8	395.4407
K = 11	16	67.5	395.4407
K = 20	9	49.38639	415.2342
K = 50	7	24.319	458.6667
Average			379.8273818

Table A. 7 represents the relationship between the numbers of clusters, iteration, sum square error, and CCS using the K- Mean algorithm, while Table A. 8 shows the prediction of Compressive Strength of Concrete (CCS) by implementation both KSOM and K-means utilize WEKA tool. It is found that the actual average of Compressive Strength of Concrete (CCS) is 280.84 and by making a comparison between the results of the Compressive Strength of Concrete (CCS) of the two algorithms, it is found to be similar between both algorithms.

Table A. 8: Results of the Compressive Strength of Concrete (CCS) of K-Mean and KSOM algorithms.

Name	**CCS (Average Actual Data) = 280.84**	
Number of Clusters	**K Mean CCS high-value prediction**	**KSOM CCS high-value prediction**
K = 3	313.8732	333.3681
K= 4	327.0286	343.9022
K = 5	328.7086	353.9022
K = 6	374.1795	351.9931
K = 7	386.1516	341.9731
K = 8	391.6887	351.9631
K = 9	391.6887	371.9731
K = 10	395.4407	391.3
K = 11	395.4407	393.0685
K = 20	415.2342	405.0412
K = 50	458.6667	420.0784
Average	**379.8273818**	**368.9593636**

The values above in Table A. 8 show that the results for the prediction of Concrete Compressive Strength (CCS) of both algorithms are very close or similar sometimes. The result shows that the K-mean algorithms can be swimmingly utilized to have a more accurate prediction for improving the Concrete Compressive Strength (from average actual data 280.84 to average prediction data 379.82). These algorithms were implemented to Palestinian Governorates datasets and compared these results were obtained from these algorithms to find the main factors that affect Palestinian Concrete Compressive Strength (PCCS) using the WEKA tool. From the analysis results, it is shown that EM and KSOM are the best accurate and effective algorithms to find these factors that affect Palestinian governorates Concrete Compressive Strength (PCCS), and they can obtain that K-Means and KSOM algorithms are effective algorithms for predicting the Palestinian Governorates Concrete Compressive Strength (PCCS). Form the results in this section, could be used to predict the effects of the main component of Palestine Concrete Compressive Strength (PCCS). Table A. 3 shows the primary factors that predict the Palestinian governorates Concrete Compressive Strength (PCCS), and these factors are (SP, w/c ratio, Location, and Age). The analysis of the information from Table A. 7 and Table A. 8 shows great links between the main factors that affect Palestinian governorates Concrete Compressive Strength (PCCS) and the prediction of Palestinian governorates Concrete Compressive Strength (PCCS) and the values for these components are very similar among these algorithms. These results provide the threshold value that increases and improves the Palestinian governorate's Concrete Compressive Strength (PCCS). Also, these factors increase the Palestinian governorate's Concrete Compressive Strength (PCCS) from average actual data 280.84 to average prediction data 379.82, and this leads to make a percentage of increasing the performance technique from 28% to 38% of PCCS. Table A. 9 shows the summary of the main factors that improve the performance of Palestinian governorates Concrete Compressive Strength (PCCS) that results were obtained from these algorithms.

Table A. 9: Summary of the main factors that improve the Performance of Concrete Compressive Strength (PCCS).

Predictive factors	EM	KSOM	K-Means	Average value	Mean
SP	1.3037	1.3052	1.3086	4.2	1.9
W/c ratio	0.043	0.0507	0.4991	0.53	495.
Age	0.0896	0	28	15.5	12.937
Location	0.8399	.08859	4.8352	4	4.024
Coarse Aggregates	-	-	1221.4286	1095.5	1189.79
Fine Aggregates	-	-	594.14	775	555.455

In general, the implementation of these algorithms and obtained results show that these algorithms are effective models for improving the prediction of the Palestinian governorate's Concrete Compressive Strength (PCCS) and detecting the main factors that affect the Palestinian governorate's Concrete Compressive Strength (PCCS). Some notes can be obtained from these results, first note, it is important to note that the cost of super plasticizer (SP) and w/c ratio is beyond Palestinian governorates Concrete Compressive Strength (PCCS) and second note represents logically, location and age are also influencing factors in the study. Here it should be noted that all the factors involved affect the compressive strength of concrete, but the study confirmed that there are main factors that affect directly, and when Artificial Neural Networks (ANNs) are designed, all the inputs must be entered, because they are all influential.

4.1.2 Detecting Main Factors that Affect CCS in Jenin Governorate:

Jenin dataset is first selected, using these algorithms (EM, KSOM, and K-Means) were implemented on the Jenin dataset to investigate the most important factors that affect the Concrete Compressive Strength (CSS), In general, the dataset of Jenin consists of 7 input parameters (Fine Aggregates, Coarse Aggregates, Super plasticizer, Water, Cement, W/C ratio, and Age) that were examined against Concrete Compressive Strength (CCS) in Jenin Governorate. Table J. 1 shows the main factors obtained from EM algorithms. To extract these results, different datasets were used with different numbers of clusters (k=3, 5, 7, and 9) as represented in Table J. 1. For each cluster as mentioned above, mathematical values are used, called standard deviation. In our method, this parameter was used to detect the main factors that affect the Compressive Strength of Concrete, for example when K=3, it has three factors (W/C ratio, Age, and Super plasticizer). Table J. 2 represents the standard deviation values of the main factors that affect CSS.

Table J. 1: Jenin Governorate “EM Results”:

Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	W/C ratio , SP , Age
K = 5	W/C ratio , SP , Age
K=7	W/C ratio , SP , Age
K = 9	W/C ratio , SP , Age

Table J. 2: list of the main factors with their standard deviation in Jenin Governorate.

Number of Clusters	Standard Deviation	Main factors that affect Concrete Compressive Strength
K = 3	1.4727 0.0392 2.2064	SP W/C ratio Age
K= 5	1.5118 0.0343 1.6189	SP W/C ratio Age
K = 7	1.5113 0.0343 1.919	SP W/C ratio Age
K = 9	0.0857 0.015 0.1479	SP W/C ratio Age

KSOM algorithm is also utilized for detecting the main factors that affect Concrete Compressive Strength, after running KSOM on the Jenin dataset several times with different numbers of clusters and the main parameters are identified based on their standard deviation values, similar to the EM measures. Table J. 3 shows the main factors that affect Concrete Compressive Strength and these parameters (Coarse Aggregates, Fine Aggregates, SP, W/c ratio, and Age).

Table J. 3: Jenin Governorate dataset "KSOM Results".

Name	**KSOM**
Number of Clusters	**Main factors that affect Concrete Compressive Strength**
k=3	Coarse Aggregates, Fine Aggregates, SP ,w/c Ratio , Age
K=5	Coarse Aggregates, Fine Aggregates, SP ,w/c Ratio , Age
K = 7	Coarse Aggregates, Fine Aggregates, SP ,w/c Ratio , Age
K=9	Coarse Aggregates, Fine Aggregates, SP ,w/c Ratio , Age

K-mean algorithm is also used for detecting the main factors that affect Concrete Compressive Strength after applying K-Means to the Jenin governorate dataset several times with different numbers of clusters, as shown in Table J. 4 with their values. The main factors that affect Concrete Compressive Strength are identified. Table J. 5 represents the main factors that affect Concrete Compressive Strength and these parameters (Coarse Aggregates, Fine Aggregates, SP, w/c ratio, and Age).

Table J. 4: Results for K-Means algorithm based on some different number of clusters (K=3, 5, 7 and 9) in Jenin Governorate.

Number of clusters	Result for K=3, 5 , 7 ,9

K=3

Final cluster centroids:

Attribute	Full Data (71.0)	Cluster# 0 (18.0)	1 (39.0)	2 (14.0)
Coarse agg	1276.2254	1126.7222	1327	1327
Fine Aggegate	584.6761	751.5556	528	588
Water	174.5211	165.3333	175.7436	172.9286
Cement	351.9718	275	375.8974	344.2857
SP	2.3789	4.4556	1.8051	2.5071
w/c Ratio	0.5011	0.5906	0.4687	0.6764
Age	11.7465	14.0556	4.8462	28
CS	193.8031	57.9456	192.9231	370.9286

K=5

Final cluster centroids:

Attribute	Full Data (71.0)	Cluster# 0 (15.0)	1 (26.0)	2 (8.0)	3 (15.0)	4 (7.0)
Coarse agg	1276.2254	1086.6667	1327	1327	1327	1327
Fine Aggegate	584.6761	796.2667	528	528	528	528
Water	174.5211	171.8	163.0769	190.625	169.6	193.5714
Cement	351.9718	280	360	381.25	345.3333	392.8571
SP	2.3789	4.9467	2.8615	0	2.3833	0
w/c Ratio	0.5011	0.5967	0.4596	0.5	0.682	0.4929
Age	11.7465	14.3333	4.8462	3	28	7
CS	193.8031	33.5347	189.4231	140.5	363.5333	250.7143

K=7

Final cluster centroids:

Attribute	Full Data (71.0)	Cluster# 0 (7.0)	1 (6.0)	2 (3.0)	3 (14.0)	4 (6.0)	5 (9.0)	6 (26.0)
Coarse agg	1276.2254	1141.2857	1327	1327	1327	1327	1070.8889	1327
Fine Aggegate	584.6761	737.2857	528	528	588	528	812.3333	528
Water	174.5211	166.7143	182	204	172.9286	196	171.4444	163.0769
Cement	351.9718	275.7143	366.6667	400	344.2857	400	280	360
SP	2.3789	4.3929	0	0	2.3871	0	5.05	2.8615
w/c Ratio	0.5011	0.5957	0.4967	0.51	0.6764	0.49	0.5933	0.4596
Age	11.7465	28	3.6667	3	28	7	5.2222	4.8462
CS	193.8031	74.5257	157.6667	134.6667	370.9286	254.8333	26.8156	189.4231

K=9

Final cluster centroids:

Attribute	Full Data (71.0)	Cluster# 0 (7.0)	1 (5.0)	2 (3.0)	3 (14.0)	4 (6.0)	5 (8.0)	6 (20.0)	7 (2.0)	8 (6.0)
Coarse agg	1276.2254	1141.2857	1317.2	1327	1327	1327	1045	1327	1327	1327
Fine Aggegate	584.6761	737.2857	542.8	528	588	528	838.625	528	528	528
Water	174.5211	166.7143	177.2	204	172.9286	196	171.625	161.6	188	168
Cement	351.9718	275.7143	336	400	344.2857	400	280	348	400	400
SP	2.3789	4.3929	0.45	0	1.3871	0	5.4	2.76	0	3.2
w/c Ratio	0.5011	0.5957	0.53	0.51	0.6764	0.49	0.5912	0.4715	0.47	0.42
Age	11.7465	28	3.8	3	28	7	5.5	4.8	3	5
CS	193.8031	74.5257	129.436	134.6667	370.9286	254.8333	26.895	174.2	162.5	240.1667

Table J. 5: Main factors that affect Concrete Compressive Strength by K-Mean Algorithm in Jenin Governorate.

Name	**K-mean**
Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Age
K= 5	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Age
K = 7	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Age
K = 9	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Age

Table J. 6: Summary of the main parameters that affect Compressive Strength of concrete using the three algorithms in Jenin Governorate.

Name	**EM**	**KSOM**	**K-Means**	**Intersection**
Main parameters	SP, W/c ratio , Age	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Age	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Age	SP, W/c ratio , Age

Table J. 6 represents a comparison of the three algorithms for detecting the main factors that affect Concrete Compressive Strength. The study concludes that these factors (SP, w/c ratio, and Age) are typical and they are the three primary components that affect Concrete Compressive Strength in the Jenin governorate.

4.1.3 Detecting Main Factors that Affect CSS in Ramallah Governorate:

The Ramallah Governorate dataset was chosen first and then these algorithms were applied to the Ramallah dataset. The dataset, which contains seven input parameters (Fine Aggregates, Coarse Aggregates, Superplasticizer, Water, Cement, W/C ratio, and Age) were compared to Concrete Compressive Strength (CCS) in Ramallah Governorate using these algorithms (EM, KSOM, K-Means). This section compares and tests these algorithms using the Ramallah Governorate dataset to find the main factors that affect the concrete mix. Table R. 1 displays the key results of EM algorithms, various datasets with different numbers of clusters (k=3, 5, 7, and 9) were used to obtain these results as shown in table R. 1. After applying EM on the Ramallah

Governorate dataset several times with different numbers of clusters, the main parameters were detected on their standard deviation values. Table R. 2 represents the main factors that affect Concrete Compressive Strength and their standard deviation values.

Table R. 1: Main factors that affect Concrete Compressive Strength in Ramallah Governorate by EM.

Number of Clusters	**Main factors that affect Concrete Compressive Strength in Ramallah Governorate By EM Algorithm**
K = 3	W/C ratio , SP , Age
K = 5	W/C ratio , SP , Age
K=7	W/C ratio , SP , Age
K = 9	W/C ratio , SP , Age

Table R. 2: List of the main factors with their standard deviation in Ramallah Governorate.

Number of Clusters	**Standard Deviation**	**Main factors that affect Concrete Compressive Strength**
K = 3	1.2764 0.0312 1.0462	SP W/C ratio Age
K= 5	1.7264 0.0312 0.369	SP W/C ratio Age
K = 7	0.6507 0.0191 1.0462	SP W/C ratio Age
K = 9	0.2254 0.0147 1.0462	SP W/C ratio Age

Table R. 3 shows the main factors that affect Concrete Compressive Strength produced by KSOM clustering and KSOM is utilized to detect the main factors that affect Concrete Compressive Strength, After applying KSOM on the dataset of Ramallah governorate several times with different numbers of clusters, the main factors are detected based on their standard deviation values, these parameters (SP, W/c ratio, and Age) were obtained from this algorithm.

Table R. 3: Main factors that affect Concrete Compressive Strength by KSOM algorithm In Ramallah Governorate.

Name	**KSOM**
Number of Clusters	**Main factors that affect Concrete Compressive Strength by KSOM algorithm In Ramallah Governorate.**
k=3	SP ,w/c Ratio , Age
K=5	SP ,w/c Ratio , Age
K = 7	SP ,w/c Ratio , Age
K=9	SP ,w/c Ratio , Age

Table R. 5 represents the main factors that affect concrete compressive strength obtained from the K-mean clustering algorithm and the K-mean algorithm is another algorithm that used for detecting the main factors that affect Concrete Compressive Strength, and these factors (Coarse Aggregates, Fine Aggregates, SP, w/c ratio and Age) were obtained from this algorithm after applying K-Means on the dataset of Ramallah governorate many times with different numbers of clusters as shown in Table R. 4 with their values.

Table R. 4: Main factors that affect Concrete Compressive Strength by K-Mean Algorithm in Ramallah Governorate.

Name	K-mean
Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	Coarse Aggregates, water , cement, SP , W/C ratio, Age
K= 5	Coarse Aggregates, water , cement, SP , W/C ratio, Age
K = 7	Coarse Aggregates, water , cement, SP , W/C ratio, Age
K = 9	Coarse Aggregates, water , cement, SP , W/C ratio, Age

Table R. 5: Results for K-Means algorithm based on some different numbers of clusters (K=3, 5, 7 and 9) in Ramallah Governorate.

Number of clusters	Result for K=3, 5 , 7 ,9

K=3

Final cluster centroids:

Attribute	Full Data	Cluster# 0	1	2
	(102.0)	(49.0)	(27.0)	(26.0)
Coarse agg	1225.2255	1173.3469	1375	1167.4615
Fine Aggegate	592.8922	617.9592	525	616.1538
Water	161.3824	162.6531	156.6296	163.9231
Cement	346.6569	354.0816	324.4074	355.7692
SP	1.6769	1.4286	2.5941	1.1923
w/c Ratio	0.4675	0.461	0.4844	0.4683
Age	12.6275	4.8776	11.8889	28
CS	257.7112	197.8069	253.1222	375.3731

K=5

Final cluster centroids:

Attribute	Full Data	Cluster# 0	1	2	3	4
	(102.0)	(23.0)	(14.0)	(22.0)	(25.0)	(18.0)
Coarse agg	1225.2255	1179.1304	1317.8571	1166.0909	1165.56	1367.2222
Fine Aggegate	592.8922	618.6957	549.2857	617.2727	616.4	531.3889
Water	161.3824	161.1304	175.8571	162.8182	165.36	143.1667
Cement	346.6569	369.5652	393.8571	347.7273	344	283.0556
SP	1.6769	2.5652	3.1486	0.8273	0.36	2.2644
w/c Ratio	0.4675	0.4343	0.4757	0.47	0.4832	0.5022
Age	12.6275	4.7391	19.9286	28	4.92	8.9444
CS	257.7112	207.7913	350.0714	366.5727	188.9536	212.1056

K=7

Final cluster centroids:

Attribute	Full Data	Cluster# 0	1	2	3	4	5	6
	(102.0)	(18.0)	(6.0)	(26.0)	(25.0)	(5.0)	(10.0)	(12.0)
Coarse agg	1225.2255	1155	1375	1167.4615	1165.56	1266	1375	1363.3333
Fine Aggegate	592.8922	610	525	616.1538	616.4	650	525	534.5833
Water	161.3824	162	169.5	163.9231	165.36	158	172.5	134.75
Cement	346.6569	375	364.8333	355.7692	344	350	374	256.6667
SP	1.6769	2.5	2.9167	1.1923	0.36	2.8	2.99	2.0533
w/c Ratio	0.4675	0.43	0.4783	0.4623	0.4832	0.45	0.465	0.5117
Age	12.6275	4.7778	28	28	4.92	4.6	5.4	8.8333
CS	257.7112	210.0111	393.9167	375.3731	188.9536	199.8	236.23	191.5

K=9

Final cluster centroids:

Attribute	Full Data	Cluster# 0	1	2	3	4	5	6	7	8
	(102.0)	(9.0)	(11.0)	(20.0)	(20.0)	(3.0)	(2.0)	(7.0)	(15.0)	(15.0)
Coarse agg	1225.2255	1175	1282.4545	1256.1	1173.2	1266	1266	1375	1141.6667	1375
Fine Aggegate	592.8922	610	578.6364	614	618	650	650	525	612	525
Water	161.3824	170	169.8182	163.3	169.2	158	158	129.5714	151.0667	163.3333
Cement	346.6569	400	381.7273	347.5	355	350	350	238.5714	326.6667	346
SP	1.6769	3.2	3.0527	1.6810	0	2.8	2.8	1.9086	1.8133	2.7667
w/c Ratio	0.4675	0.42	0.4445	0.472	0.479	0.45	0.45	0.5171	0.4653	0.4767
Age	12.6275	4.7778	28	28	4.8	3	7	14.8571	5.1333	5.1333
CS	257.7112	222.2667	405.6636	365.18	187.492	154	268.5	220.6429	196.22	218.9067

Table R. 6: Summary of the main parameters that affect Compressive Strength of concrete using the three algorithms in Ramallah Governorate.

Name	EM	KSOM	K-Means	Intersection
Main parameters	SP, W/c ratio , Age	SP, W/c ratio , Age	Coarse Aggregates, water , cement, SP , W/C ratio, Age	SP, W/c ratio , Age

K-Mean algorithm produces these factors (Coarse Aggregates, Water, Cement, w/c ratio, and Age) and according to the results obtained from EM algorithms (SP, w/c ratio and Age) and KSOM (SP, w/c ratio and Age), K-Mean shows distinguished factors which are the (Coarse Aggregates, Fine Aggregates, Water, and Cement). Table R. 6 represents the comparison between three algorithms to detect the main factors that affect the Compressive Strength of concrete, it is clear that the three algorithms show intersection and provide different results and the analysis concludes these factors (SP, w/c ratio are Age) are common factors, and they are the factors that affect Concrete Compressive Strength.

4.1.4 Detecting Main Factors that Affect CSS in Tubas Governorate:

Tubas Governorate dataset is first selected, and then these algorithms (EM, KSOM, and K-Means) were implemented on Tubas Governorate dataset to detect the main factors that affect Concrete Compressive Strength. Table T. 1 shows the main factors that affect CSS by EM algorithms in Tubas. To extract these results, different datasets were used with different numbers of clusters (k=3, 5, 7, and 9) as represented in Table T. 1. After the implementation of the EM algorithm on the Tubas Governorate dataset many times with different numbers of clusters, the main parameters were detected on their standard deviation values and table T. 2 represents the standard deviation values of the main factors that affect Concrete Compressive Strength. The dataset consists of 7 input parameters as same each governorate was examined against Concrete Compressive Strength (CCS) in Tubas Governorate.

Table T. 1: Main factors that affect Concrete Compressive Strength in Tubas Governorate by EM Algorithm.

Number of Clusters	**Main factors that affect Concrete Compressive Strength in Tubas Governorate By EM Algorithm**
K = 3	SP ,W/C Ratio
K = 5	SP ,W/C Ratio
K=7	SP ,W/C Ratio , Coarse Aggregates, Fine Aggregates
K = 9	SP ,W/C Ratio

Table T. 2: List of the main factors with their standard deviation in Tubas Governorate.

Number of Clusters	**Standard Deviation**	**Main factors that affect Concrete Compressive Strength**
K = 3	0.2622 0.0083	SP W/C ratio
K= 5	0.2921 0.0089	SP W/C ratio
K = 7	0.0114 0.0033 0.1128 0.1128	SP W/C ratio Coarse Aggregates Fine Aggregates
K = 9	1.1226 0.0553	SP W/C ratio

KSOM algorithm is also used to detect the main factors that affect Concrete Compressive Strength. KSOM was implemented on Tubas Governorate Dataset many times with several clusters and the main parameters were detected on their standard deviation values. Table T. 3 represents the main factors that affect Concrete Compressive Strength and these parameters (Coarse Aggregates, Water, SP, Cement, W/C ratio, and Age) as an intersection between clusters.

Table T. 3: Main factors that affect Concrete Compressive Strength by KSOM algorithm In Tubas Governorate.

Name	**KSOM**
Number of Clusters	**Main factors that affect Concrete Compressive Strength by KSOM algorithm In Tubas Governorate.**
k=3	Coarse Aggregates, SP, age, water , cement
K=5	Coarse Aggregates ,water ,SP, cement , W/C ratio, age
K = 7	Coarse Aggregates ,water, SP, cement , W/C ratio, age
K=9	Coarse Aggregates, water, SP , cement , W/C ratio, age

K-mean algorithm is also utilized to detect the main factors that affect Concrete Compressive Strength, and K-Means was applied on the dataset of Tubas governorate many times with different numbers of clusters as shown in Table T. 4 with their values. Table T. 5 represents the main factors that affect Concrete Compressive Strength. These parameters (Coarse Aggregates, Fine Aggregates, SP, w/c ratio, and Age) are produced by K mean.

