Arab American University

Faculty of Graduate Studies

Prediction and Classification Analytics of Obesity Datasets Using a Hybrid Model of Clustering and Neuro-Fuzzy Methods

By

Younes Nedal Younes Hantoli

Supervisor

Prof. Dr. Mohammed Awad

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

March / 2021

Prediction and Classification Analytics of Obesity Datasets Using a Hybrid Model of Clustering and Neuro-Fuzzy Methods

By
Younes Nedal Younes Hantoli

This Thesis was Defended Successfully on **09/03/2021** and Approved By:

Committee Members	Signature
1. Supervisor: **Prof. Dr. Mohammed Awad**	……………………………….
2. Internal Examiner: **Dr. Rami Hadrob**	……………………………….
3. External Examiner: **Dr. Yousef Awwad**	……………………………….

Dedication

I dedicate this thesis to my family, fiancée and friends for their unconditional love and support that they have shown and given to me. Special thanks to my grandfather for his support and encouragement all the time. Also, I dedicate this thesis to my thesis supervisor Prof. Dr. Mohammed Awad.

Name: Younes Nedal Hantoli

Sig:

Date: 10.1.2022

Acknowledgments

I would like to seize 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 valuable suggestions that have had a significant impact and have helped in overcoming many obstacles in writing this thesis in the best way. I would like to thank Dr. Ahmad Batran (Faculty of Nursing in AAUP) for providing us with the dataset of obesity, we declare he will participate with us in any future paper publication that is based on this thesis.

Abstract

Human health is the most important and valuable field in life. Obesity is one of the main causes of many diseases that might cause death such as diabetes, hypertension, and stroke. Artificial Intelligence (AI), which has offered inference tools to support clinical decision-making. The integration of Artificial Intelligence in the field of Healthcare diagnostic is an effective method in a large number of health care applications. The diagnostic procedures for health care problems can be categorized as intelligent data, prediction, and classification tasks. Artificial intelligence techniques can be utilized to predict and classify obesity disease to give appropriate assistance to physicians in decision-making. Several intelligence techniques are being used to diagnose diseases such as a neural network, fuzzy logic, expert systems, etc.

In this research, different Artificial intelligence techniques were used to classify and predict the child's obesity. The dataset was collected from 4 cities in Palestine. The Collected data passed through data preprocessing and applying feature extraction that most closely affect the child's obesity. Where the final step is applying the AI methods to recognize the patterns in the dataset. Decision-tree, k-nearest neighbor, support vector machine, logistic regression, neural network, and a hybrid adaptive neuro-fuzzy inference system that combines fuzzy logic and neural networks, were used to recognize the pattern on the dataset and improve the results of the classification of obesity in children with high accuracy. For this neuro-fuzzy hybrid model, the membership functions used are trimf, trapmf, gaussmf, and gauss2mf, where two types of Neuro-fuzzy structures were used; grid partitioning and clustering structure. The total of fuzzy rules was 512 obtaining as an output and the degree of belonging of a child to obesity or not.

Based on the obtained results for the applied dataset that represents all cities , the hybrid adaptive Neuro-fuzzy inference system achieved a prediction accuracy of 98.33% using grid partition and using neural networks which achieved an accuracy of 98.40% which are very good results and deserved to be used in a real application to help specialists in making decisions, while the other models got accuracy as follows: Logistic Regression 97.50%, d-tree 92.30%, KNN 93.60%, and SVM 97.10%. Also another results were obtained when the same techniques were applied for each city alone, city 1 results (ANFIS-Grid 99.97% , ANFIS-Cluster 98.58%, Logistic Regression 95.20% and Neural Network 97.40%), the other detailed results are shown in chapter 4.

Table of Contents

List of Figures

List of Tables

List of Abbreviation

AI	Artificial Inelegance
ANNs	Artificial Neural Networks
FIS	Fuzzy Inference System
ANFIS	Adaptive Neuro-Fuzzy Inference System
SVM	Support Vector Machine
KNN	K-Nearest Neighbor
FCM	Fuzzy C- Means Clustering
WHTR	Waist-To-Height Ratio
WHR	Waist-To-Hip Ratio
FRBS	Fuzzy Rule-Based System
NNs	Neural Networks
SVR	Support Vector Regression
RMSE	Root-Mean-Square Error
MUAC	Mid-Upper Arm Circumference
WC	Waist Circumference
HC	Hip Circumference
MLP	Multilayer Perception
RNN	Recurrent Neural Networks
MF	Membership Function
AUC	Area Under Curve
ROC	Receiver Operating Characteristic Curve
TP	True Positive
TN	True Negative
FP	False-Positive
FN	False-Negative
TPR	True Positive Rate