Table T. 4: Main factors that affect Concrete Compressive Strength by K-Mean Algorithm in Tubas Governorate.

Name	K-mean
Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	Coarse Aggregates ,water , SP ,cement , W/C ratio, age
K= 5	Coarse Aggregates ,water, SP , cement , W/C ratio, age
K = 7	Coarse Aggregates ,water , SP, cement , W/C ratio, age
K = 9	Coarse Aggregates ,water , SP, cement , W/C ratio, age

Table T. 5: Results for K-Means algorithm based on some different number of clusters (K=3, 5, 7 and 9) in Tubas Governorate.

Number of clusters	Result for K=3, 5 , 7 ,9

K=3

Final cluster centroids:

Attribute	Full Data (65.0)	Cluster# 0 (42.0)	1 (16.0)	2 (7.0)
Coarse agg	1319.4923	1315.381	1327	1327
Fine Aggegate	543.0154	551.2381	528	528
Water	146.5538	131.3333	185.875	148
Cement	298.3077	295.7143	326.25	250
SP	1.3209	2.0443	1.921	0
w/c Ratio	0.5051	0.4662	0.57	0.59
Age	13.0769	12.7619	13.375	14.2857
CS	213.2462	217.7143	224.6875	160.2857

K=5

Attribute	Full Data (65.0)	Cluster# 0 (29.0)	1 (10.0)	2 (4.0)	3 (8.0)	4 (14.0)
Coarse agg	1319.4923	1327	1327	1327	1266	1327
Fine Aggegate	543.0154	528	528	528	650	528
Water	146.5538	118.6552	184.6	148	170.125	163.2857
Cement	298.3077	270.6897	324	250	377.5	305.7143
SP	1.3209	1.7655	0	0	3.0325	1.7429
w/c Ratio	0.5051	0.4586	0.57	0.59	0.495	0.5364
Age	13.0769	9.9655	4.6	4	13.375	28
CS	213.2462	179.931	163.7	106.5	274.375	313.2143

K=7

Final cluster centroids:

Attribute	Full Data (65.0)	Cluster# 0 (8.0)	1 (10.0)	2 (4.0)	3 (8.0)	4 (9.0)	5 (11.0)	6 (15.0)
Coarse agg	1319.4923	1327	1327	1327	1266	1327	1327	1327
Fine Aggegate	543.0154	528	528	528	650	528	528	528
Water	146.5538	144	184.6	148	170.125	174.6667	154.3636	109
Cement	298.3077	325	324	250	377.5	303.3333	297.2727	250
SP	1.3209	2.2	0	0	3.0325	0	1.8182	1.6
w/c Ratio	0.5051	0.455	0.57	0.59	0.495	0.5767	0.5618	0.46
Age	13.0769	5.5	4.6	4	13.375	28	28	5.1333
CS	213.2462	198.875	163.7	106.5	274.375	294.8889	295.8182	140.2667

K=9

Final cluster centroids:

Attribute	Full Data (65.0)	Cluster# 0 (3.0)	1 (10.0)	2 (4.0)	3 (8.0)	4 (9.0)	5 (5.0)	6 (15.0)	7 (5.0)	8 (6.0)
Coarse agg	1319.4923	1327	1327	1327	1266	1327	1327	1327	1327	1327
Fine Aggegate	543.0154	528	528	528	650	528	528	528	528	528
Water	146.5538	143.3333	184.6	148	170.125	174.6667	142.8	109	144.4	109
Cement	298.3077	316.6667	324	250	377.5	303.3333	310	250	330	250
SP	1.3209	2.1333	0	0	3.0325	1.912	2.08	1.6	2.24	1.6
w/c Ratio	0.5051	0.46	0.57	0.59	0.495	0.5767	0.464	0.46	0.452	0.46
Age	13.0769	3	4.6	4	13.375	28	28	5.1333	7	28
CS	213.2462	138.3333	163.7	106.5	274.375	294.8889	346.2	140.2667	235.2	253.8333

Table T. 6: Summary of the main parameters that affect compressive strength of concrete using the three algorithms in Tubas Governorate.

Name	EM	KSOM	K-Means	Intersection
Main parameters	W/C Ratio , SP	Coarse Aggregates, Fine Aggregates, SP ,w/c Ratio	Coarse Aggregates ,water, SP , cement , W/C ratio, age	**W/C Ratio , SP**

Table T. 6 represents the comparison between the three algorithms To detect the main factors that affect Compressive Strength of concrete. K-Mean algorithm produces these parameters (Coarse Aggregates, Water, SP, Cement, W/C ratio, and Age). According to the results obtained from EM algorithms (SP and w/c ratio) and KSOM (Coarse Aggregates, Fine Aggregates, SP and w/c Ratio), this leads to KSOM and K-Mean to show distinguishing factors which are the (Coarse Aggregates, Fine Aggregates, Water and Cement). It is clear that the three algorithms show intersection and provide different results and the analysis concludes that these factors (SP and w/c ratio) are common factors that affect Concrete Compressive Strength.

4.1.5 Detecting Main Factors that Affect CCS in Salfit Governorate:

The dataset of Salfit Governorate is first selected and then implemented these algorithms on Salfit Governorate dataset. In general, the dataset consists of 7 input parameters (Fine Aggregates, Coarse Aggregates, Superplasticizer, Water, Cement, W/C ratio, and Age) that were examined against Concrete Compressive Strength (CCS) in Salfit Governorate using these algorithms (EM, KSOM, K-Means). This section compares and evaluates these algorithms using Salfit Governorate dataset to investigate the main or primary factors that affect the concrete mix. Table S. 1 shows the main results of EM algorithms. To extract these results, different datasets were used with different numbers of clusters (k=3, 5, 7, and 9) as represented in Table S. 1. After applying EM on the Salfit Governorate dataset many times with a different number of clusters, the main parameters were detected on their standard deviation values and Table S2 represents the main factors that affect Concrete Compressive Strength and their standard deviation values.

Table S. 1: Main factors that affect Concrete Compressive Strength in Salfit Governorate by EM Algorithm.

Number of Clusters	**Main factors that affect concrete compressive strength Salfit Governorate By EM Algorithm**
K = 3	W/C Ratio , SP , age
K = 5	W/C Ratio , SP , age
K=7	W/C Ratio , SP , age
K = 9	W/C Ratio , SP , age

Table S. 2: List of the main factors with their standard deviation in Salfit Governorate.

Number of Clusters	**Standard Deviation**	**Main factors that affect Concrete Compressive Strength**
K = 3	1.3884 0.065 0.0031	SP W/C ratio age
K= 5	0.969 0.0378 0.08	SP W/C ratio age
K = 7	0.3771 0.0067 1.7427	SP W/C ratio age
K = 9	0.5636 0.028 1.912	SP W/C ratio age

Table S. 3 represents the main factors that affect Concrete Compressive Strength produced by KSOM, and these parameters (W/C Ratio, SP, and age) as an intersection between clusters. KSOM is another algorithm that is utilized to detect the main factors that affect Concrete Compressive Strength, after applying KSOM on Salfit Governorate dataset many times with different numbers of clusters. The main factors were detected on their standard deviation values.

Table S. 3: Main factors that affect Concrete Compressive Strength by KSOM algorithm In Salfit Governorate.

Name	**KSOM**
Number of Clusters	**Main factors that affect Concrete Compressive Strength by KSOM algorithm In Salfeet Governorate.**
k=3	W/C Ratio , SP , age
K=5	W/C Ratio , SP , age
K = 7	W/C Ratio , SP , age
K=9	W/C Ratio , SP , age

K-mean algorithm is another algorithm that is used to detect the main factors that affect Concrete Compressive Strength, after applying K-Means on the dataset of Salfit governorate many times with different numbers of clusters as shown in Table S. 4 with their values. Table S. 5 represents the main factors that affect Concrete Compressive Strength and these parameters (Coarse Aggregates, W/C Ratio, and age) as an intersection between clusters were obtained from this algorithm. Table S. 6 represents the comparison between three algorithms to detect the main factors that affect Compressive Strength of concrete. K-Mean algorithm produces these parameters (Coarse Aggregates, W/C Ratio, and age) and according to the results obtained from EM algorithms (W/C Ratio, SP and age) and KSOM (w/c Ratio, SP and Age). K-Mean shows a distinguishing factor which is the (Coarse Aggregates and SP). It is clear that the three algorithms show intersection and provide different results and the analysis concludes that these factors (Age and w/c ratio) are common factors and are the two primary components that affect Concrete Compressive Strength.

Table S. 4: Results for K-Means algorithm based on some different number of clusters (K=3, 5, 7 and 9) in Salfit Governorate.

Number of clusters	Result for K=3, 5 , 7 ,9

K=3

Final cluster centroids:

Attribute	Full Data (71.0)	Cluster# 0 (27.0)	1 (22.0)	2 (22.0)
Coarse agg	938.0507	955.7296	1005.9091	848.4955
Fine Aggegate	942.4789	923.3	878.4091	1030.0864
Water	159.4507	159.4444	158.8182	160.0909
Cement	295.9577	291.4815	282.5	314.9091
SP	4.3521	4.163	3.5636	5.3727
w/c Ratio	0.5263	0.533	0.5527	0.4918
Age	14.0282	28	5.9091	5
CS	181.0031	230.0219	211.5095	90.3373

K=5

Final cluster centroids:

Attribute	Full Data (71.0)	Cluster# 0 (10.0)	1 (20.0)	2 (18.0)	3 (17.0)	4 (6.0)
Coarse agg	938.0507	889.66	1010.5	846.9444	1010.2941	845.8333
Fine Aggegate	942.4789	997.1	870.25	1032.7778	868.2353	1031.6667
Water	159.4507	173.6	158.5	155.8333	158.8235	151.6667
Cement	295.9577	271	285.75	322.6667	286.1765	319.1667
SP	4.3521	4.7	3.52	5.4778	3.4941	5.6
w/c Ratio	0.5263	0.641	0.543	0.4589	0.5365	0.4533
Age	14.0282	14.6	6	4.7778	28	28
CS	181.0031	34.972	229.915	103.8833	315.5353	111.5333

K=7

Final cluster centroids:

Attribute	Full Data (71.0)	Cluster# 0 (7.0)	1 (19.0)	2 (9.0)	3 (15.0)	4 (5.0)	5 (9.0)	6 (7.0)
Coarse agg	938.0507	885.6714	1015	841.6667	1023.3333	851	895.2111	840
Fine Aggegate	942.4789	985.5857	862.3684	993.3333	864.6667	1012	1011.3222	1080
Water	159.4507	186.4286	158.4211	148.8889	151.3333	146	164.5556	169.2857
Cement	295.9577	271.4286	286.3158	358.6667	287.6667	331	272.7778	288.5714
SP	4.3521	4.7714	3.4421	6.7778	3.2933	5.92	4.6889	4
w/c Ratio	0.5263	0.6014	0.5411	0.4178	0.5307	0.444	0.6044	0.5
Age	14.0282	28	5.9474	5.2222	28	28	5.2222	4.7143
CS	181.0031	186.6414	227.4895	29.2333	308.0067	56.8	53.3811	224.9714

K=9

Final cluster centroids:

Attribute	Full Data (71.0)	Cluster# 0 (5.0)	1 (19.0)	2 (9.0)	3 (4.0)	4 (2.0)	5 (9.0)	6 (7.0)	7 (9.0)	8 (7.0)
Coarse agg	938.0507	895.94	1015	841.6667	846.25	845	895.2111	840	1043.3333	980
Fine Aggegate	942.4789	999.82	862.3684	993.3333	1053.75	987.5	1011.3222	1080	883.3333	827.1429
Water	159.4507	171	158.4211	148.8889	152.5	150	164.5556	169.2857	135.5556	188.5714
Cement	295.9577	273	286.3158	358.6667	295	367.5	272.7778	288.5714	243.3333	342.8571
SP	4.3521	4.88	3.4421	6.7778	4.8	7.2	4.6889	4	3.0111	3.9
w/c Ratio	0.5263	0.626	0.5411	0.4178	0.475	0.41	0.6044	0.5	0.5511	0.5314
Age	14.0282	28	5.9474	5.2222	28	28	5.2222	4.7143	28	28
CS	181.0031	114.258	227.4895	29.2333	129.95	74.7	53.3811	224.9714	208.5222	441.9143

Table S. 5: Main factors that affect Concrete Compressive Strength by K-Mean Algorithm in Salfit Governorate.

Name	K-mean
Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	Coarse Aggregates, W/C Ratio , age
K= 5	Coarse Aggregates, W/C Ratio , age
K = 7	Coarse Aggregates, W/C Ratio , age
K = 9	Coarse Aggregates ,water , cement , W/C Ratio , age

Table S. 6: Summary of the main parameters that affect Compressive Strength of Concrete using the three algorithms in Salfit Governorate.

Name	EM	KSOM	K-Means	Intersection
Main parameters	W/C Ratio , SP , age	w/c Ratio , SP , Age	Coarse Aggregates, W/C Ratio, age	**w/c Ratio, Age**

4.1.6 Detecting Main Factors that Affect CCS in Hebron Governorate:

The dataset of Hebron Governorate is first selected and then implemented these algorithms on the Hebron dataset. In general, the dataset consists of 7 input parameters (Fine Aggregates, Coarse Aggregates, Superplasticizer, Water, Cement, W/C ratio, and Age) which were examined against Concrete Compressive Strength (CCS) in Hebron Governorate using these algorithms (EM, KSOM, and K-Means). This section compares and evaluates these algorithms using Hebron Governorate dataset to investigate the main or primary factors that affect the concrete mix. Table H. 1 shows the main results of EM algorithms. To extract these results, different datasets were used with different numbers of clusters (k=3, 5, 7, and 9) as represented in Table H. 1. After applying EM on the Hebron Governorate dataset many times with different numbers of clusters and the main parameters were detected on their standard deviation values. Table H. 2 represents the main factors that affect Concrete Compressive Strength and their standard deviation values.

Table H. 1: Main factors that affect Concrete Compressive Strength in Hebron Governorate by EM Algorithm.

Number of Clusters	Main factors that affect Concrete Compressive Strength in Hebron Governorate By EM Algorithm
K = 3	Fine Aggregates, W/C Ratio , SP, age
K = 5	Fine Aggregates, W/C Ratio , SP, age
K=7	Fine Aggregates, W/C Ratio , SP, age
K = 9	Coarse Aggregates ,Cement , W/C Ratio , SP, age

KSOM algorithm is also used to detect the main factors that affect Hebron Concrete Compressive Strength. After applying KSOM on Hebron Governorate Dataset many times with different numbers of clusters and the main parameters were detected on their standard deviation values. Table H. 3 represents the main factors that affect concrete compressive strength and these parameters (Fine Aggregates, SP, w/c Ratio and Age) as an intersection between clusters were obtained by KSOM

Table H. 2: List of the main factors with their standard deviation in Hebron Governorate.

Number of Clusters	**Standard Deviation**	**Main factors that affect Concrete Compressive Strength**
K = 3	0.0003 1.1351 0.0335 1.0155	Fine Aggregates SP W/C ratio Age
K= 5	0.0003 1.133 0.0335 0.4565	Fine Aggregates SP W/C ratio Age
K = 7	0.0001 1.567 0.438 0.4994	Fine Aggregates SP W/C ratio age
K = 9	0.0001 0 0.0002 1.566 0.451 0.4774	Coarse Aggregates Fine Aggregates Cement SP W/C ratio age

Table H. 3: Main factors that affect Concrete Compressive Strength by KSOM algorithm In Hebron Governorate.

Name	**KSOM**
Number of Clusters	**Main factors that affect Concrete Compressive Strength by KSOM algorithm In Hebron Governorate.**
k=3	Fine Aggregates, W/C Ratio , SP, age
K=5	Fine Aggregates, W/C Ratio , SP, age
K = 7	Fine Aggregates, W/C Ratio , SP, age
K=9	Fine Aggregates, W/C Ratio , SP, age

Table H. 5 represents the main factors that affect Concrete Compressive Strength which were obtained from this K-mean algorithm and K-mean is used to detect the main factors that affect Concrete Compressive Strength after applying K-Means on the dataset of Hebron governorate many times with different numbers of clusters as shown in Table H. 4 with their values. These factors (Coarse Aggregates, Fine Aggregates, W/C ratio, and Age) as an intersection between clusters were obtained from this algorithm. K-Mean algorithm produces these parameters (Coarse Aggregates, Fine Aggregates, W/C ratio, and Age) and according to the results obtained from EM algorithms (Fine Aggregates, W/C Ratio, SP and Age) and KSOM (Fine Aggregates, SP, w/c Ratio and Age). K-Mean, KSOM, and EM show a distinguishing factor which is the (Coarse Aggregates and SP). Table H. 6 represents the comparison between the three algorithms to detect the main factors that affect the Compressive Strength of Concrete. It is clear that the three algorithms show intersection and provide different results and the analysis concludes that these factors (Fine Aggregates, W/C Ratio, and Age) are common factors and they are the three primary components that affect Concrete Compressive Strength.

Table H. 4: Results for K-Means algorithm based on some different number of clusters (K=3, 5, 7 and 9) in Hebron Governorate.

number of clusters	Result for K=3, 5 , 7 ,9

K=3

Final cluster centroids:

Attribute	Full Data (99.0)	Cluster# 0 (76.0)	1 (10.0)	2 (13.0)
Coarse agg	1167.5253	1147.1053	1235	1235
Fine Aggegate	616.9697	610	640	640
Water	155.7374	161.9474	135	135.3846
Cement	324.7475	347.3684	250	250
SP	1.1273	1.1789	1	0.9231
w/c Ratio	0.4856	0.4688	0.54	0.5415
Age	13.4646	12.9737	28	5.1538
CS	246.3929	259.4579	273.45	149.2

K=5

Final cluster centroids:

Attribute	Full Data (99.0)	Cluster# 0 (6.0)	1 (10.0)	2 (13.0)	3 (36.0)	4 (34.0)
Coarse agg	1167.5253	1175	1235	1235	1189.4444	1150.2941
Fine Aggegate	616.9697	610	640	610	640	610
Water	155.7374	170	135	135.3846	153.8889	169.0588
Cement	324.7475	400	250	250	331.9444	354.4118
SP	1.1273	3.2	1	0.9231	1.9556	0
w/c Ratio	0.4856	0.42	0.54	0.5415	0.4961	0.4803
Age	13.4646	5	28	5.1538	15.3333	11.8824
CS	246.3929	228	273.45	149.2	276.1444	247.3412

K=7

Final cluster centroids:

Attribute	Full Data (99.0)	Cluster# 0 (6.0)	1 (5.0)	2 (5.0)	3 (36.0)	4 (34.0)	5 (7.0)	6 (6.0)
Coarse agg	1167.5253	1175	1235	1235	1139.4444	1150.2941	1235	1235
Fine Aggegate	616.9697	610	640	640	610	610	640	640
Water	155.7374	170	130	140	153.8889	169.0588	140	130
Cement	324.7475	400	250	250	331.9444	354.4118	250	250
SP	1.1273	3.2	1.1	0	1.9556	0	0	2
w/c Ratio	0.4856	0.42	0.52	0.56	0.4661	0.4803	0.56	0.52
Age	13.4646	5	28	28	15.3333	11.8824	5.2857	5
CS	246.3929	228	283.46	263.44	276.1444	247.3412	146.5143	152.3333

K=9

Final cluster centroids:

Attribute	Full Data (99.0)	Cluster# 0 (6.0)	1 (5.0)	2 (5.0)	3 (21.0)	4 (18.0)	5 (3.0)	6 (6.0)	7 (31.0)	8 (4.0)
Coarse agg	1167.5253	1175	1235	1235	1142.619	1163.8889	1235	1235	1135	1235
Fine Aggegate	616.9697	610	640	640	610	610	640	640	610	640
Water	155.7374	170	130	140	157.5238	176.6667	140	130	154.8387	140
Cement	324.7475	400	250	250	338.0952	386.1111	250	250	320.9677	250
SP	1.1273	3.2	1.3612	0	1.6381	0	0	2	1.1613	0
w/c Ratio	0.4856	0.42	0.52	0.56	0.4686	0.4583	0.56	0.52	0.4845	0.56
Age	13.4646	5	28	28	28	12.1667	3	5	4.8065	7
CS	246.3929	228	283.46	263.44	362	278.9556	106.7667	152.3333	184.7613	176.325

Table H. 5: Main factors that affect Concrete Compressive Strength by K-Mean Algorithm in Hebron Governorate.

Name	K-mean
Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	Coarse Aggregates, Fine Aggregates , W/C ratio , Age
K= 5	Coarse Aggregates, Fine Aggregates , W/C ratio , Age ,SP
K = 7	Coarse Aggregates, Fine Aggregates , W/C ratio , Age
K = 9	Coarse Aggregates, Fine Aggregates , W/C ratio , Age

Table H. 6: Summary of the main parameters that affect Compressive Strength of concrete using the three algorithms in Hebron Governorate.

Name	EM	KSOM	K-Means	Intersection
Main parameters	Fine Aggregates , W/C Ratio , SP , Age	Fine Aggregates, SP ,w/c Ratio ,Age	Coarse Aggregates, Fine Aggregates, W/C ratio, Age	**Fine Aggregates, W/C Ratio , Age**

4.1.7 Detecting Main Factors that Affect CCS in Nablus Governorate:

The Nablus Governorate is first selected, and then these algorithms were implemented on the Nablus dataset, In general, the dataset consists of 7 input parameters (Fine Aggregates, Coarse Aggregates, Superplasticizer, Water, Cement, W/C ratio, and Age) were examined against Concrete Compressive Strength (CCS) in Nablus Governorate using these algorithms (EM, KSOM, and K-Means). T

This section compares and evaluates these algorithms using Nablus governorate dataset to investigate the main or primary factors that affect the concrete mix. Table N. 1 shows the main results of EM algorithms. To extract these results, different datasets were used with different numbers of clusters (k=3, 5, 7, and 9) as represented in Table N. 1. After applying EM on the Nablus Dataset many times with different numbers of clusters, the main parameters were detected on their standard deviation values and

Table N. 2 represents the main factors that affect concrete compressive strength and their standard deviation values.

Table N. 1: Main factors that affect CSS Strength in Nablus Governorate by EM Algorithm.

Number of Clusters	Main factors that affect Concrete Compressive Strength in Nablus By EM Algorithm
K = 3	Fine Aggregates, Age, W/C Ratio , SP
K = 5	Age, W/C Ratio , SP
K=7	Fine Aggregates, Coarse Aggregates, Cement, Water, W/C Ratio , SP , Age
K = 9	Fine Aggregates, Water, W/C Ratio , SP

Table N. 2: List of the main factors with their standard deviation in Nablus Governorate.

Number of Clusters	Standard Deviation	Main factors that affect Concrete Compressive Strength
K = 3	0 1.38 0.0229 0.0849	Fine Aggregates SP W/C ratio Age
K= 5	1.3868 0.0229 0.2444	SP W/C ratio Age
K = 7	0.0802 0.1604 0.401 0.1069 0.0032 0.0003 0.0118	Fine Aggregates Coarse Aggregates, Cement Water SP W/C ratio Age
K = 9	0.00879 0.025 0.008 0.0001	Fine Aggregates, Water SP W/C Ratio

Table N. 3 represents the main factors that affect Concrete Compressive Strength and these parameters (Fine Aggregates, SP, and w/c Ratio) as an intersection between clusters were obtained from this KSOM.

Table N. 3: Main factors that affect Concrete Compressive Strength by KSOM algorithm In Nablus Governorate.

Name	**KSOM**
Number of Clusters	**Main factors that affect Concrete Compressive Strength by KSOM algorithm In Nablus Governorate.**
k=3	Fine Aggregates, SP ,w/c Ratio
K=5	Fine Aggregates, SP ,w/c Ratio
K = 7	Fine Aggregates, SP ,w/c Ratio
K=9	Fine Aggregates, SP ,w/c Ratio

K-mean algorithm is also used to detect the main factors that affect Concrete Compressive Strength, after applying K-Means on the dataset of Hebron governorate many times with different numbers of clusters as shown in table K. 4.

Table N. 5 represents the main factors that affect Concrete Compressive Strength and these parameters (water, Cement, SP, W/C ratio, and age) as an intersection between clusters were obtained from this K- mean.

Table N. 6 represents the comparison between the three algorithms to detect the main factors that affect the Compressive Strength of Concrete. K-Mean algorithm produces these parameters (W/C ratio, Cement, SP, and age) and according to the results obtained from EM algorithms (Fine Aggregates, Water, W/C Ratio and SP) and KSOM (Fine Aggregates, SP and W/C Ratio). This leads K-Mean, KSOM, and EM to show a distinguishing factor which is the (Fine Aggregates, Cement, and age). It is clear that the three algorithms show intersection and provide different results and the analysis concludes that these factors (SP and W/C Ratio) are common and they are the two primary components that affect concrete compressive strength.

Table N. 4: Results for K-Means algorithm based on some different number of clusters (K=3, 5, 7 and 9) in Nablus Governorate.

Number of clusters	Result for K=3, 5 , 7 ,9

K=3

Final cluster centroids:

Attribute	Full Data	Cluster# 0	1	2
	(54.0)	(20.0)	(22.0)	(12.0)
Coarse agg	1188.3333	1157	1235	1155
Fine Aggegate	622.2222	610	640	610
Water	153.6667	166.6	134.5455	167.1667
Cement	325	377.5	250	375
SP	1.4815	1.89	1.0909	1.5167
w/c Ratio	0.4811	0.44	0.5382	0.495
Age	13.9815	5	14.5	28
CS	247.4889	211.19	201.1364	392.9667

K=5

Final cluster centroids:

Attribute	Full Data	Cluster# 0	1	2	3	4
	(54.0)	(9.0)	(11.0)	(12.0)	(12.0)	(10.0)
Coarse agg	1188.3333	1175	1142.2727	1155	1235	1235
Fine Aggegate	622.2222	610	610	610	640	640
Water	153.6667	170	163.8182	167.1667	130	140
Cement	325	400	359.0909	375	250	250
SP	1.4815	3.2	0.8182	1.5167	2	0
w/c Ratio	0.4811	0.42	0.4564	0.485	0.52	0.56
Age	13.9815	4.7778	5.1818	28	14.75	14.2
CS	247.4889	222.2444	202.1455	392.9667	209.9583	190.55

K=7

Attribute	Full Data	Cluster# 0	1	2	3	4	5	6
	(54.0)	(9.0)	(11.0)	(4.0)	(5.0)	(4.0)	(13.0)	(8.0)
Coarse agg	1188.3333	1175	1149.5455	1175	1235	1235	1235	1135
Fine Aggegate	622.2222	610	610	610	640	640	640	610
Water	153.6667	170	172.3636	170	130	140	134.6154	154
Cement	325	400	368.1818	400	250	250	250	350
SP	1.4815	3.	3.2	0	2	0	1.0769	1.8
w/c Ratio	0.4811	0.42	0.4691	0.42	0.52	0.56	0.5385	0.44
Age	13.9815	4.7778	15.8182	28	28	28	5.1538	13.375
CS	247.4889	222.2444	285.3636	423.075	283.46	263.55	150.2692	263.4875

K=9

Final cluster centroids:

Attribute	Full Data	Cluster# 0	1	2	3	4	5	6	7	8
	(54.0)	(9.0)	(6.0)	(4.0)	(5.0)	(4.0)	(6.0)	(5.0)	(8.0)	(7.0)
Coarse agg	1188.3333	1175	1148.3333	1175	1235	1235	1235	1135	1145	1235
Fine Aggegate	622.2222	610	610	610	640	640	640	610	610	640
Water	153.6667	170	172	170	130	140	140	154	165.75	130
Cement	325	400	366.6667	400	250	250	250	350	362.5	250
SP	1.4815	3.2	0	3.2	2	0	0	1.8	0.675	2
w/c Ratio	0.4811	0.42	0.47	0.49	0.52	0.56	0.56	0.44	0.4575	0.52
Age	13.9815	4.7778	5.6667	28	28	28	5	4.6	28	5.2857
CS	247.4889	222.2444	209.7833	423.075	283.46	263.55	141.8833	192.98	377.9125	157.4571

Table N. 5: Main factors that affect CCS by K-Mean Algorithm in Nablus Governorate.

Name	K-mean
Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	water , Cement, w/c ratio ,SP , age
K= 5	water , Cement, w/c ratio ,SP , age
K = 7	water , Cement, w/c ratio, SP , age
K = 9	water , Cement, w/c ratio , SP , age

Table N. 6: Summary of the main parameters that affect Compressive Strength of concrete using the three algorithms in Nablus Governorate.

Name	EM	KSOM	K-Means	Intersection
Main parameters	Fine Aggregates , Water, W/C Ratio , SP	Fine Aggregates, SP ,w/c Ratio	W/C ratio , Cement, ,Water ,SP , age	**SP ,w/c Ratio**

4.1.8 Detecting main Factors that CCS in Tulkram Governorate:

The dataset of Tulkram Governorate is first selected, and then these algorithms were implemented on the Tulkarem dataset, In general, the dataset consists of 7 input parameters (Fine Aggregates, Coarse Aggregates, Superplasticizer, Water, Cement, W/C ratio, and Age) were examined against Concrete Compressive Strength (CCS) in Tulkarm Governorate using these algorithms (EM, KSOM, and K-Means). This section compares and evaluates these algorithms using Tulkram Governorate dataset to investigate the main or primary factors that affect the concrete mix. Table K. 1 shows the main results of EM algorithms and to extract these results, different datasets were used with different numbers of clusters (k=3, 5, 7, and 9) as represented in Table K. 1. After applying EM on the Tulkarem Dataset many times with different numbers of clusters and the main parameters were detected on their standard deviation values. Table K. 2 represents the main factors that affect Concrete Compressive Strength and their standard deviation values.

Table K. 1: Main factors that affect Concrete Compressive Strength in Tulkram Governorate by EM Algorithm.

Number of Clusters	**Main factors that affect Concrete Compressive Strength in Tulkram By EM Algorithm**
K = 3	w/c Ratio , SP , Age
K = 5	w/c Ratio , SP , Age
K=7	w/c Ratio , SP , Age
K = 9	w/c Ratio , SP , Age

Table K. 2: List of the main factors with their standard deviation in Tulkram Governorate.

Number of Clusters	**Standard Deviation**	**Main factors that affect Concrete Compressive Strength**
K = 3	1.2709 0.0498 2.111	SP W/C ratio age
K= 5	1.5775 0.0416 2.271	SP W/C ratio age
K = 7	1.5775 0.0417 2.2204	SP W/C ratio age
K = 9	1.7913 0.0348 0.0239	SP W/C ratio age

Table K. 3 represents the main factors that affect Concrete Compressive Strength and these parameters (W/C Ratio, SP, and age) as an intersection between clusters were obtained from this algorithm by using standard devotion values.

Table K. 3: Main factors that affect CCS by KSOM algorithm In Tulkram Governorate.

Name	**KSOM**
Number of Clusters	**Main factors that affect Concrete Compressive Strength by KSOM algorithm In Tulkram Governorate.**
k=3	w/c Ratio , SP , Age
K=5	w/c Ratio , SP , Age
K = 7	w/c Ratio , SP , Age
K=9	w/c Ratio , SP , Age

K-mean algorithm is also used to detect the main factors that affect Concrete Compressive Strength, after applying K-Means on the dataset of Hebron governorate many times with different numbers of clusters as shown in Table K. 4 with their values. Table K. 5 represents the main factors that affect CSS, and these parameters (Coarse Aggregates, fine Aggregates, water, cement, and SP) as an intersection between clusters were obtained from this algorithm.

Table K. 4: Results for K-Means algorithm based on some different number of clusters (K=3, 5, 7 and 9) in Tulkram Governorate.

# of cluster	Result for K=3, 5 , 7 ,9

K=3

Final cluster centroids:

Attribute	Full Data (42.0)	Cluster# 0 (7.0)	1 (28.0)	2 (7.0)
Coarse agg	1227.6905	1300.8571	1178.5714	1351
Fine Aggegate	622.781	580.2857	664.8143	497.1429
Water	159.4524	173.5714	146.75	196.1429
Cement	312.1429	402.8571	278.2143	357.1429
SP	2.0657	3.2429	2.0379	1
w/c Ratio	0.5262	0.53	0.5343	0.55
Age	11.7619	11.8571	11.1071	14.2857
CS	158.354	279.5714	125.1811	169.8286

K=5

Final cluster centroids:

Attribute	Full Data (42.0)	Cluster# 0 (2.0)	1 (24.0)	2 (3.0)	3 (12.0)	4 (1.0)
Coarse agg	1227.6905	1266	1202.275	1355	1232.0333	1327
Fine Aggegate	622.781	650	636.575	492	631.25	528
Water	159.4524	193	147.625	194.6667	164.9167	205
Cement	312.1429	460	285.8333	356.6667	325	360
SP	2.0657	3.7	2.1533	1.1667	2.015	0
w/c Ratio	0.5262	0.53	0.5258	0.5467	0.5208	0.57
Age	11.7619	15.5	5.5	3	28	7
CS	158.354	248.5	135.8146	90.4267	200.0867	222

K=7

Final cluster centroids:

Attribute	Full Data (42.0)	Cluster# 0 (2.0)	1 (13.0)	2 (3.0)	3 (12.0)	4 (1.0)	5 (9.0)	6 (2.0)
Coarse agg	1227.6905	1266	1212.4923	1355	1232.0333	1327	1235.3556	987
Fine Aggegate	622.781	650	587.2769	492	631.25	528	650.8	893
Water	159.4524	193	147.6154	194.6667	164.9167	205	142.7778	169.5
Cement	312.1429	460	306.1538	356.6667	325	360	258.8889	275
SP	2.0657	3.7	1.3708	1.1667	2.015	0	2.2511	6.8
w/c Ratio	0.5262	0.51	0.4862	0.5467	0.5208	0.57	0.5633	0.615
Age	11.7619	28	5.4615	3	5	7	5.6667	5
CS	158.354	248.5	170.1308	90.4267	200.0867	222	79.9756	164.035

K=9

Final cluster centroids:

Attribute	Full Data (42.0)	Cluster# 0 (2.0)	1 (6.0)	2 (3.0)	3 (12.0)	4 (1.0)	5 (4.0)	6 (2.0)	7 (7.0)	8 (5.0)
Coarse agg	1227.6905	1266	1302.9	1355	1232.0333	1327	1327	987	1135	1162.04
Fine Aggegate	622.781	650	560.7667	492	631.25	528	528	893	610	749.04
Water	159.4524	193	139.8333	194.6667	164.9167	205	133	169.5	154.2857	150.6
Cement	312.1429	460	313.3333	356.6667	325	360	250	275	300	266
SP	2.0657	3.7	2.37	1.1667	2.015	0	2	6.8	0.5143	2.452
w/c Ratio	0.5262	0.53	0.4533	0.5467	0.5208	0.57	0.56	0.615	0.5143	0.566
Age	11.7619	28	5.6667	3	5	7	5	5	5.2857	6.2
CS	158.354	248.5	155.2333	90.4267	200.0867	222	138.5	164.035	182.9	33.156

Table K. 5: Main factors that affect Concrete Compressive Strength by K-Mean Algorithm in Tulkram Governorate.

Name	K-mean
Number of Clusters	Main factors that affect Concrete Compressive Strength
K = 3	coarse Aggregates, water , cement ,w/c ratio , SP, Age
K= 5	coarse Aggregates, fine Aggregates, water , w/c ratio , cement , SP , Age
K = 7	coarse Aggregates, fine Aggregates, water , w/c ratio, cement , SP , Age
K = 9	coarse Aggregates, fine Aggregates, water , w/c ratio , cement , SP , Age

Table K. 6: Summary of the main parameters that affect Compressive Strength of concrete using the three algorithms in Tulkram Governorate.

Name	EM	KSOM	K-Means	Intersection
Main parameters	SP , W/C Ratio , Age	w/c Ratio , SP , Age	Coarse Aggregates, water ,SP , w/c ratio , cement , Age	**w/c Ratio , SP**

Table K. 6 represents the comparison between the three algorithms to detect the main factors that affect Compressive Strength of concrete, K-Mean algorithm produces these parameters (W/C ratio, Cement, SP, and age) and according to the results obtained from EM algorithms (SP, W/C Ratio and Age), and KSOM (SP, W/C Ratio and Age) This leads K-Mean EM to show a distinguishing factor which is the (Coarse Aggregates). It is clear that the three algorithms show intersection and provide different results, and the analysis concludes that these factors (SP, W/C Ratio, and Age) are common factors and they are the three primary components that affect concrete compressive strength.

Table of Palestine Governorates Summary 1 represents the summary of the comparison between the three algorithms to detect the main factors that affect compressive strength of concrete on each governorate of Palestinian Governorates which are the K-Mean algorithm, EM algorithm, and KSOM. It is clear that the three algorithms show intersection and provide different results and the analysis concludes that different factors are common factors that affect concrete compressive strength on each Governorate.

Table of Palestine Governorates Summary 1: Table Palestinian Governorates Summary: Summary of main factors that affect each governorate of Palestinian Governorates.

Name	**EM**	**K-mean**	**KSOM**	
	Similar items	Similar items	Similar items	**Intersection of algorithms**
Jenin	SP , W/C Ratio, Age	Coarse Aggregates, Fine Aggregates, SP, w/c ratio , Age	Coarse Aggregates , Fine Aggregates, SP ,w/c Ratio , Age	**SP , W/C ratio , Age**
Ramallah	W/C Ratio , SP , Age	Coarse Aggregates, water , cement, SP , W/C ratio, Age	w/c Ratio , SP , Age	**w/c Ratio , SP , Age**
Tubas	W/C Ratio , SP	Coarse Aggregates ,water, SP , cement , W/C ratio, age	Coarse Aggregates , Fine Aggregates, SP ,w/c Ratio	**W/C Ratio , SP**
Salfit	W/C Ratio , SP , age	Coarse Aggregates, , W/C Ratio , age	w/c Ratio , SP , Age	**w/c Ratio, Age**
Hebron	Fine Aggregates , W/C Ratio , SP , Age	Coarse Aggregates, Fine Aggregates , W/C ratio , Age	Fine Aggregates, SP ,w/c Ratio ,Age	**Fine Aggregates, w/c Ratio ,Age**
Nablus	Fine Aggregates , Water, W/C Ratio , SP	W/C ratio , Cement, ,Water ,SP , age	Fine Aggregates, SP ,w/c Ratio	**SP ,w/c Ratio**
Tulkarem	SP , W/C Ratio , Age	Coarse Aggregates, water ,SP, w/c ratio , cement	w/c Ratio , SP , Age	**w/c Ratio , SP**

4.2 Classification Results

In this section and from results, it was discussed to make a classification for the datasets that were collected from Palestinian governorates laboratories for all types of concrete like B200, B250, B300, B350, and B400. These datasets were applied to the classification techniques like Multilayer Perceptron Neural Networks (MLPNNs), Linear Support Vector Machine (SVM), and Ensemble algorithm (ES). The results show the accuracy for each type of concrete by using these techniques. Table 4.2.1 represents the summary of results obtained from these techniques and the results showed that MLPNNs accuracy is more accurate than other techniques for each type of concrete.

Table 4.2. 1: Accuracy Results for Classification Models.

Type / Algorithm	**Neural Networks Accuracy**	**Linear support vector machine (SVM) Accuracy**	**Ensemble Algorithm Accuracy**
B200	93.5%	80.4%	90.2%
B250	90.0%	66.5%	75.5%
B300	93.3%	68.3%	79.2%
B350	90.6%	83.3%	85.6%
B400	90.0%	80.6%	78.6%

Table 4.2.2 shows that the number of Neuron of Neural Networks has the best accuracy for each type which is classified in this table below by using Neural Networks Technique and the ranges of neurons for each type was between [2–20].

Figure C1 represents the Chart of Accuracy Results for Classification Techniques, the results showed that MLPNNs accuracy is more accurate than SVM and ensemble for each type of concrete.

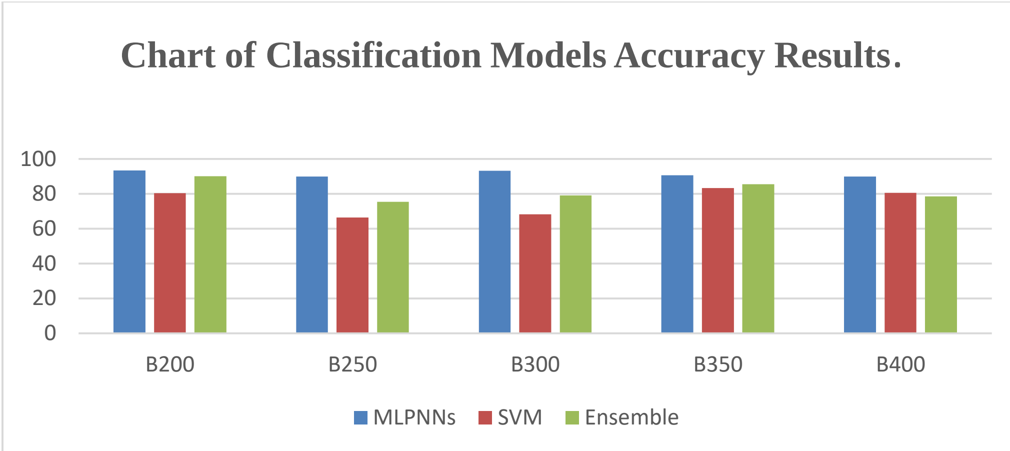


Figure C1: Chart of Accuracy Results for Classification Models.

Table 4.2. 2: Summary Accuracy Results of MLPNNs.

Type / Algorithm	**MLPNNs Accuracy**	**Number of Neurons**
B200	93.5%	4 , 8, 10
B250	90.0%	18
B300	93.3%	10
B350	90.6%	18
B400	90.0%	20

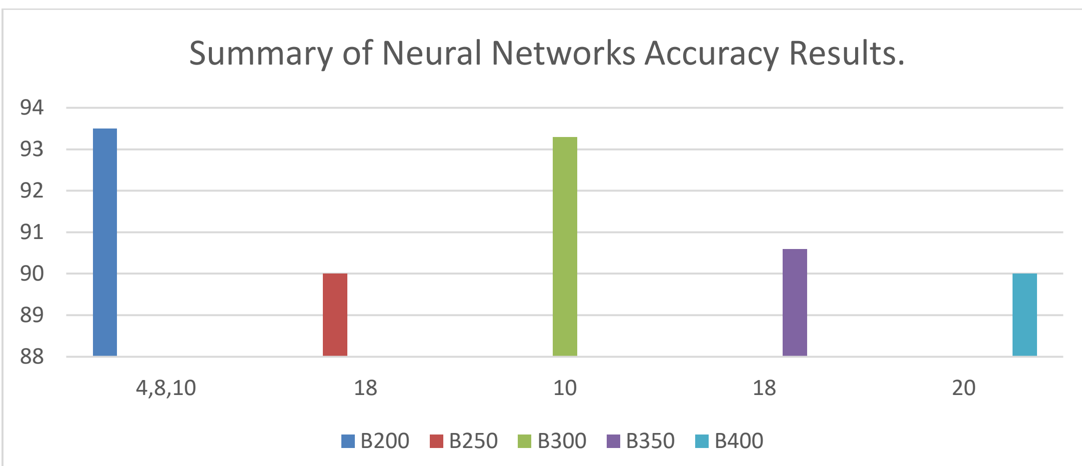


Figure C2: Chart of Summary Accuracy Results of MLPNNs for each type of concrete based on the best number of neurons.

Figure C2 shows that the number of neurons of MLPNNs was used for each type of concrete. Table 4.2.3 represents that the range of the number of neurons was used for B200 concrete which was classified in this table below by using the MLPNNs technique, and this table below shows some parameters that can be obtained from Confusion Matrix based on (TP, TN, FP, and FN). These parameters were Sensitivity, Specificity, Precision, and Negative Prediction. The best accuracy was 93.5% at N = 4.

Table 4.2.4 represents the range of MLPNNs Neurons used for B250 concrete which was classified in this table below by using the MLPNNs Technique. The best accuracy was 90.0% at N = 18 and it shows some parameters that can be obtained from Confusion Matrix based on (TP, TN, FP, and FN). These parameters were Sensitivity, Specificity, Precision, and Negative Prediction.

Table 4.2. 3: B200 MLPNNs Accuracy.

MLPNNs Accuracy										
Accuracy	93.5%									
N	Training Accuracy	Testing Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
2	72.5%	77.4%	12	5	12	2	85.7%	70.6%	70.6%	85.7%
4	90.8%	93.5%	19	0	10	2	90.5%	100.0%	100.0%	83.3%
6	95.1%	90.3%	18	2	10	1	94.7%	83.3%	90.0%	90.9%
8	89.4%	93.5%	16	1	13	1	94.1%	92.9%	94.1%	92.9%
10	90.1%	93.5%	16	0	13	2	88.9%	100.0%	100.0%	86.7%
12	88.0%	83.9%	13	0	13	5	72.2%	100.0%	100.0%	72.2%
14	89.4%	87.1%	15	2	12	2	88.2%	85.7%	88.2%	85.7%
16	90.1%	87.1%	16	2	11	2	88.9%	84.6%	88.9%	84.6%
18	93.0%	87.1%	13	2	14	2	86.7%	87.5%	86.7%	87.5%
20	92.3%	90.3%	12	3	16	0	100.0%	84.2%	80.0%	100.0%

Table 4.2.5 represents the range of Neural Networks Neuron has been used for B300 concrete which was classified in this table below by using the Neural Networks technique. The table below shows some parameters that can be obtained from Confusion Matrix based on (TP, TN, FP, and

FN). These parameters were Sensitivity, Specificity, Precision, and Negative Prediction and the best accuracy was 93.3% at N = 10.

Table 4.2. 4: B250 MLPNNs Accuracy.

MLPNNs Accuracy										
Accuracy	90.0%									
N	Training Accuracy	Testing Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
2	72.9%	80.0%	15	4	9	2	88.2%	69.2%	78.9%	81.8%
4	55.7%	46.7%	10	10	4	6	62.5%	28.5%	50.0%	40.0%
6	80.7%	70.0%	12	6	9	3	80.0%	60.0%	66.7%	75.0%
8	79.3%	70.0%	11	4	10	5	68.8%	71.4%	73.4%	66.7%
10	62.1%	50.0%	13	8	2	7	65.0%	20.0%	61.9%	22.2%
12	67.1%	66.7%	12	6	8	4	75.0%	57.1%	66.7%	66.7%
14	75.7%	80.0%	14	6	10	0	100.0%	62.5%	70.0%	100%
16	79.3%	73.3%	11	0	11	8	57.9%	100.0%	100.0%	57.9%
18	78.6%	90.0%	17	3	10	0	100.0%	76.9%	85.0%	100.0%
20	67.1%	70.0%	9	1	12	8	52.9%	92.3%	90.0%	60.0%

Table 4.2. 5: B300 MLPNNs Accuracy.

MLPNNs Accuracy										
Accuracy	93.3%									
N	Training Accuracy	Testing Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
2	68.1%	60.0%	5	2	13	10	33.4%	86.7%	71.4%	56.5%
4	74.5%	70.0%	8	3	13	6	57.1%	81.3%	72.7%	68.4%
6	76.6%	80.0%	14	2	10	4	77.8%	83.3%	87.5%	71.4%
8	73.8%	70.0%	10	5	11	4	71.4%	68.8%	66.7	73.3%
10	97.2%	93.3%	17	1	11	1	94.4%	91.6%	94.4%	91.7%
12	76.6%	66.7%	8	5	12	5	61.5%	70.6%	61.5%	70.6%
14	80.1%	66.7%	8	3	12	7	53.3%	80.0%	72.7%	63.2%
16	77.3%	66.7%	10	1	10	9	52.6%	90.9%	90.9%	52.6%
18	73.8%	70.0%	11	6	10	3	78.6%	62.5%	64.7%	76.9%
20	78.0%	73.3%	11	5	11	3	78.6%	68.8%	68.8%	78.6%

Table 4.2. 6: B350 MLPNNs Accuracy.

MLPNNs Accuracy										
Accuracy	90.6%									
N	Training Accuracy	Testing Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
2	87.5%	78.1%	10	4	15	3	76.9%	78.9%	71.4%	83.3%
4	86.2%	75.0%	10	0	14	8	55.6%	100.0%	100.0%	63.6%
6	86.8%	87.5%	17	1	11	3	85.0%	91.7%	94.4%	78.6%
8	86.8%	78.1%	10	2	15	5	66.7%	88.2%	83.3%	75.0%
10	87.5%	87.5%	15	1	13	3	83.3%	92.9%	93.8%	81.3%
12	85.5%	81.3%	11	0	15	6	64.7%	100.0%	100.0%	71.4%
14	76.3%	81.3%	7	0	19	6	53.8%	100.0%	100.0%	76.0%
16	87.5%	87.5%	11	2	17	2	84.6%	89.5%	84.6%	89.5%
18	85.5%	90.6%	18	1	11	2	90.0%	91.7%	94.7%	85.0%
20	86.2%	84.4%	13	0	14	5	72.2%	100.0%	100.0%	74.0%

Table 4.2.6 represents MLPNNs Neurons range that has been used for B350 concrete which shows some parameters that can be obtained from Confusion Matrix based on (TP, TN, FP, and FN), like Sensitivity, Specificity, Precision, and Negative Prediction and the best accuracy was 90.6% at N = 10.

Table 4.2.7 shows the range of MLPNNs Neurons which has been used for B400 concrete which was classified in this table below by using MLPNNs, the best accuracy was 90.0% at N = 20, which shows some parameters that can be obtained from Confusion Matrix based on (TP, TN, FP, and FN). These parameters were Sensitivity, Specificity, Precision, and Negative Prediction.

Table 4.2.8 represents the results of applying the Support Vector Machine Technique for each type of concrete which was classified in this table below and this table below shows some parameters that can be obtained from Confusion Matrix based on (TP, TN, FP, and FN). These parameters were Sensitivity, Specificity, Precision, and Negative Prediction.

Table 4.2. 7: B400 MLPNNs Accuracy.

MLPNNs Accuracy										
Accuracy	90.0%									
N	Training Accuracy	Testing Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
2	67.4%	70.0%	9	2	12	7	56.3%	85.7%	81.8%	63.2%
4	78.0%	83.3%	12	4	13	1	92.3%	76.5%	75.0%	92.9%
6	84.4%	73.3%	11	4	11	4	73.3%	73.3%	73.3%	73.3%
8	83.7%	76.7%	12	5	11	2	85.7%	68.8%	70.6%	84.6%
10	82.3%	86.7%	13	4	13	0	100.0%	76.5%	76.5%	100.0%
12	85.1%	80.0%	12	6	12	0	100.0%	66.7%	66.7%	100.0%
14	84.4%	83.3%	11	4	14	1	91.7%	77.8%	73.3%	93.3%
16	84.4%	80.0%	16	6	8	0	100.0%	57.1%	72.7%	100.0%
18	84.4%	80.0%	17	4	7	2	89.5%	63.6%	81.0%	77.8%
20	81.6%	90.0%	19	3	8	0	100.0%	72.7%	86.4%	100.0%

Table 4.2. 8: SVM for All Types of Concrete Accuracy.

SVM Accuracy									
Type	Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
B200	80.4%	103	11	61	29	78.0%	84.7%	90.4%	67.8%
B250	66.5%	61	56	72	11	84.7%	56.3%	52.1%	86.8%
B300	68.3%	61	43	77	21	74.4%	64.2%	58.7%	78.6%
B350	83.3%	82	26	99	9	90.1%	79.2%	75.9%	91.7%
B400	80.6%	90	16	72	23	79.6%	81.8%	84.9%	75.8%

Table 4.2.9 represents the results of applying the Ensemble algorithm technique for each type of concrete which was classified in this table below and shows some parameters that can be obtained from accuracy. These parameters were Sensitivity, Specificity, Precision, and Negative Prediction.

Table 4.2. 9: Ensemble for All Types of Concrete Accuracy.

Ensemble Accuracy									
Type	Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
B200	90.2%	102	12	82	8	92.7%	87.2%	89.5%	91.1%
B250	75.5%	88	29	63	20	81.5%	68.5%	75.2%	76.0%
B300	79.2%	88	16	72	26	77.2%	81.8%	84.6%	73.4%
B350	85.0%	88	20	96	12	88.0%	82.8%	81.5%	88.9%
B400	78.6%	87	19	71	24	78.4%	78.9%	82.1%	75.0%

4.2.1 B200 Concrete Classification

MLPNNs technique was more accurate than the linear support vector machine (SVM) and Ensemble algorithm (ES) in B200 concrete type.

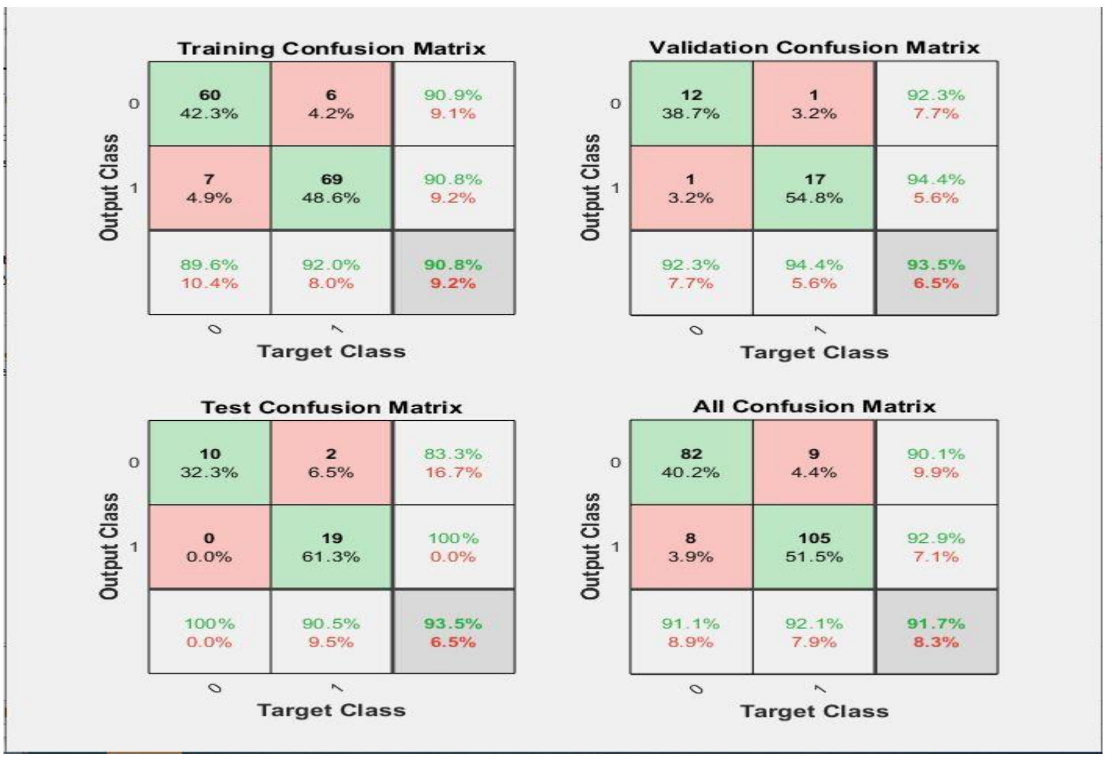


Figure B200 - 1: Confusion Matrix with B200 concreteby MLPNNswhen N=4.

In this Figure, B200 - 1, some experiments were made by changing the number of neurons, and the results are more accurate than others when several neurons equal 4, 8, or 10. The results have obtained from these figures show the confusion matrix and accuracy percentages training, validation, and testing, the percentages were 90.8%, 87.1%, and 93.5% respectively. These figures represent the accuracy was obtained for the B200 concrete type when using the MLPNNs technique is more accurate than other techniques. Finally, after some calculations from these parameters (TP, TN, FP, and FN) in CM, it is obtained that sensitivity, specificity, and precision percentages equal 90.5%, 100.0%, and 100.0% in order. Figure B200-2, shows the receiver operating characteristic (ROC) curve for B200 concrete type.

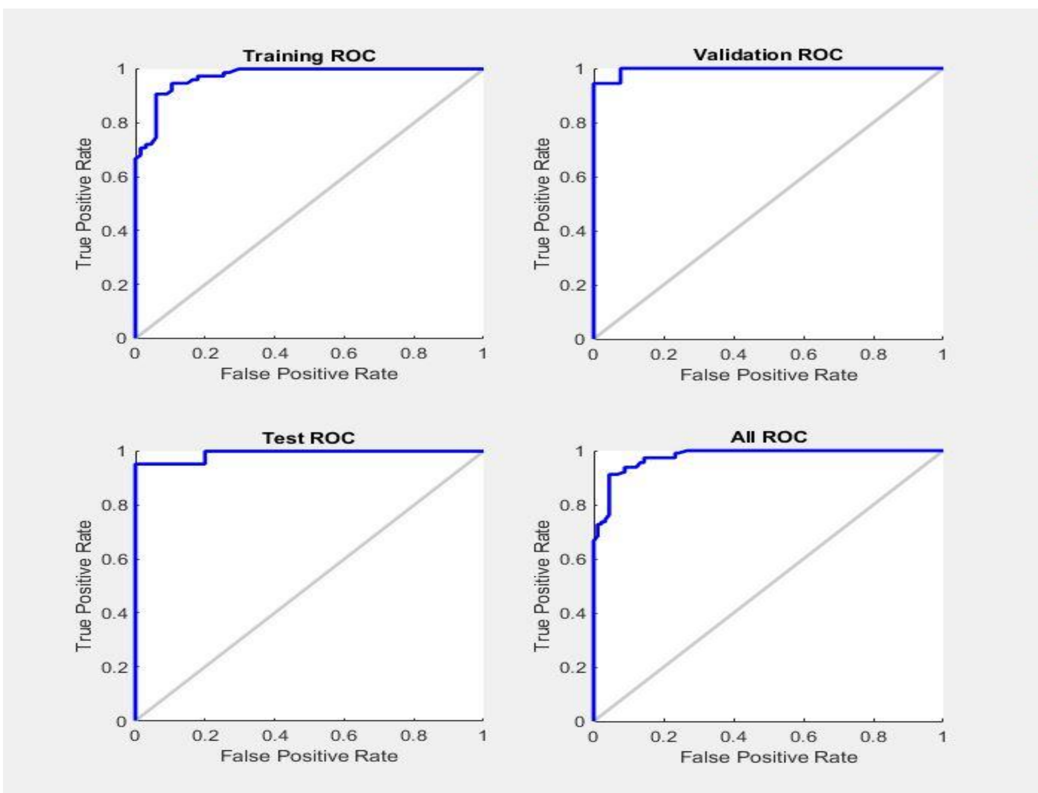


Figure B200 - 2: ROC Curve with B200 concrete by MLPNNswhen N=4.

The whole line with blue color for all figures represents the ROC curve for training, validations, and testing datasets. The ROC curve represents sensitivity versus (1-Specificity) and for B200 concrete type the sensitivity is 90.5% and the Specificity is

100%, these results are very well in network performance and good when the points are in the upper- left corner.

B200 – SVM: In this Figure, B200 -3, which was a snapshot from classification application in mat-lab and represents the Confusion matrix for B200 concrete dataset that was taken from Palestinian Governorates produced by SVM.

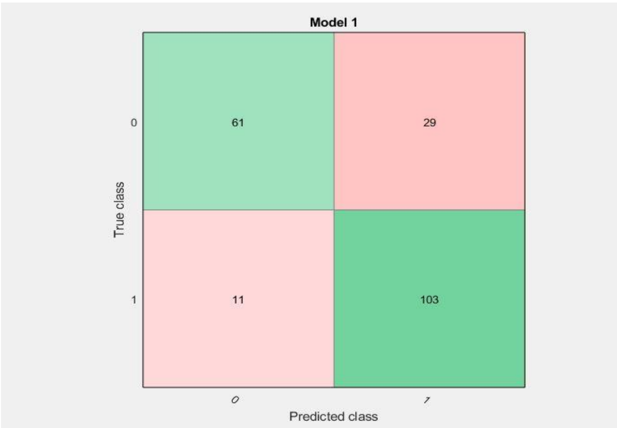


Figure B200 - 3: Confusion Matrix with B200concrete by SVM technique.

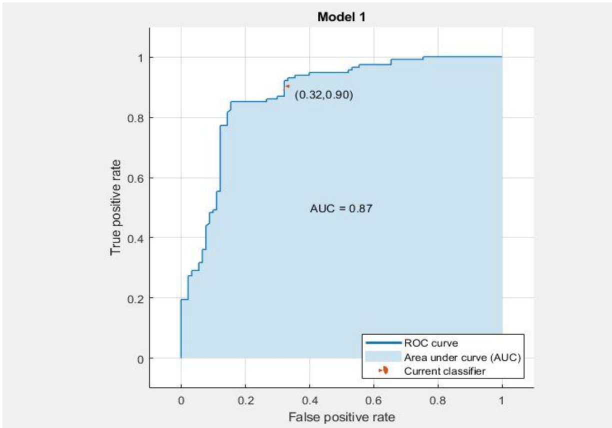


Figure B200 - 4: ROC curve with B200 concreteby SVM technique.

In Figure B200 -3, it is shown that the percentage for testing accuracy was 80.4%. These figures represent that the accuracy was obtained for the B200 concrete type when using Support Vector Machine (SVM). Some calculations were obtained from these results based on these parameters (TP, TN, FP, and TN). It is obtained that sensitivity, specificity, and precision percentages equal 78.0%, 84.7%, and 90.4% respectively.

In figure B200 -4, The whole line with blue color represents the ROC curve, the ROC curve represents sensitivity versus (1-specificity), the sensitivity was 78.0% and the specificity was 84.7% for B200 concrete type. These results are very well in network performance. A good result when points were in the upper- left corner area under curve occupies 87% from this curve, and the current classifier is (0.32,0.90).

B200 Ensemble:

Figure B200 -5 represents the Confusion matrix for the B200 concrete type produced by the Ensemble technique (ES).

Figure B200 - 5: Confusion Matrix with B200 concrete by ES technique.

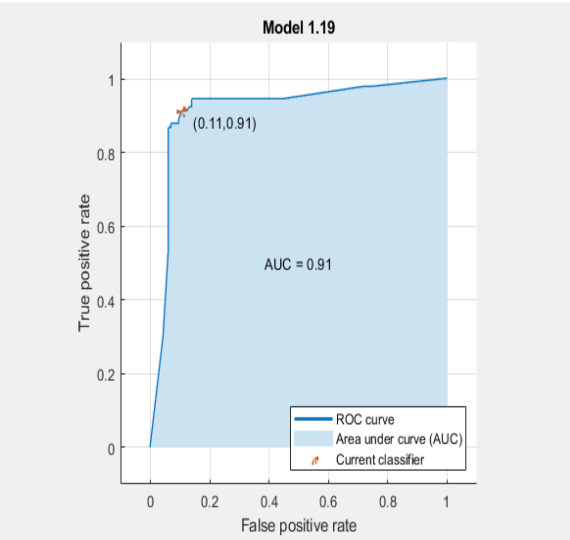


Figure B200 - 6: ROC curve with B200 concreteby ES technique.

In this figure, B200-5, the percentage for testing accuracy was 90.2%. This figure represents That the accuracy was obtained for the B200 concrete type when using the Ensemble algorithm (ES) is more accurate than support Vector Machine (SVM). Some calculations from these results are based on these parameters (TP, TN, FP, and TN). It is obtained that sensitivity, specificity, and precision percentage equal 92.7%, 87.2%, and 89.5 respectively.

In this figure, B200 -6, the whole line with blue color represents the ROC curve which represents sensitivity versus (1- Specificity). The sensitivity was 92.7% and the specificity was 87.2% for B200 concrete type. These results are very well in network performance and good results are when points were in the upper- left corner, AUC occupies 91% from this curve, and the current classifier is (0.11,0.91).

4.2.2 B250 Concrete Classification

In the B250 concrete type, the MLPNNs are more accurate than the linear support vector machine (SVM) and Ensemble algorithm(ES)when N=18 produce by using MLPNNs techniques. In this Figure B250-1, some experiments were made by changing the number of neurons, and the results are more accurate than others when the number of neurons equals 18. From the confusion matrix, training, validation, and testing percentages produced by the MLPNNs technique can be known. The percentage was 78.6%, 73.3%, and 90.0% respectively.

These figures show that the MLPNNs technique is more accurate than other techniques used for accuracy and some calculations obtained from these results. It is known that the Sensitivity, Specificity, and Precision percentages equal 100.0%, 76.9%, and 85.0% in order based on TP, TN, FP, and FN. In this Figure, B250-2 was taken from classification application in mat-lab and represents the ROC curve for the B250 concrete type.

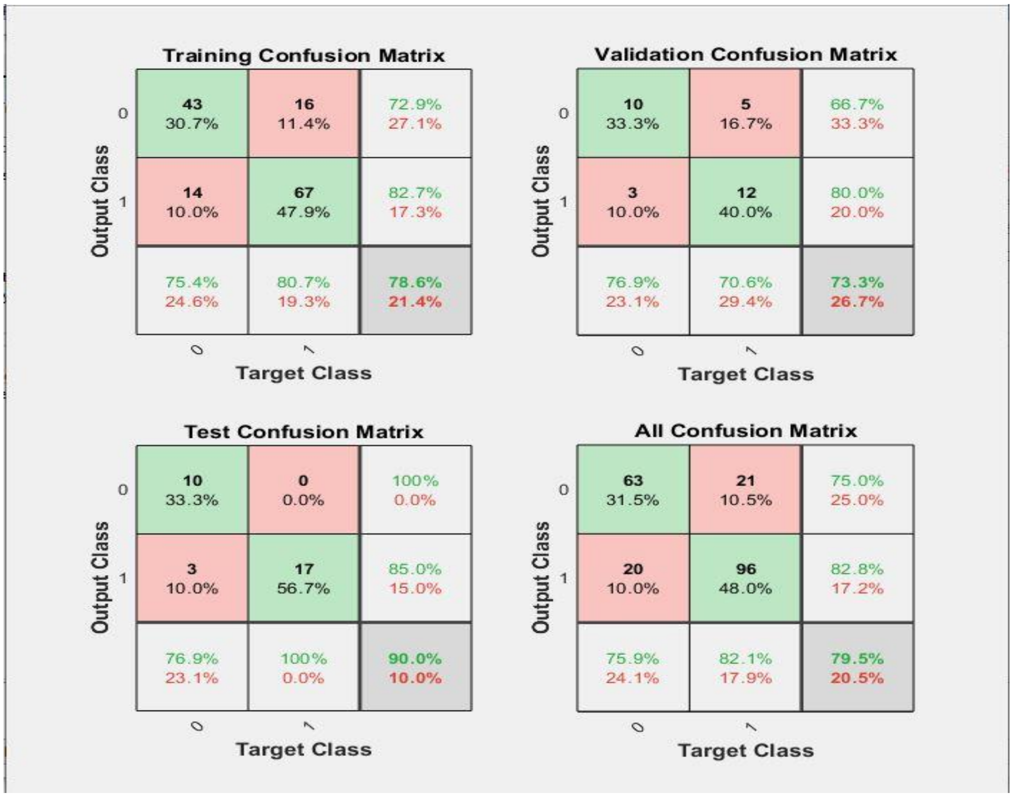


Figure B250 - 1: Confusion Matrix with B250 concrete by MLPNNswhen N=18.

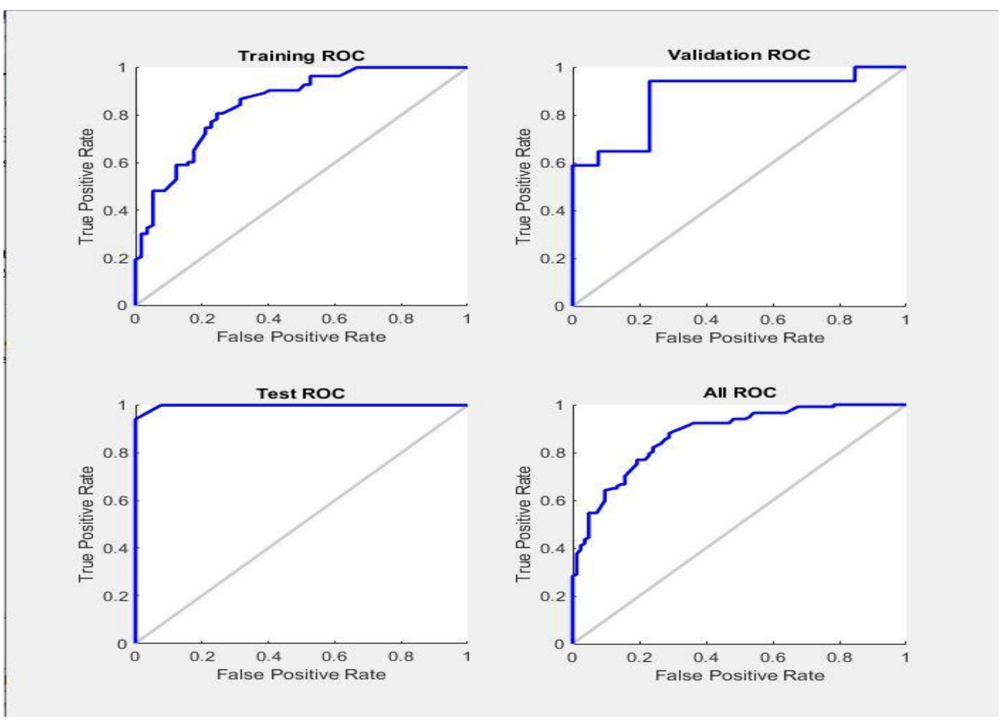


Figure B250 - 2: ROC curve with B250 concrete by MLPNNswhen N=18.

The whole line with blue color for all figures represents the ROC curve for training, validations, and testing datasets. The ROC curve is a representing sensitivity versus (1-specificity); for B250 the sensitivity is 100% and the Specificity equals 76.9%. These results are very well in network performance and good when the points are in the upper-left corner.

B250-SVM:

In this figure, B250- 3, the snapshot from classification application in mat-lab represents the Confusion matrix for B250 concrete type produced by SVM.

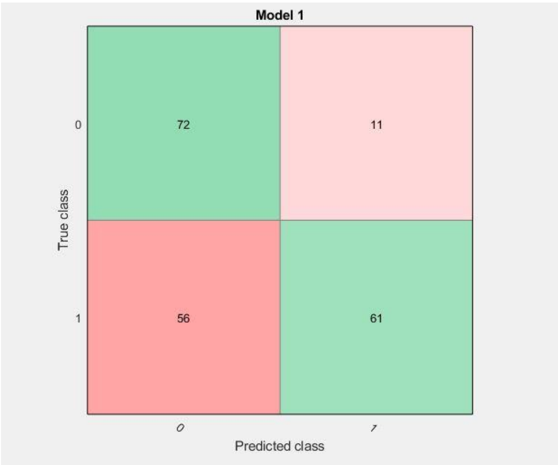


Figure B250 - 3: Confusion Matrix with B250 concrete by SVM technique.

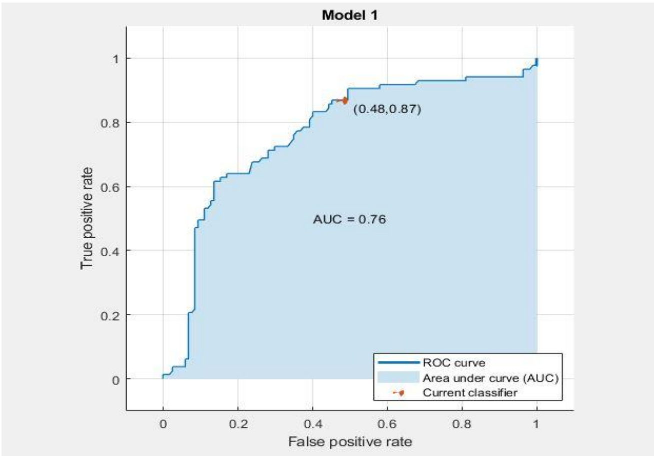


Figure B250 - 4: ROC curve with B250 concrete by SVM technique.

In this figure, B250 -3, it is shown that the percentage for testing accuracy was 66.5% produced by SVM. These figures represent that the accuracy was obtained for the B250 concrete type when using Support Vector Machine (SVM). After some calculations from these results based on these parameters (TP, TN, FP, and TN), It is obtained that Sensitivity, Specificity, and Precision percentages equal 84.7%, 56.3%, and 52.1% respectively.

In this figure, B250 -4, the whole line with blue color represents the ROC curve, which represents sensitivity versus (1- Specificity). The sensitivity was 84.7% and the specificity was 56.3% for the B250 concrete type. These results are not well in network performance and a good result when points were in the upper- left corner; here area under curve occupies 76% from this axis, and the current classifier is (0.48,0.87).

B250 Ensemble:

Figure B250\-5, which was taken from classification application by Ensemble technique (ES), represents the Confusion matrix for B250 concrete type.

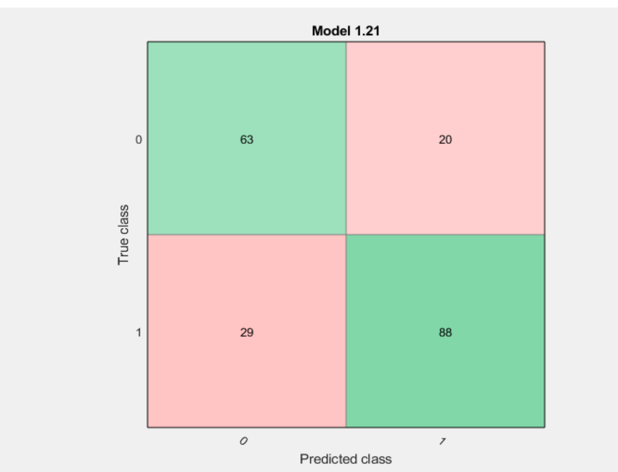


Figure B250 - 5: Confusion Matrix with B250 concrete by ES technique.

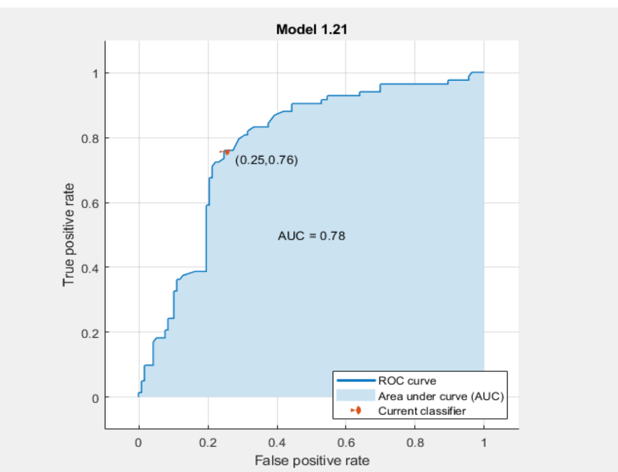


Figure B250 - 6: ROC curve with B250 concrete by ES technique.

Figure B250 -5 shows that the results from used Ensemble algorithm (ES) technique. The percentage for testing accuracy was 75.5% and it also represents the accuracy which was obtained for B250 concrete type when using Ensemble algorithm, (ES) is more accurate than support Vector Machine (SVM). It can be obtained that sensitivity, specificity, and precision percentage equal 81.5%, 68.5%, and 75.2% respectively depend on some calculations from these results based on these parameters (TP, TN, FP, and TN).

Figure B250 -6, shows that the whole line with blue color represents the ROC curve, for the B250 concrete type; the sensitivity was 81.5% and the specificity was 68.5%. These results are not well in network performance and a good result when points were in the upper- left corner; here urea under curve occupies 78% from this curve, and the current classifier is (0.25,0.76).

4.2.3 B300 Concrete Classification

MLPNNs technique is more accurate than SVM and Ensemble techniques when applying B300 concrete dataset on classification application in mat-lab. Figure B300-1 shows the confusion matrix concrete when N=10 produced by using MLPNNs technique.

Figure B300-1 shows that changing the number of neurons is done to reach the best accuracy results; the accuracy when N equals 10 is more accurate than others. The results can be obtained from training, validation, and testing percentages from the confusion matrix and these percentages were 97.2%, 96.7 %, and 93.3% respectively. These figures represent that the MLPNNs technique is more accurate than other techniques used for accuracy.

Based on some calculations from these results, it produces sensitivity, specificity, and precision percentages, and these percentages equal 94.4%, 91.6%, and 94.6% in order. Figure B300- 2 shows the ROC curve for B300 concrete type.

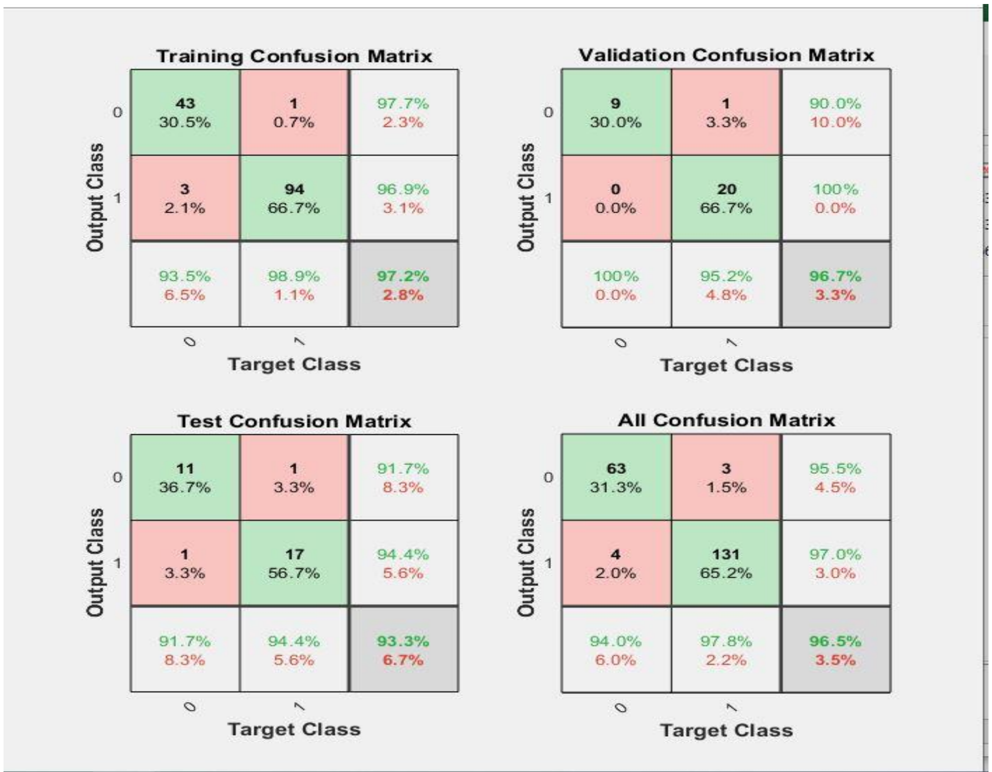


Figure B300 - 1: Confusion Matrix with B300 concrete by MLPNNs when N=10.

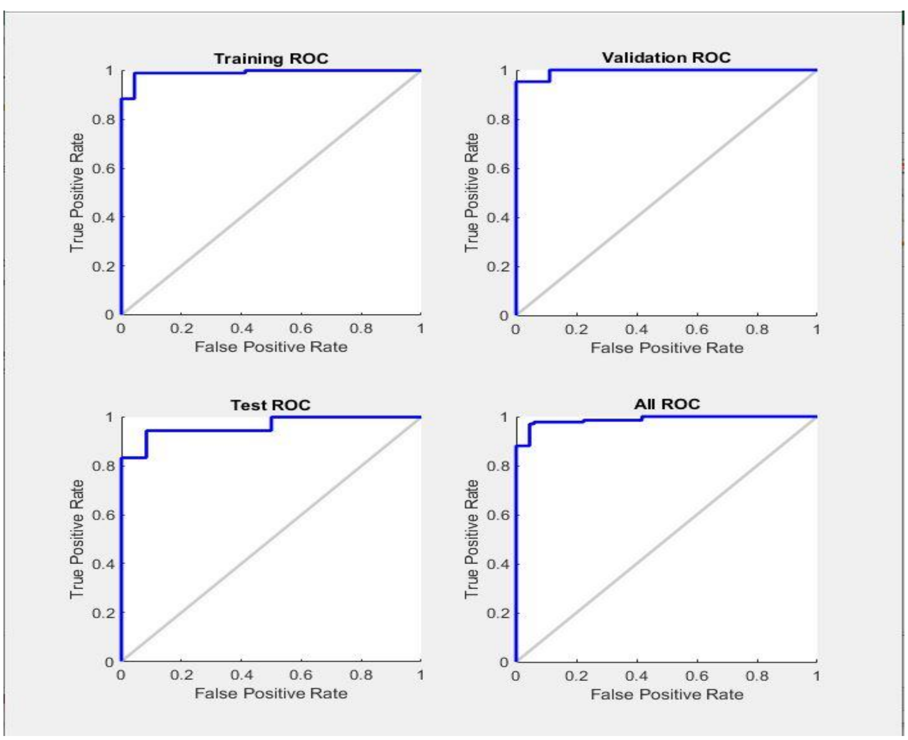


Figure B300 - 2: ROC curve with B300 concrete by MLPNNs when N=10.

The whole line with blue color for all figures represents the ROC curve for training, validations, testing, and all datasets. The sensitivity is 94.4% and the specificity equals 91.6%. These results are very well in network performance for B300 concrete type and a good result when the points are in the upper- left corner.

B300 SVM: this figure B300–3 is a snapshot from the classification application in mat-lab and represents the Confusion matrix for the B300 type.

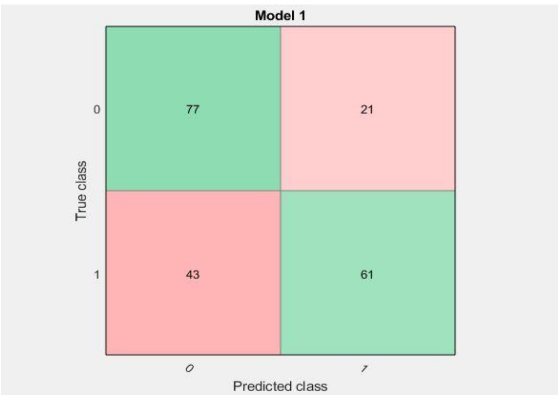


Figure B300 - 3: Confusion Matrix with B300 concrete by SVM technique.

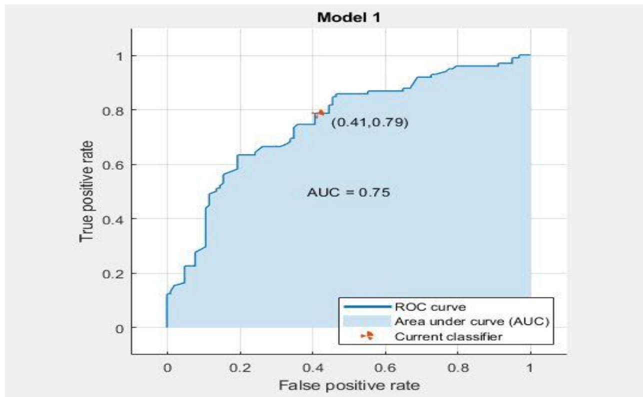


Figure B300 - 4: ROC curve with B300 concrete by SVM technique.

Figure B300-3 shows that the percentage for testing accuracy was 68.3% which is produced by SVM. These figures represent the accuracy for the B300 concrete type when using the Support Vector Machine (SVM). It can be obtained that sensitivity, specificity, and precision percentages equal 74.7%, 64.2%, and 58.7% respectively by some calculations from these results based on these parameters (TP, TN, FP, and TN).

In this figure B300 -4, the whole line with blue color represents the ROC curve, which represents sensitivity versus (1- Specificity); the sensitivity was 84.7% and the specificity was 74.7% for B300 concrete type. These results are not good in network performance and good results are when points were in the upper-left corner; here area under the curve occupies 75% of this curve, and the current classifier equals (0.41,0.79).

B300 Ensemble: Figure B300 – 5 was a snapshot from the classification application in mat-lab and represents the Confusion matrix for the B300 type.

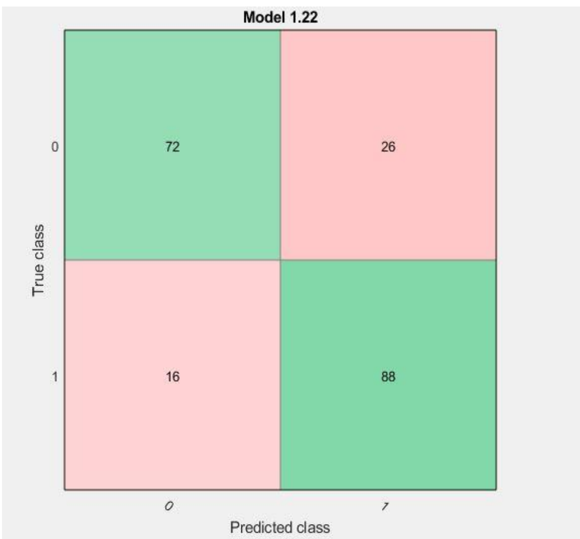


Figure B300 - 5: Confusion Matrix with B300 concrete by ES technique.

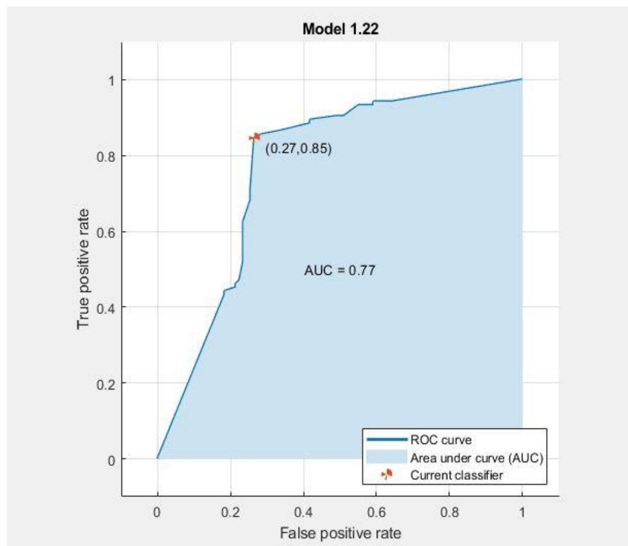


Figure B300 - 6: ROC curve with B300 concrete by ES technique.

Figure B300-5 shows That the percentage for testing accuracy was 79.2% that is produced by ES technique, which is more accurate than SVM, some calculations from these results based on these parameters (TP, TN, FP, and TN). Produce sensitivity, specificity, and precision percentages equal 77.2%, 81.8%, 84.6% respectively.

In Figure B300-6, the whole line with blue color represents the ROC curve, the sensitivity was 77.2% and the specificity was 81.8% for B300 concrete type. These results are not good in network performance and good results when points were in the upper-left corner, here the area under curve occupies 77% from this axis, and the current classifier is equals (0.27,0.85).

4.2.4 B350 Concrete Classification

MLPNNs technique is more accurate than SVM and ES techniques. Figure B350-1 represents the Confusion Matrix with B350 concrete when the Number of neurons equals 18 is produced by the MLPNNs technique.

Figure B350-1 shows MLPNNs technique when the number of neurons equals 18 is more accurate than others. The results show the Confusion Matrix and percentages for training, validation, and testing, and these percentages were 85.8%, 93.8%, and 90.6% respectively. These figures represent the accuracy for the B350 concrete type when using the MLPNNs technique and this technique is more accurate than other techniques used for accuracy. Some calculations from these results produce Sensitivity, Specificity, and Precision percentages which equal 90.0%, 91.7%, and 94.7% in order based on TP, TN, FP, and FN. Figure B350-2 was taken from classification application in mat-lab and represents the receiver operating characteristic (ROC) curve for the B350 concrete type.

The whole line with blue color for all figures represents the ROC curve for training, validations, testing datasets, and for B350 concrete type the sensitivity is 90.0% and the Specificity equals 91.7%. These results are very well in network performance and a good result is when the points are in the upper-left corner. B350 SVM: this figure, B350-3 shows a snapshot from classification application in mat-lab and represents the Confusion matrix for B350 concrete type.

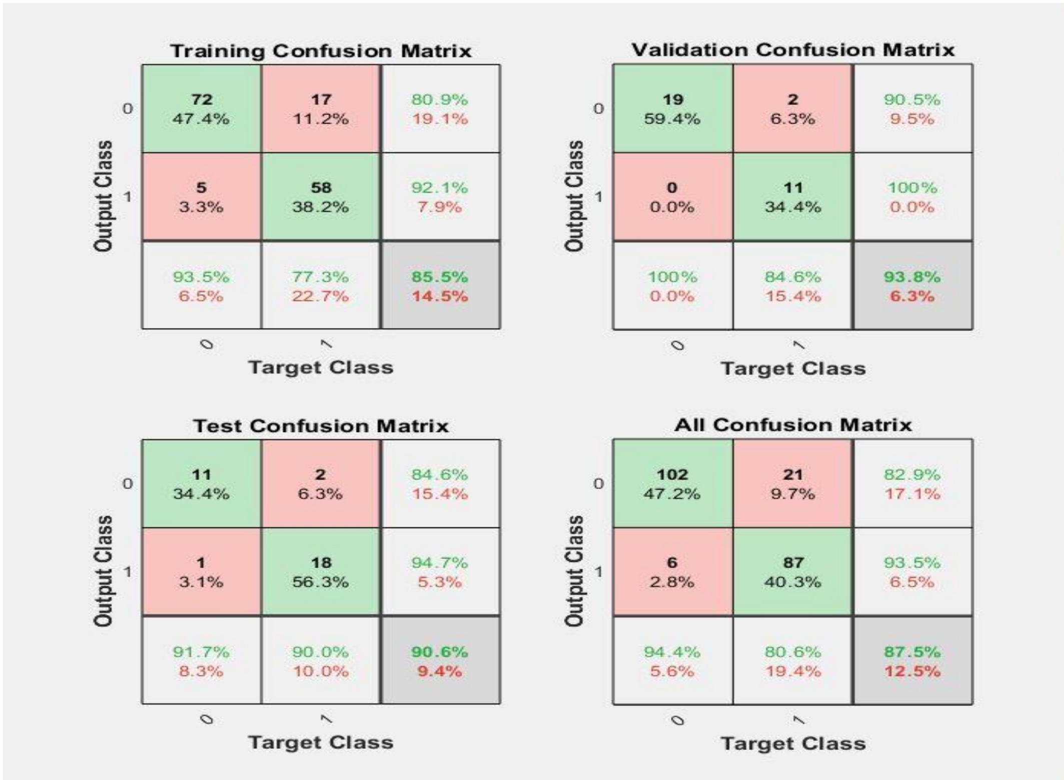


Figure B350 - 1: Confusion Matrix with B350 concrete by MLPNNs when N=18.

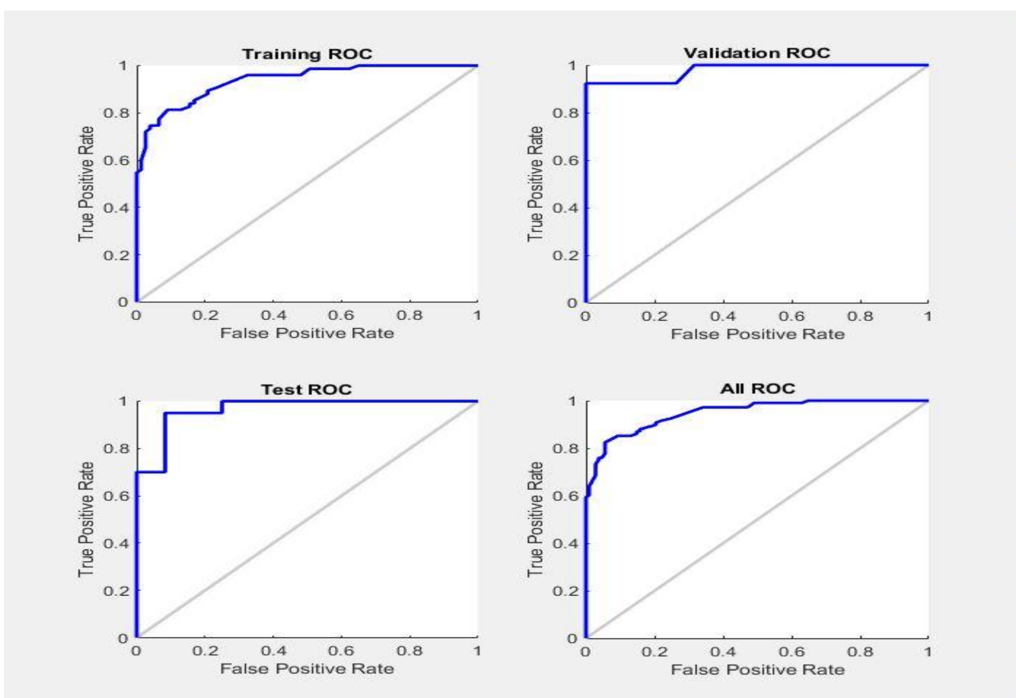


Figure B350 - 2: ROC curve with B350 concrete by MLPNNs when N=18.

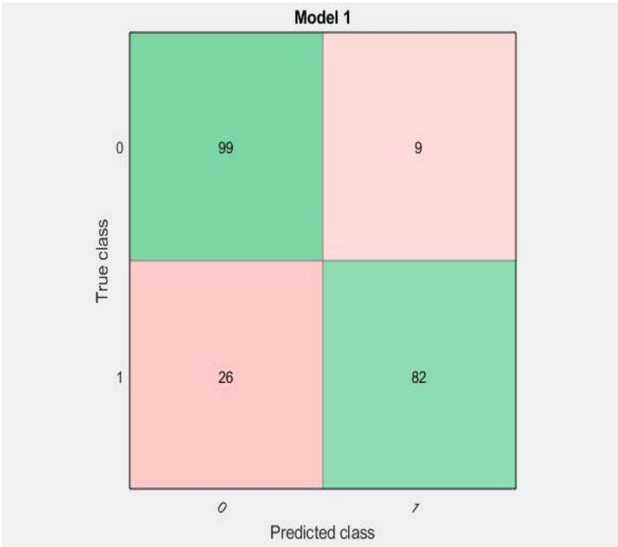


Figure B350 - 3: Confusion Matrix with B350 concrete by SVM technique.

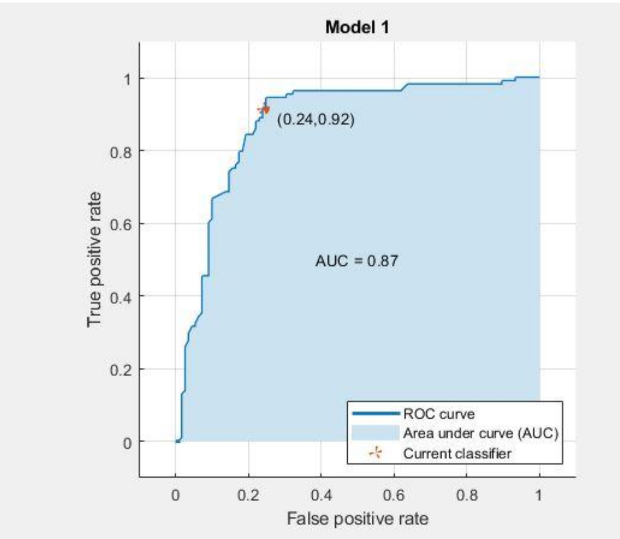


Figure B350 - 4: ROC curve with B350 concrete by SVM technique.

Figure B350-3 shows That the percentage for testing accuracy was 83.3% which is produced by SVM. This figure represents the accuracy was obtained for the B350 concrete type when using Support Vector Machine (SVM). It is obtained that Sensitivity, Specificity, and Precision percentages equal 90.1%, 79.2%, and 75.9% respectively by some calculations from these results based on these parameters (TP, TN, FP, and TN). In figure B350-4, the whole line with blue color represents the ROC curve, the sensitivity was 90.1% and the specificity was 79.2% for the B350 concrete type. These results are very well in network performance and good results when points were in the upper- left corner; here area under curve occupies 87% from this axis, and the current classifier equals (0.24,0.92). B350-Ensemble: In figure B350-5 was a snapshot from classification application in mat-lab and represents the Confusion matrix for B350 concrete dataset was taken from Palestinian Governorates.

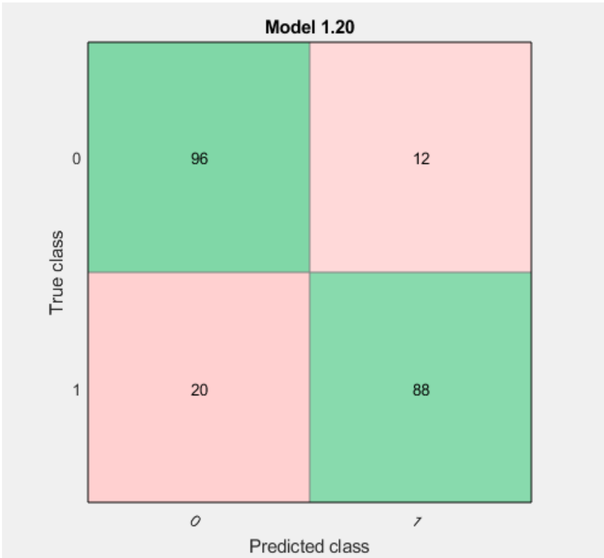


Figure B350 - 5: Confusion Matrix with B350 concrete by ES technique.

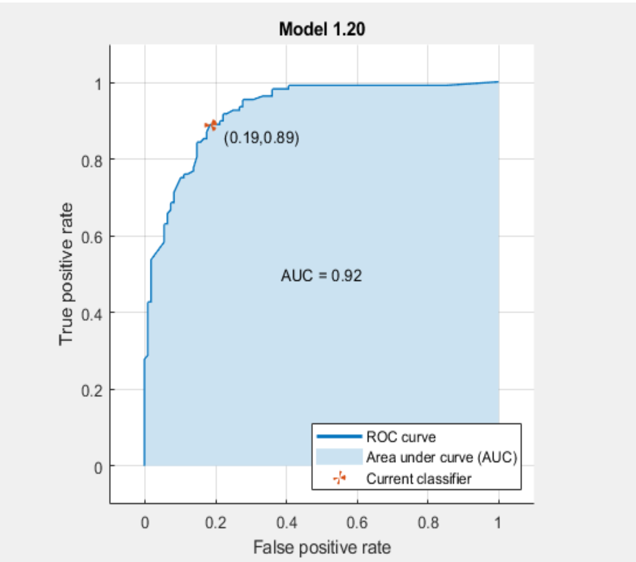


Figure B350 - 6: ROC curve with B350 concrete by ES technique.

Figure B350-5 shows that the percentage for testing accuracy was 85.0%. It also represents That the best accuracy was obtained for the B350 concrete type when using the Ensemble algorithm (ES) and it is more accurate than support Vector Machine (SVM). Finally, after some calculations from these results based on these parameters (TP, TN, FP, and TN), it can be obtained that sensitivity, specificity, and precision percentage equal 88.0%, 82.8%, and 81.5% respectively. In figure B350-6, the whole line with blue color represents the ROC curve, for B350 concrete type, the sensitivity was 88.0% and the specificity was 82.8%. These results are good in network performance, and a good result is when points were in the upper-left corner, here the area under curve occupies 92% from this axis, and the current classifier equals (0.19,0.89).

4.2.5 B400 Concrete Classification

MLPNNs technique is more accurate than the linear support vector machine (SVM) and Ensemble algorithm. Figure B400-1 shows the confusion matrix with B400 concrete when N=20 which is produced by MLPNNs technique. In figure B400-1, some experiments were made by changing the number of neurons, and the results are more accurate than others when several neurons equal 20. These results have been obtained from the MLPNNs technique. It also shows the Confusion Matrix and percentages for training, validation, and testing, the percentage was 81.6%, 90.0 %, and 90.0% respectively. These figures represent the accuracy for the B400 concrete type when using the MLPNNs technique which is more accurate than other techniques.

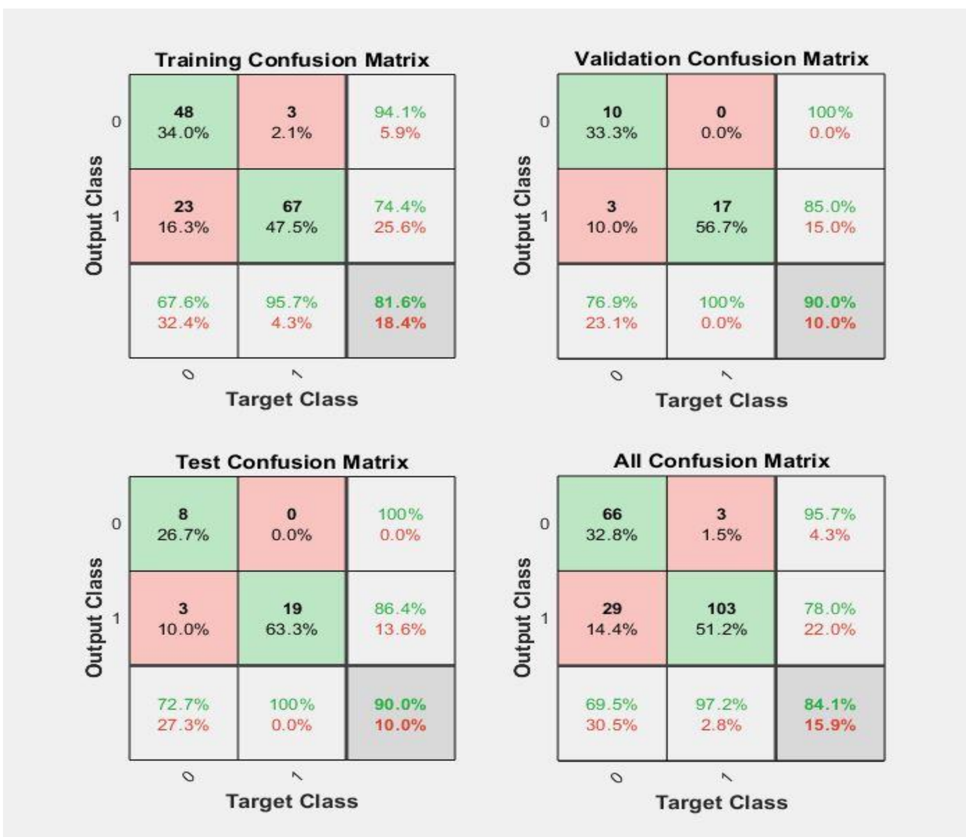


Figure B400 - 1: Confusion Matrix with B400 concrete by MLPNNs when N=20.

From these results, it can be obtained that sensitivity, specificity, and precision percentage equal 100.0%, 72.7%, and 86.4% in order. Figure B400-2 was a snapshot from classification application in mat-lab and represents the receiver operating characteristic (ROC) curve for the B400 concrete dataset which was taken from Palestinian Governorates. The whole line with blue color for all figures represents the ROC curve for training, validations, testing datasets. ROC curve is a representing sensitivity versus (1-specificity); for B400 the sensitivity is 100% and the specificity is equal to 72.7%. These results are very well in network performance and a good result when the points are in the upper-left corner.

B400 SVM: Figure, B400-3 was a snapshot from classification application in mat-lab and represents the Confusion matrix for B400 concrete type.

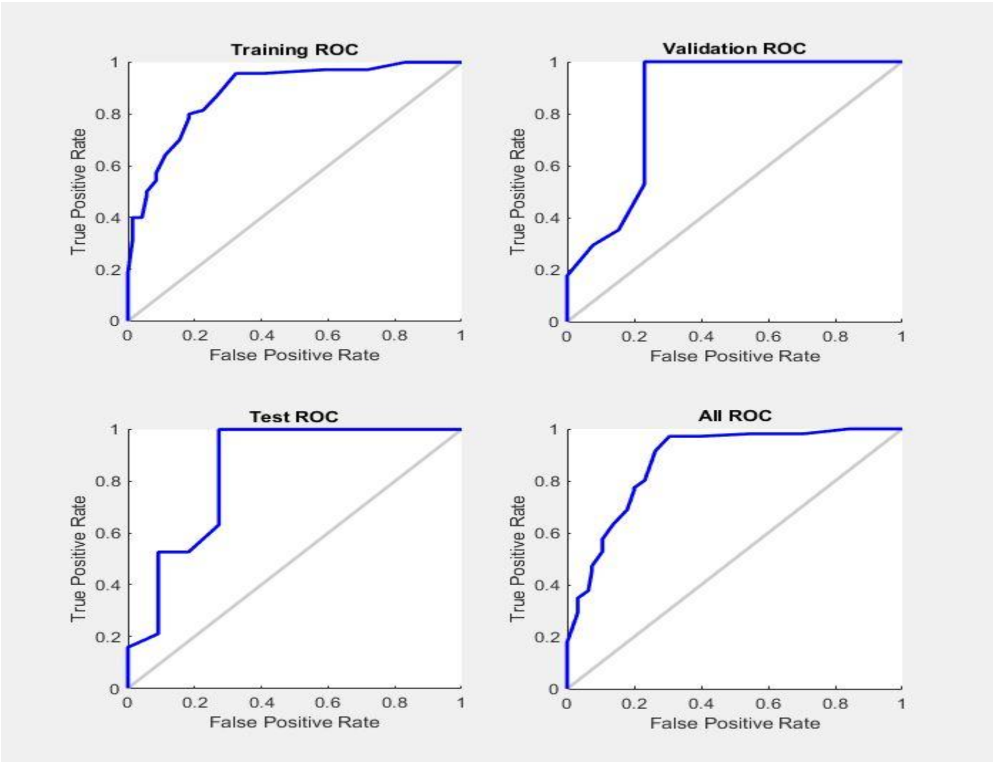


Figure B400 - 2: ROC curve with B400 concrete by MLPNNs when N=20.

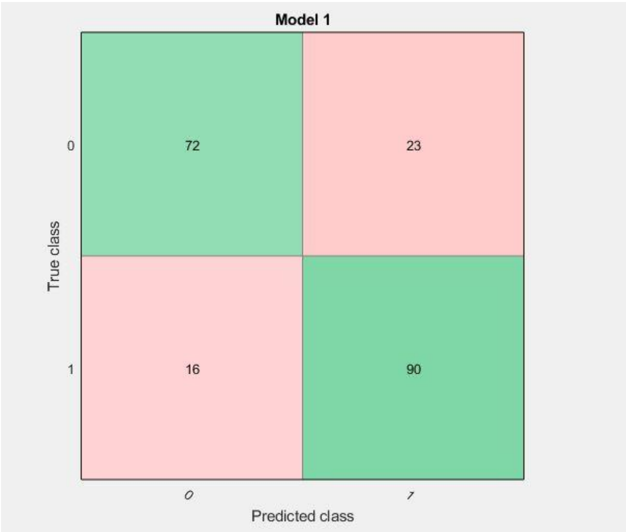


Figure B400 - 3: Confusion Matrix with B400 concrete by SVM technique.

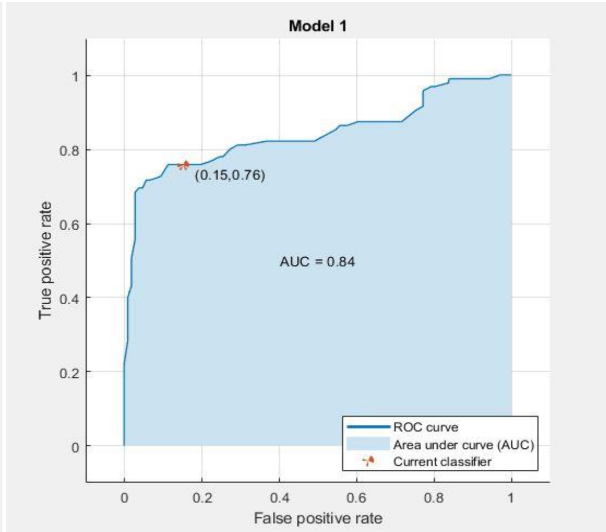


Figure B400 - 4: ROC curve with B400 concrete by SVM technique.

figure B400-3, which shows the results from using the Support Vector Machine (SVM) technique by using a classification application. It also shows the percentage for testing accuracy which was 80.6%. This technique is more accurate than the ES technique. After some calculations from these results based on these parameters (TP, TN, FP, and TN), it can be obtained that sensitivity, specificity, and precision percentage equal 79.6%, 81.8%, and 84.9% respectively. In figure B400-4, the whole line with blue color represents the ROC curve, the ROC curve represents sensitivity versus (1-Specificity), the sensitivity was 79.6% and the specificity was 81.87% for B400 concrete type. These results are very well in network performance AND good results when points were in the upper- left corner; here area under curve occupies 84% from this axis, and the current classifier equals (0.15,0.76). B400 Ensemble: In this figure, B400-6 was a snapshot from classification application in mat-lab and represents the Confusion matrix for B400 concrete dataset.

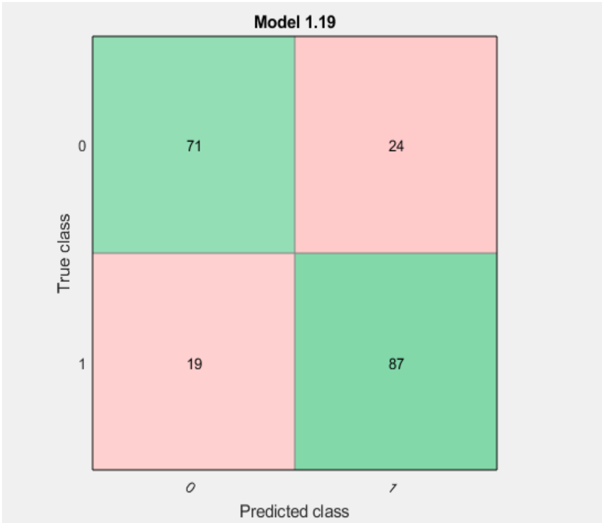


Figure B400-5: Confusion Matrix with B400 concrete by ES technique.

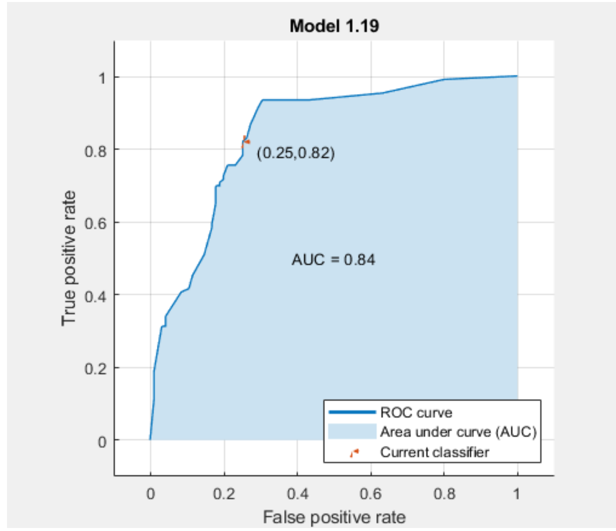


Figure B400-6: ROC curve with B400 concrete by ES technique

FigureB400 -5 represents the percentage for testing accuracy was 78.6%, also it represents the accuracy that was obtained for B400 concrete type when using the Ensemble algorithm (ES). After some calculations from these results based on these parameters (TP, TN, FP, and TN), it is obtained that sensitivity, specificity, and precision percentage equal 78.4%, 78.9%, and 82.1% respectively. In Figure B400 -6, the whole line with blue color represents the ROC curve, which represents sensitivity versus (1- Specificity); the sensitivity was 78.4% and the specificity was 78.9% for B400 concrete type. These results are very well in network performance and good results are when points were in the upper- left corner; here area under curve occupies 84% from this curve, and the current classifier equals (0.25,0.82).

4.2.6 Classification for Main Factors that Affect CCS in Palestinian Governorates.

This section from results discusses making a classification for the datasets which was collected from Palestinian governorates laboratories after it removes other parameters and remains only factors that affect Palestinian Concrete Compressive Strength (PCCS), and the new dataset was implemented on the classification models like MLPNNs, linear support vector machine, and Ensemble algorithm show the results which are close to previous experiences that were implemented on pervious datasets on Table 4.2.10

Table 4.2. 10: Summary of All Models of Concrete New Dataset Accuracy.

Type / Algorithm	**MLPNNs Accuracy**	**Linear support vector machine (SVM) Accuracy**	**Ensemble Algorithm Accuracy**
New Dataset	92.5%	75.4%	88.0%

Figure ND-1 represents the chart of accuracy results of all models used for the classification process that applied on the new dataset (ND) after removing other parameters that haven't the same effect, the results are close to previous experiences that implement on previous datasets.

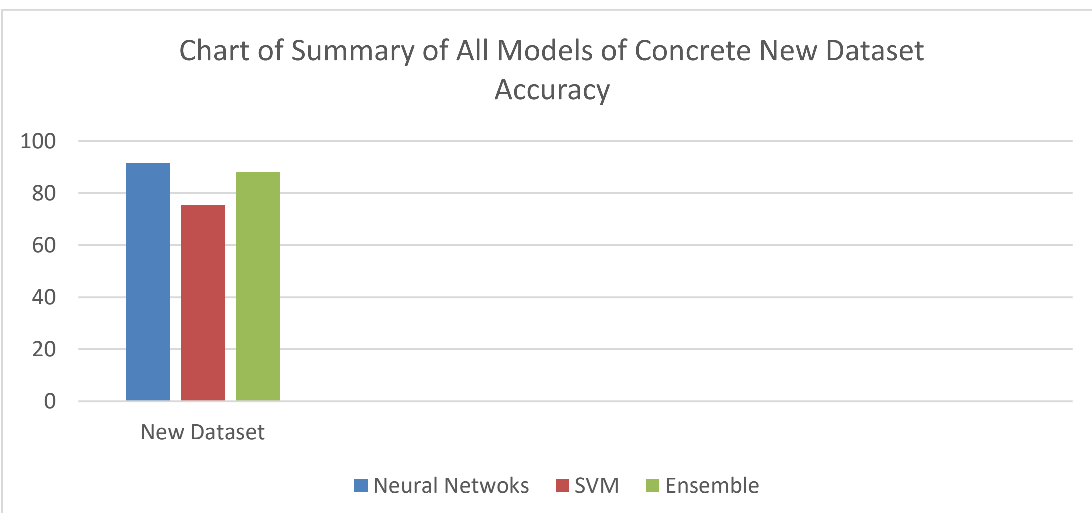


Figure ND - 1: Chart of summary of All Models of Concrete New Dataset Accuracy.

Figure ND-2 represents the chart of comparison between accuracy results for all types of concrete and the new dataset, the results show that ND accuracy is close to the accuracy of all types of Concrete.

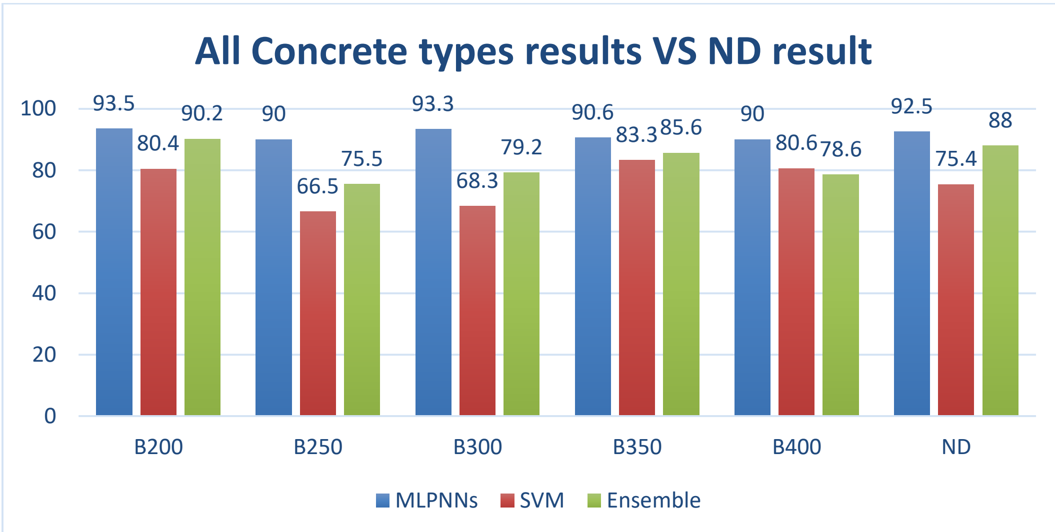


Figure ND - 2: Chart of comparison between accuracy all types of Concrete VS ND accuracy.

Table 4.2.11 shows the number of Neuron of Neural Networks that has the best accuracy for each type which is classified in this table below by using MLPNNs technique and the ranges of neurons for the New Dataset was between [2 – 20]. Table 4.2.12 represents the summary of Accuracy Results for Classification Techniques, the results showed that MLPNNs accuracy is more accurate than SVM and Ensemble for New dataset concrete.

Table 4.2. 11: Analysis of Accuracy of MLPNNs of Concrete New Dataset Accuracy.

MLPNNs Accuracy										
Accuracy	92.5%									
N	Training Accuracy	Testing Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
2	84.0%	80.4%	44	18	42	3	93.6%	70.0%	70.9%	93.3%
4	79.4%	73.9%	34	21	45	7	82.9%	68.2%	61.2%	86.5%
6	79.4%	76.6%	25	5	57	20	55.5%	91.9%	83.3%	74.0%
8	81.4%	76.6%	36	12	46	12	75.0%	79.3%	75.0%	79.3%
10	87.2%	81.3%	37	11	50	9	80.4%	81.9%	77.1%	84.7%
12	84.6%	81.3%	40	13	47	7	85.1%	78.3%	75.5%	87.0%
14	83.8%	79.4%	33	8	52	14	70.2%	86.6%	80.5%	78.8%
16	85.2%	86.9%	47	9	46	5	90.4%	83.6%	83.9%	90.2%
18	74.9%	79.4%	30	7	55	15	66.5%	88.7%	81.1%	78.6%
20	89.6%	92.5%	36	3	63	5	87.8%	95.5%	92.3%	92.7%

Table 4.2. 12: Analysis of Accuracy of All Models on Concrete New Dataset.
Summary of All Models of Concrete New Dataset Accuracy.

All Models									
Accuracy									
Type	Accuracy	TP	FP	TN	FN	Sensitivity	Specificity	Precision	Negative Prediction
MLPNNs	92.5%	36	3	63	5	87.8%	95.5%	92.3%	92.7%
SVM	75.4%	201	129	338	47	81.0%	72.3%	60.1%	87.8%
Ensemble	88.0%	296	34	333	52	85.1%	90.7%	89.7%	86.5%

MLPNNs classification on the new dataset:

In New Dataset concrete, the MLPNNs technique is more accurate than linear support vector machine (SVM) and Ensemble algorithm. Figure ND-1 represents the confusion matrix with New Dataset concrete when N=20 is produced by using MLPNNs techniques.

Figure ND - 3: Confusion Matrix with New Dataset concrete By MLPNNswhen N= 20.

In Figure ND – 3, some experiments were made by changing the number of neurons, when a number of neurons equal 20, it is more accurate than others. This figure shows the Confusion Matrix and percentages for training, validation, and testing produced by MLPNNs, the percentages were 89.6%, 88.8 %, and 92.5% respectively. It also represents the accuracy that was obtained for the New Dataset concrete type when using the MLPNNs, technique is more accurate than others. Some calculations from these results, it is obtained that sensitivity, specificity, and precision percentage equal 87.8%, 95.5%, and 92.3% in order.

Figure ND-4 was a snapshot from the classification application in mat-lab and represents the receiver operating characteristic (ROC) curve for the New Dataset concrete dataset.

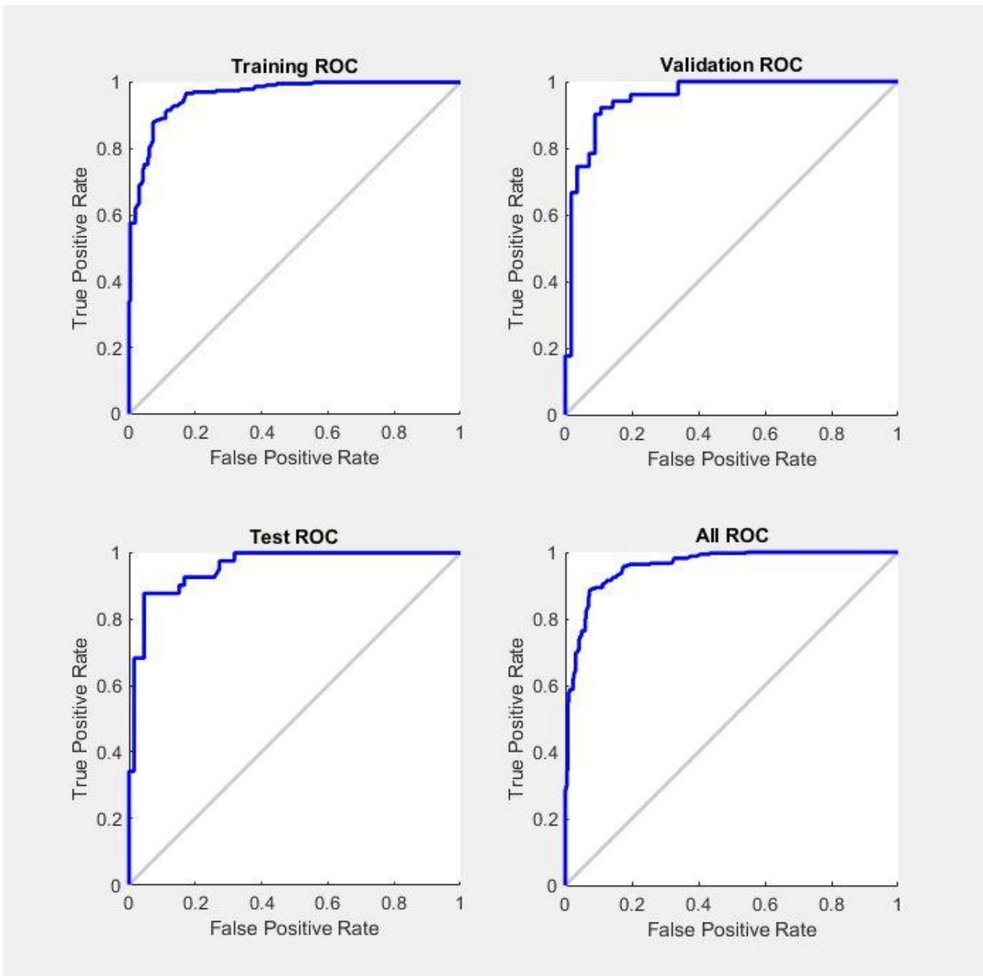
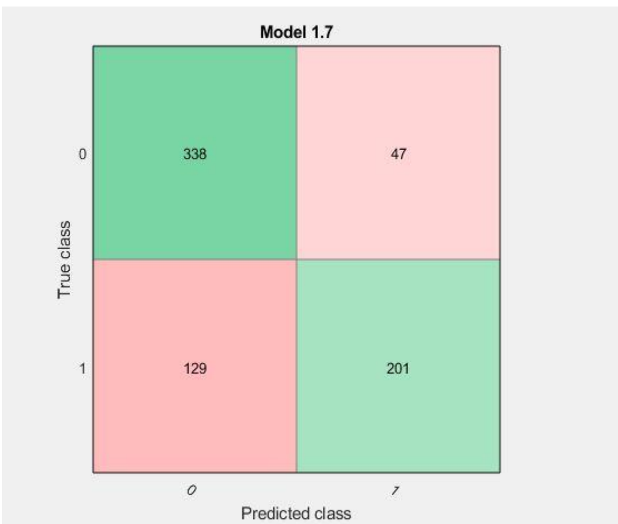


Figure ND - 4: ROC curve with New Dataset concrete by MLPNNswhen N=20.

The whole line with blue color for all figures represents the ROC curve for training, validations, testing datasets which is a representing sensitivity versus (1-specificity). For the new dataset, the sensitivity is 87.8% and the specificity equals 95.5%. These results are very well in network performance and a good result when the points are in the upper-left corner.

ND SVM: this figure, ND-5, was a snapshot from the classification application in mat-lab and represents the Confusion matrix for new dataset concrete dataset was taken from Palestinian Governorates.



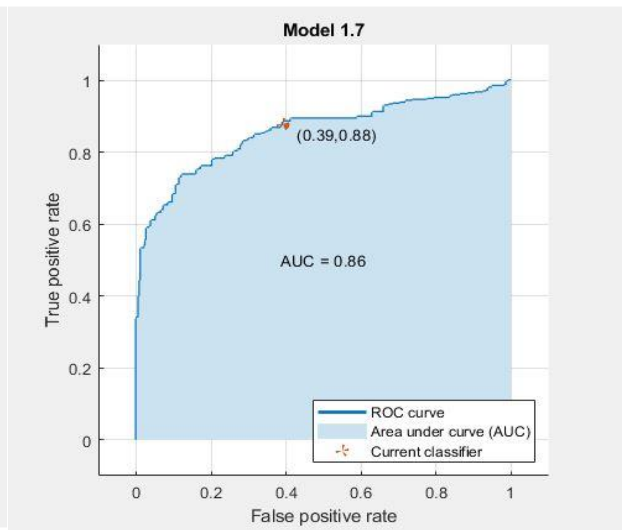


Figure ND - 5: Confusion Matrix with New Dataset concrete by SVM technique.

Figure ND - 6: ROC curve with New Dataset concrete by SVM technique.

Figure ND-5 shows the results from the Support Vector Machine (SVM) technique by classification application in mat-lab, it also shows the percentage for testing accuracy which was 75.4%. This figure represents the accuracy for the new dataset concrete type when using Support Vector Machine (SVM). After some calculations from these results based on these parameters(TP, TN, FP, and TN), it is obtained that sensitivity, specificity, and precision percentage equal 81.0%, 72.4%, and 60.9% respectively.

In Figure ND -6, the whole line with blue color represents the ROC curve, which represents sensitivity versus (1- Specificity); the sensitivity was 81.0% and the specificity was 72.4% for the new dataset concrete type. These results are very well in network performance and good results are when points were in the upper- left corner; here area under curve occupies 86% from this curve, and the current classifier equals (0.39,0.88).

ND Ensemble: this figure, ND-7, was a snapshot from the classification application in mat-lab and represents the Confusion matrix for New Dataset concrete dataset was taken from Palestinian Governorates.

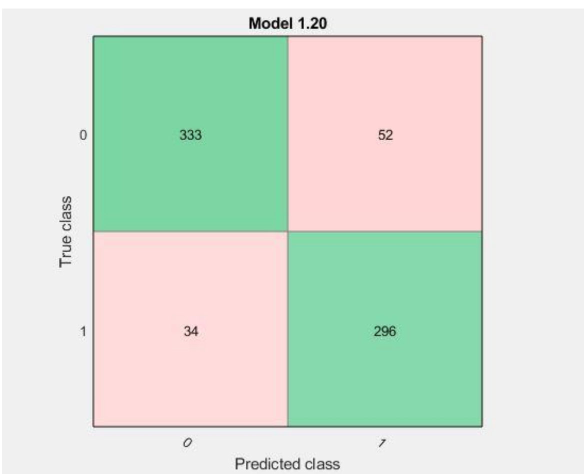


Figure ND - 7: Confusion Matrix with New dataset concrete by ES technique.

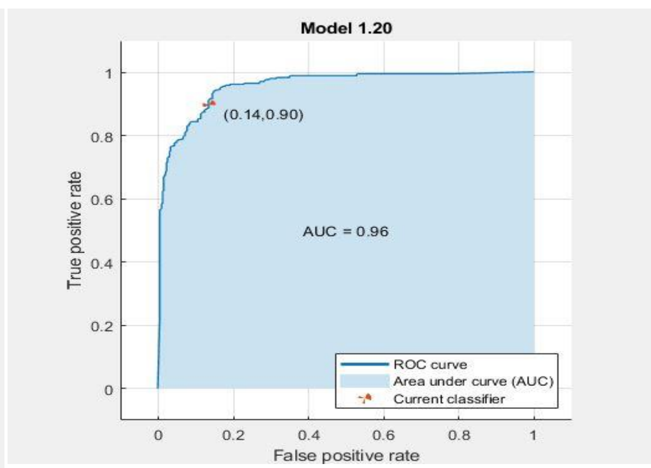


Figure ND - 8: ROC curve with New dataset concrete by ES technique.

Figure ND-7 shows the results from using the Ensemble algorithm (ES) technique by classification application in mat-lab, it also shows the percentage for testing accuracy which was 88.0%. This figure represents the best accuracy that was obtained for the new dataset concrete type when using the Ensemble algorithm (ES). It is more accurate than Support Vector Machine (SVM), and finally, after some calculations from these results based on these parameters, it is obtained that sensitivity, specificity, and precision percentages equal 85.1%, 90.7%, and 89.7% respectively. In Figure ND- 8, the whole line with blue color represents the ROC curve which represents sensitivity versus (1- Specificity). The sensitivity was 85.1% and the specificity was 90.7% for the new dataset concrete type, These results are very well in network performance and good results are when points were in the upper- left corner; here area under curve occupies 96% from this curve, and the current classifier equals (0.14,0.90).

4.3 Prediction Results

Initially, Min-Max normalization methods were applied to the Palestinian Concrete Compressive Strength dataset to optimize the dataset. The experiments were performed three times on the dataset. These experiments were performed on the Palestinian Concrete Compressive Strength dataset using MLPNNs, RBFNNs, and RNNs. The result showed that the Min-Max normalization got better performance regarding prediction. MLPNNs, RNNs, and RBFNNs were designed and the results were shown in the table based on the average mean square error. As it showed that the recurrent layers over 22n obtained the lowest mean square error as shown in Table AP 1 and Figure P1.

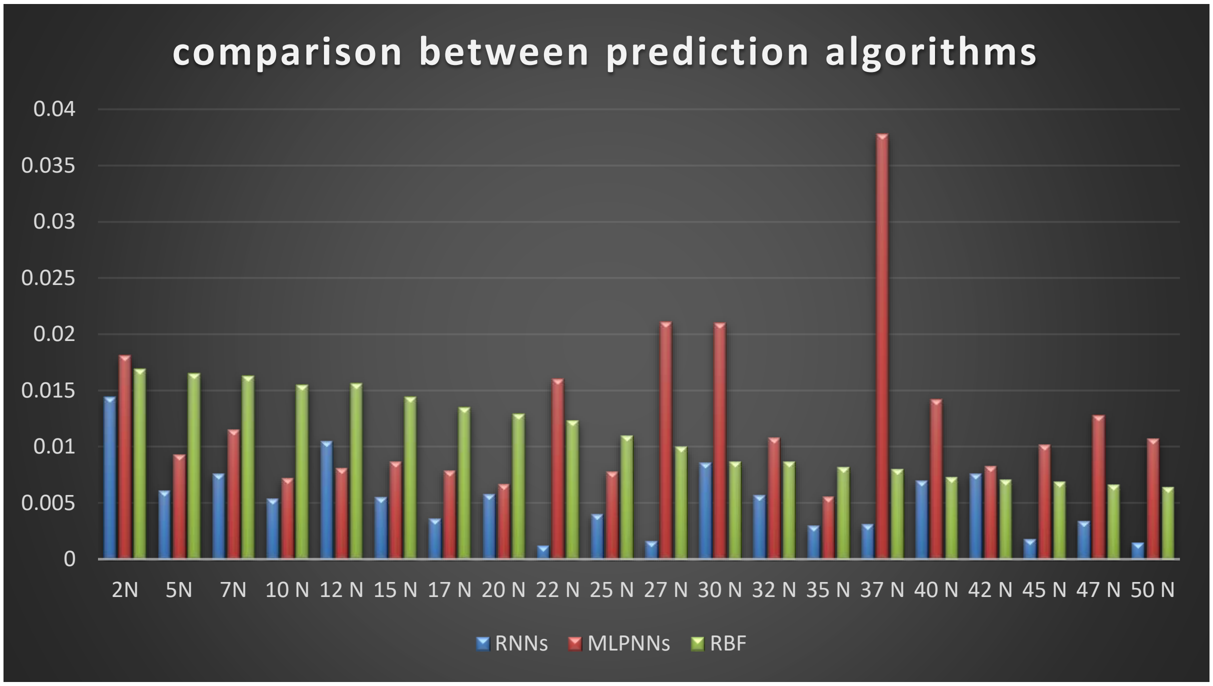


Figure P1: Comparison between Mean Square Error with all prediction Models.

Table AP 1: Comparison between Mean Square Error with all Models.

Number of neurons	RNNs	MLPNNs	RBF
2N	0.0144	0.0181	0.0169
5N	0.0061	0.0093	0.0165
7N	0.0076	0.0115	0.0163
10 N	0.0054	0.0072	0.0155
12 N	0.0105	0.0081	0.0156
15 N	0.0055	0.0087	0.0144
17 N	0.0036	0.0079	0.0135
20 N	0.0058	0.0067	0.0129
22 N	0.0012	0.0160	0.0123
25 N	0.0040	0.0078	0.0110
27 N	0.0016	0.0211	0.0100
30 N	0.0086	0.0210	0.0087
32 N	0.0057	0.0108	0.0087
35 N	0.0030	0.0056	0.0082
37 N	0.0031	0.0378	0.0080
40 N	0.0070	0.0142	0.0073
42 N	0.0076	0.0083	0.0071
45 N	0.0018	0.0102	0.0069
47 N	0.0034	0.0128	0.0066
50 N	0.0015	0.0107	0.0064

Figure P2 shows that the highest percentage for Testing, Training, and validation from MLPNNs when several neurons equal 35. MSE is 0.0056 when the number of neurons equals 35, the testing percentage was 93.3% as shown in this figure below, and the training percentage was 97.7%, validation percentage was equal to 87.9% and overall percentage together was 95.3%. The testing, training, validation, and overall percentage were 93.3%, 97.7%, 87.9%, and 95.3% respectively. The best figures that were obtained from trying and changing several neurons from (0, 2, 5, 7, 10, 12, ….., 50) and the least mean square error was obtained from these techniques as shown in Table AP1 when a number of neurons were 35 and the mean square error was 0.0056 when a comparison is made between these algorithms, the least means square error was obtained from RNNs was 0.0012 at several neurons was 22.

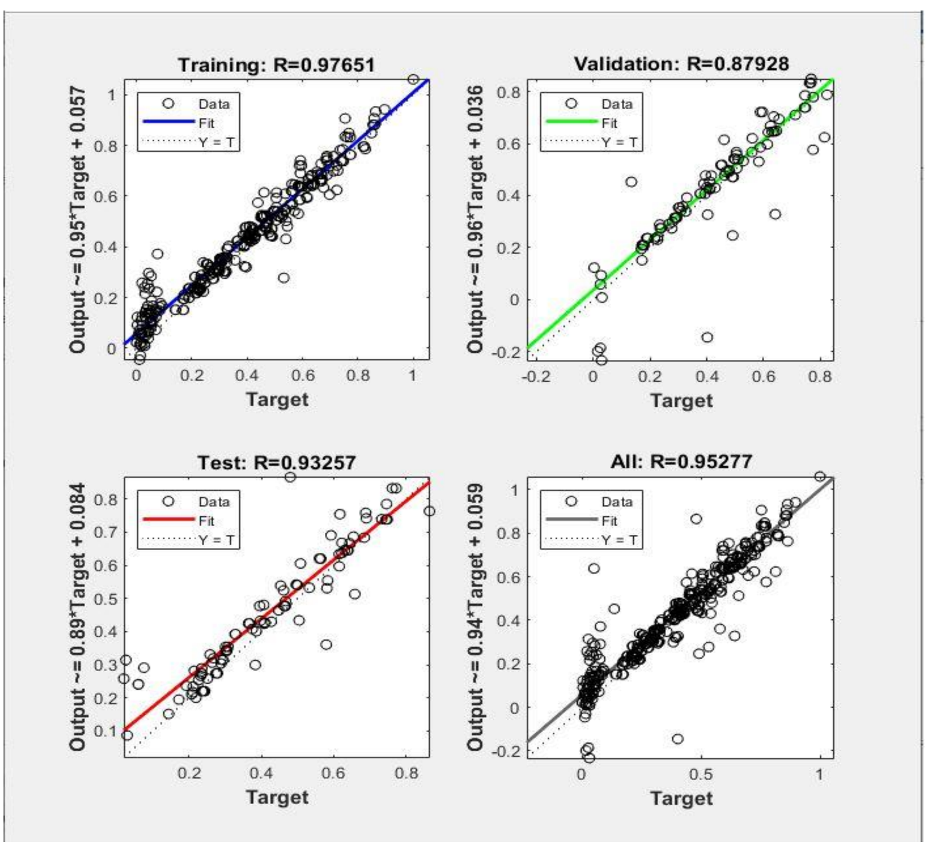


Figure P2: prediction of linear regression of Neural Fitting Tool (NFTOOL) when number of neurons was 35.

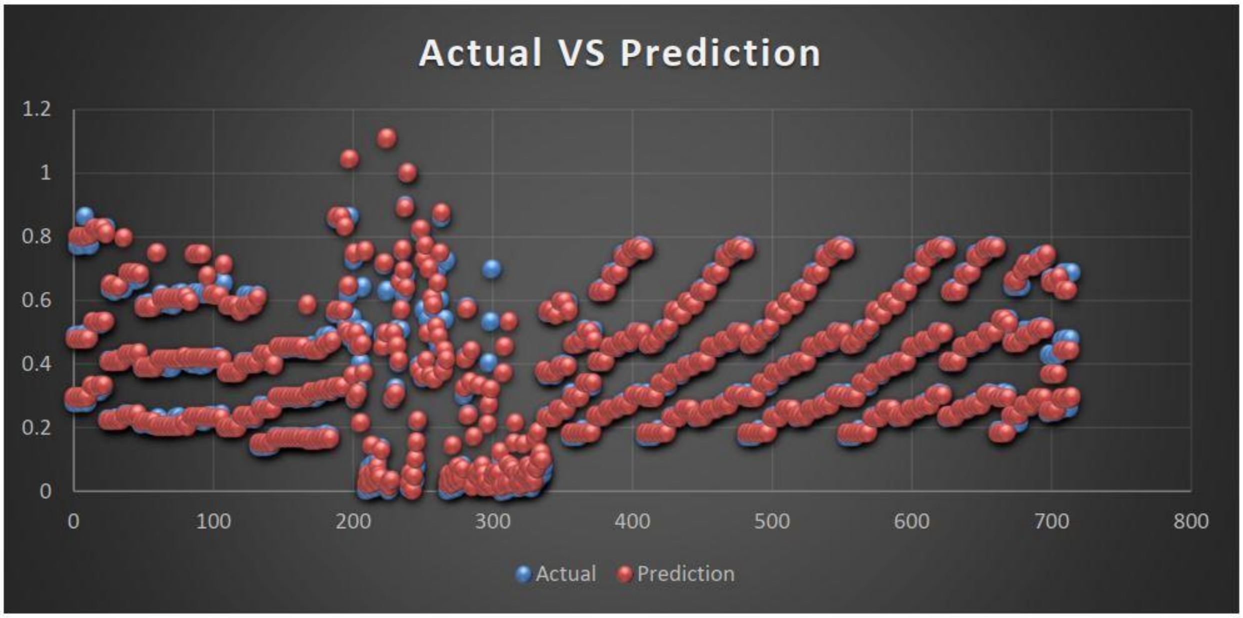


Figure P3: Comparison between Actual and Prediction on Palestinian Governorates Concrete Compressive Strength.

Figure P3 is a graphic that represents the comparison between the actual data with the prediction dataset generated from the RNNs, and the MSE was 0.0012 when the number of neurons equals 22 the comparison between Actual and the Prediction on Palestinian Governorates Concrete Compressive Strength (PCCS).

4.4 Challenges and Limitation

Several challenges and obstacles have appeared during the collection of data used in this research. The most important problem faced in this research is the quality of the classification and the prediction which are related to the dataset quality and the nature of the data. This issue was solved using the preprocessing and feature selection techniques which in general aims to organize the data to be suitable for Machine Learning techniques. Data normalization is one of the most important types of preprocessing data that were used to improve the dataset that was collected in the prediction process.

Chapter 5

Conclusion and Future Work

5.1 Conclusion and Future Works

Machine Learning Techniques (MLT) in the fields of prediction and classifications of Concrete Compressive Strength in the Palestinian governorates have excellent effectiveness in this research and have shown excellent results that can be applied in the real life.

Many algorithms have been used in the prediction phase of the strength of Concrete Compressive in the Palestinian governorates including MLPNNs, RBFNNs, and RNNs, and the study proved that the RNNs have a mean square error (MSE) of 0.0012 at 22 neurons.

In the process of classifiying all types of concrete in the Palestinian governorates, many algorithms were used to give accurate results, including MLPNNs, SVM, and Ensemble and it has been noticed that the MLPNNs were the best algorithm with an accuracy of about 90% for all types of concrete.

In the process of detecting the most important factors that affect the Compressive Strength of concrete in the Palestinian governorates, the following clustering algorithms were used: K-Mean, KSOM, and EM. In EM and KSOM algorithms, EM algorithm is completely identical to the KSOM algorithms it depends on the standard deviation of the input that was entered so that mathematically, according to a special analysis of the standard deviation in the algorithms, then account these factors are considered and considering them the factors affecting the strength of the concrete's compressive. In the K-Means algorithm, the data were divided on a certain number of clusters, then the largest value of the concrete compressive strength is taken, and the whole cluster is also taken so that it is looked at the specific results that have been determined in the cluster of the largest value if the recorded value is greater than the original value in the first column, then its factors influencing concrete Compressive Strength are considered.

Based on the results of each algorithm, different factors were deduced from each algorithm, then the common elements found by the algorithms were adopted. This work was applied to the entire dataset of Palestine, and also was applied to each of the Palestinian governorates. The results of which showed that the obtained factors differ from one governorate to another.

After applying the factors affecting the Compressive Strength of concrete in the Palestinian governorates in the classification process, the data set consisted of 4 variables which are w/c ratio, superplasticizer, location, and age, it showed great agreement with the results that appeared in the classifications in the Compressive Strength of concrete in the Palestinian governorates so that the accuracy was also about 90% when using neural networks.

Future Work

A hybrid system that integrates Machine Learning Techniques (MLT) like thinking of integrates Genetic algorithm (GA) with Multilayer Perceptron Neural networks (MLP) will be created to have a GA-MLP hybrid system. Another hybrid system that can contain Particle Swarm Optimization (PSO) algorithm with Multilayer Perceptron Neural networks (MLP) and Genetic algorithm (GA) can be made. Moreover, another hybrid system that consists of Genetic algorithm (GA) with Particle Swarm Optimization (PSO) namely GA-PSO can be made so that it shows better results and accuracy than using single techniques, and Palestinian laboratories will be communicated to inform them of the effectiveness of using MLT in the process of prediction and classification of Concrete Compressive Strength in the Palestinian governorates, and inform each governorate about any factor affecting mainly the Compressive Strength of Concrete for each governorate separately.

Bibliography

Richart, F. E., Brandtzæg, A., & Brown, R. L. (1928). A study of the failure of concrete under combined compressive stresses. University of Illinois at Urbana Champaign, College of Engineering. Engineering Experiment Station.

Vairagade, V. S., & Kene, K. S. (2012). Introduction to steel fiber reinforced concrete on engineering performance of concrete. International journal of scientific & technology research, 1(4), 141.

Malhotra, H. L. (1956). The effect of temperature on the compressive strength of concrete. Magazine of concrete research, 8(23), 85-94.

Bartlett, F. M., & MacGregor, J. G. (1996). Statistical analysis of the compressive strength of concrete in structures. Materials Journal, 93(2), 158-168.

Lee, "Prediction of concrete strength using artificial neural networks," Engineering Structures, vol. 25, no. 7, pp. 849–857,2003.

Atici, "Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network," Expert Systems with Applications, vol. 38, no. 8, pp. 9609–9618, 2011.

I. B. Topc¸u and M. Saridemir, "Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic," Computational Materials Science, vol.
41, no. 3, pp. 305–311, 2008.

Lin, C.-C. J. Lin, and K.-C. Chang, "Neural network based methodology for estimating bridge damage after major earthquakes," Journal of the Chinese Institute of Engineers, vol.
25, no. 5, pp. 415–424, 2002.

A. Oztas ¨ ¸, M. Pala, E. Ozbay, E. Kanca, N. C ¨ ¸ aglar, and M. A. Bhatti, "Predicting the compressive strength and slump of high strength concrete using neural network," Construction and Building Materials, vol. 20, no. 9, pp. 769–775, 2006.

Kewalramani and R. Gupta, "Concrete compressive strength prediction using ultrasonic pulse veloGovernorate through artificial neural networks," Automation in Construction, vol. 15, no. 3, pp. 374–379, 2006.

Alexandridis, D. Triantis, I. Stavrakas, and C. Stergiopoulos, “A neural network approach for compressive strength prediction in cement-based materials through the study of pressure stimulated electrical signals,” Construction and Building Materials, vol. 30, pp. 294–300, 2012.

Hola and Schabowicz, “Methodology of neural identification of strength of concrete,” ACI Materials Journal, vol. 102, no. 6, pp. 459–464, 2005.

Trtnik, Kavci ˇ c, and Turk, “Prediction of concrete ˇ strength using ultrasonic pulse veloGovernorate and artificial neural networks,” Ultrasonics, vol. 49, no. 1, pp. 53–60, 2009.

Gupta, Sakshi. "Using artificial neural network to predict the compressive strength of concrete containing nano-silica." Civil Engineering and Architecture 1.3 (2013): 96-102.

Ibrahim, A. S. Prediction Model of the Concrete Compressive Strength Using Neural Network Regression.

Wang, J. Z. (2000). Prediction of compressive strength of concrete by neural networks. Cement and Concrete Research, 30(8), 1245-1250.

Ming, Zhao. (2011). “The Study of BP Neural Network in Predicting Concrete Strength”. Advanced Materials Research. 243-249. 6169-6173. 10.4028/www.scientific.net/AMR.243-249.6169.

Park, Jong & Yoon, Younggeun & Oh, Tae. (2019). “Prediction of Concrete Strength with P-, S-, R-Wave Velocities by Support Vector Machine (SVM) and Artificial Neural Network (ANN)”. Applied Sciences. 9. 4053. 10.3390/app9194053.

Chopra, Palika & Sharma, Rajendra & Kumar, Maneek. (2016). “Prediction of Compressive Strength of Concrete Using Artificial Neural Network and Genetic Programming”. Advances in Materials Science and Engineering. 2016. 1-10. 10.1155/2016/7648467.

Tenza-Abril, Antonio & Satorre-Cuerda, Rosana & Rosique, Patricia & Navarro-Gonzalez, F. & Villacampa, Yolanda. (2019). “COMPRESSIVE STRENGTH CLASSIFICATION OF LIGHTWEIGHT AGGREGATE CONCRETE USING A SUPPORT VECTOR MACHINE MODEL”. 173-183. 10.2495/CMEM190171.

Prof.Yousef Saleh Abu Zir, Eng. Saleh Yousef Abu Zir “Data Mining Techniques for Prediction of Concrete Compressive Strength (CCS)”. Palestinian Journal of Technology & Applied Sciences No. 3 - January 2020

Khademi, Faezehossadat & Jamal, Sayed Mohammadmehdi. (2016). Predicting the 28 Days

Compressive Strength of Concrete Using Artificial Neural Network. i-manager's Journal on Civil Engineering. 6. 10.26634/jce.6.2.5936.

Yousif, Salim & Abdullah, Salwa. (2009). ARTIFICIAL NEURAL NETWORKS MODEL FOR PREDICTING COMPRESSIVE STRENGTH OF CONCRETE. Tikrit Journal of Eng. Sciences. Vol.6. ,(55-63).

Naderpour, H. & Rafiean, Amir Hossein & Fakharian, Pouyan. (2018). Compressive strength prediction of environmentally friendly concrete using artificial neural networks. Journal of Building Engineering. 16. 10.1016/j.jobe.2018.01.007.

Sayed Ahmed, Mahmoud. (2012). Statistical Modelling and Prediction of Compressive Strength of Concrete. Concrete Research Letters. 3. 452-458. 10.6084/M9.FIGSHARE.105905.

Duan, Zhenhua & Kou, S.C. & Poon, Chi Sun. (2013). Prediction of compressive strength of recycled aggregate concrete using artificial neural networks. Construction and Building Materials. 40. 1200-1206. 10.1016/j.conbuildmat.2012.04.063.

Alshihri, Mar & Azmy, Ahmed & Elbisy, Moussa. (2009). Neural networks for predicting compressive strength of structural light weight concrete. Construction and Building Materials - CONSTR BUILD MATER. 23. 10.1016/j.conbuildmat.2008.12.003.

Nikoo, Mehdi & Sadowski, Lukasz & Torabian, Farshid. (2015). Prediction of Concrete Compressive Strength by Evolutionary Artificial Neural Networks. Advances in Materials Science and Engineering. 2015. 10.1155/2015/849126.

Torre, Ana & Garcia, F & Moromi, I & Espinoza, P & Acuña, L. (2015). Prediction of compression strength of high performance concrete using artificial neural networks. Journal of Physics: Conference Series. 582. 10.1088/1742-6596/582/1/012010.

Bilim, Cahit & Atiş, Cengiz & Tanyildizi, Harun & Karahan, Okan. (2009). Predicting the compressive strength of ground granulated blast furnace slag concrete using artificial neural network. Advances in Engineering Software. 40. 334-340. 10.1016/j.advengsoft.2008.05.005.

Duan, Zhenhua & Kou, S.C. & Poon, Chi Sun. (2014). Using artificial neural networks for predicting the elastic modulus of recycled aggregate concrete. Construction and Building Materials. 44. 524–532. 10.1016/j.conbuildmat.2013.02.064.

Hola, J. & Schabowicz, Krzysztof. (2005). Application of artificial neural networks to determine concrete compressive strength based on non-destructive tests. Journal of Civil Engineering and Management. 11. 23-32. 10.1080/13923730.2005.9636329.

Iman & Ozbakkaloglu, Togay & Kisi, Ozgur & Xie, Tianyu. (2015). Predicting Behavior of FRP-Confined Concrete using Neuro Fuzzy, Neural Network, Multivariate Adaptive Regression Splines and M5 Model Tree Techniques. Materials and Structures. 49. 10.1617/s11527-015-0790-4.

Sadrmomtazi, A. & Sobhani, Jafar & Mirgozar, M.A.. (2013). Modeling compressive strength of EPS lightweight concrete using regression, neural network and ANFIS. Construction and Building Materials. 42. 205–216. 10.1016/j.conbuildmat.2013.01.016.

Sobhani, Jafar & Najimi, Meysam & Pourkhorshidi, Ali & Parhizkar, Tayebeh. (2010). Prediction of the compressive strength of no-slump concrete: A comparative study of regression, neural network and ANFIS models. Construction and Building Materials - CONSTR BUILD MATER. 24. 709-718. 10.1016/j.conbuildmat.2009.10.037.

Zain, M. & Abd, Suhad. (2009). Multiple Regression Model for Compressive Strength Prediction of High Performance Concrete. Journal of Applied Sciences. 9. 10.3923/jas.2009.155.160.

Yeh, Ivy. (2007). Modeling slump flow of concrete using second-order regressions and artificial neural networks. Cement and Concrete Composites. 29. 474-480. 10.1016/j.cemconcomp.2007.02.001.

Aggarwal, C. C. (2015). Data mining: the textbook. Springer.

Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman and Angela Y. Wu. "An Efficient k-means Clustering Algorithm: Analysis and Implementation"

Joaquin Perez Ortega, Ma. Del Rocio Boone Rojas and Maria J. Somodevilla Garcia. "Research issues on K-means Algorithm: An Experimental Trial Using Matlab"

The k-means algorithm - Notes by Tan, Steinbach, Kumar Ghosh.

Ahmad, Azlin & Yusof, Rubiyah. (2016). A modified kohonen self-organizing map (KSOM) clustering for four categorical data. Jurnal Teknologi. 78. 10.11113/jt.v78.9275.

R. Rojas. 1996. Unsupervised Learning and Clustering Algorithms. in Neural Networks. 99–121.

Kangas, T. K. Kohonen, and J. T. Laaksonen. 1990. Variants of self-organizing maps. IEEE Trans. Neural Netw. Berlin: Springer Berlin Heidelberg. Jan 1990. 1(1): 93–9.

Kohonen. 2000. Self-organizing Maps of Massive Document Collections. In Neural Comput. New Challenges Perspect. New Millenn. Proc. IEEE-INNS-ENNS Int. Jt. Conf. Neural Networks. 2: 3–9

Kohonen. 1982. Self-Organized Formation Of Topologically Correct Feature Maps. Biol. Cybern. 69: 59–69.

Banerjee, C. Krumpelman, and R. J. Mooney. 2005. ModelBased Overlapping Clustering. In Proceedings of the eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD '05). ACM, New York, NY, USA. 532-537.

S. Maps and T. Kohonen. 1996. New Developments and Applications of. 1: 164–172.

IEOR E4570: Machine Learning for OR&FE Spring 2015 c 2015 by Martin Haugh The EM Algorithm

Issa Allam Zaki, "2000 Neural networks, architecture, algorithms, etc." Applications "Dar Shuaa.

Flood . I, Kartam. N, 1994, " NEURAL NETWORKS IN CIVIL ENGINEERING; PRINCIPLES AND UNDERSTANDING ", Associate Members, ASCE.

Flood . I, Kartam. N, 1994, " NEURAL NETWORKS IN CIVIL ENGINEERING, "NEURAL NETWORKS IN CIVIL ENGINEERING; SYSTEMS AND APPLICATION ", Associate Members, ASCE

Mansour. M.Y, Dicleli. M, Lee. J.Y, Zhang. J, 2004, " Predicting the shear strength of reinforced concrete beams using artificial neural networks "

Brain Maker Professional, Manual of Brain Maker Software, California Scientific

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4), 115-133.

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric Environment, 32(14-15), 2627-2636.

Kumar Satapathy, Satchidananda Dehuri, Alok Kumar Jagadev, Shruti Mishra Empirical Study on the Performance of the Classifiers in EEG Classification

Author links open overlay panel Sandeep, “ EEG Brain Signal Classification for Epileptic Seizure Disorder Detection 2019, Pages 45-65 “

Phung, . Bouzerdoum and G. H. Nguyen, "*Learning Pattern Classification Tasks with Imbalanced Data Sets*," in Pattern Recognition, InTech, 2009

M. T. Hagan and M. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Transactions on Neural Networks, 5(6), 989–993, 1994

Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 39(4):561–577

الملخص

يعد إستخدام أساليب تعلم الآلة من احدى الأساليب التي تمتاز وتحقق فاعلية مميزة في عملية تنبؤ وتنصيف قوة الضغط للخرسانة في المحافظات الفلسطينية، في البداية قمنا بجمع البيانات المطلوبة للبحث من المختبرات والمصانع الخاصة بالخرسانة من سبعة محافظات فلسطينية.

لقد تم تجزئة العمل الى ثلاثة أجزاء، في الجزء الأول، تم العمل باستخدام ثلاثة خوارزميات مصممة للتجميع منها خوارزمية التجميع K ، خوارزمية EM وخوارزمية KSOM وقد أظهرت النتائج أهم العوامل المؤثرة في المحافظات الفلسطينية بشكل فردي. و أظهرت أهم العوامل المؤثرة في قوة ضغط الخرسانة في فلسطين بشكل عام يمكن الاستنباط من نتائج خوارزمية EM and KSOM وخوارزمية K التجميع تستخدم في التنبؤ بشكل أدق.

في الجزء الثاني من العمل ، تم استخدام البيانات من المختبرات الفلسطينية في عملية التصنيف بين أنواع الخرسانة في فلسطين، ومنها B200,B250,B300,B350,B400 باستخدام ثلاثة خوارزميات وهي Ensemble و الشبكات العصبونية متعددة الطبقات، و شعاع الدعم الآلي SVM ، بحيث أظهرت النتائج أن الشبكات العصبونية المتعددة الطبقات تتفوق بالدقة على غيرها، بحيث كانت الدقة تقريباً 90% لكل نوع من الخرسانة، بينما في الخوارزميات الاخرى مثل شعاع الدعم الآلي والـ Ensemble كانت 68%،80% بالترتيب، وقد تم تطبيق عملية التصنيف على البيانات المستخرجة من أهم العوامل المؤثرة في قوة ضغط الخرسانة بحيث كانت الدقة تقريبا متساوية.

في الجزء الثالث من العمل، تم استخدام البيانات من المختبرات الفلسطينية من المحافظات التي تم ذكرها في عملية التنبؤ، بحيث تم استخدام ثلاثة خوارزميات وهي الشبكات العصبونية متعددة الطبقات، RBFNNs ، RNNs بحيث كان نتائج RNNs هي أفضل نتيجة بحيث كان متوسط مربع الخطأ يساوي 0.0012بينما

متوسط مربع الخطأ في الشكبات العصبونية المتعددة الطبقات و RBFNNs كانت 0.0064 ,0.0107 بالترتيب، وقد أظهرت النتائج أن أساليب تعلم الالة في عملية التنبؤ والتصنيف هي اداة فعالة في قوة ضغط الخرسانة في المحافظات الفلسطينية.