Chapter 1

Introduction

1 Introduction

There is a close relationship between obesity and serious health risks. In the past fifty years, obesity has become a global problem, which is considered as one of the real causes that affect a person's health [1]. There have been many opinions about whether obesity is a disease or not. In 1977, the American Health Care Financing Administration confirmed that obesity was not a disease. In 2002, the Japanese Association for the Study of Obesity had published standards that were applied to its population, considering obesity as a disease [2]. In 2010, the Scottish Intercollegiate Guideline described obesity as follows: ‘Obesity is defined as a disease process characterized by excessive body fat accumulation with multiple organ-specific consequences’. Again after three years, the American Medical Association considered obesity as a disease and this was followed by several other health institutions around the world [2]. Finally, in 2015, obesity was considered as a medicinal state which required a medical intervention [2]. Obesity results from many factors that lead to an imbalance between the energy gained and lost during a long period. These aspects include genetics, epigenetic, physiological, behavioral, sociocultural, and environmental factors [1]. It is well known that some factors contribute to the obesity epidemic such as lack of sleep, quitting smoking, endocrine disruptors like chemicals in food, giving birth in old age, and obesity-related to medication intake [1] [3].

In the past years, information technology was used to develop many studies, also Artificial Intelligence (AI) has offered inference tools to support clinical decision-making, and other parts in science and engineering [4]. Simply, AI is a way to teach computers how to think, make decisions and act like humans, so, logically the computer should do things that people originally do (would be considered intelligent) [4]. AI employs the information gathered from electronic health records, to come up with a result to assist the doctors in decision making and diagnosing diseases [5].

Classifications techniques were used in many fields to gather similar data together, which make dealing with each similar group much easier and used for many other goals depending on the problem [6]. In general, the classification uses training data to build a classification model and then to be used in predicting the class for some new sample where the features are known but the class label is not [7].

AI is used in classification as supervised learning in which the algorithm learns from the input data and then uses this learning in building a model to predict some new data [6]. Many AI algorithms are used for classification. In the recent study, six algorithms are used to compare and come up with a conclusion. The first model is the logistic regression which, in general, sees how well the output is predicted from some sets of independent variables [8]. Logistic regression draws an S-shaped or sigmoid curve to fit the data as close as to the real values [9]. The second algorithm is the Decision Tree, this algorithm classify some data into levels by building a tree consist of three layers: root node layer, internal nodes layer, and leaf nodes layer [10]. The third model is the Support Vector Machine (SVM). Simply, this technique, SVM, aims to make boundaries between each class [11]. The fourth model is K-nearest neighbor. This model is very simple, gives a good result, and usually assumes that the similar data is close to each other, when classifying new data, it depends on the distance to each class [12].

The fifth model is an artificial neural networks (ANNs) which is a part of AI, ANNs and simulates the biological neural networks [4]. ANNs is a model with complex functions. It can be used whenever there is a connection between independent predictor variables (inputs) and dependent predicted variables (outputs), and it learns from known data and recognizes the patterns between them. Once the ANNs were learned the hidden relationship between inputs and outputs, they can predict the output from new given input data. After that, training is performed which makes the neural network predict with very high accuracy [13]. Mainly ANN was used in medical decision support systems because of its ability to learn the complex nonlinear relationship [14].

The sixth model is the Adaptive network fuzzy inference system (ANFIS) which is the combination of ANN and Fuzzy inference systems (FIS). FIS also is a part of AI, it is defined that there is a different membership degree (which the member can belong to), that allows to clean up the uncertain situations in the form of rules by using a decision-making mechanism [15]. ANFIS technique combines the benefits of the two machine learning algorithms fuzzy logic decision-making mechanism, learning ability, and relational structure of the ANN [15].

In recent studies, the first aim is to determine if the person is at risk of obesity using the neuro-fuzzy technique. Secondly, studying previous steps involved in predicting obesity and performing a comparison of classification algorithms to determine the more adequate classifier to improve the prediction accuracy in cases of obesity. Different data mining algorithms will be applied to the obesity datasets like; regression, decision trees, Support Vector Machine (SVM), and neural networks. In this thesis, all the experiments are performed on local data set from Palestine.

1.1 Objectives

Through this thesis, the main aim is to improve the accuracy of determining if the person is at risk of obesity or not using AI techniques. In other words, considering the appropriate mechanism for analyzing the data and searching for a more accurate method to assess obesity is the main aim of the study. Early classification and prediction of obesity can be obtained by using AI techniques that will be able to avoid many diseases, such as: vascular disease (that is, diabetes with cardiovascular complications) which is highly linked with obesity

Specific Objectives:

Other specific objectives of the study are the following:

- Using Clustering techniques which are k-means clustering and c-mean fuzzy clustering to group the homogeneous data
- Analyzing different types of AI techniques for obesity classification and prediction
- Selecting the most appropriate AI techniques for classification and prediction of obesity
- Applying the hybrid model of Neuro-fuzzy to classification and prediction of obesity

1.2 Contribution

In this research, artificial intelligence techniques which include logistic regression, KNN, SVM, decision tree, and neural network were applied to improve the classification and prediction of obesity on a Palestinian dataset among children with ages between 8-12 years. A hybrid model that combines NNs and fuzzy logic was proposed with different membership functions like trimf, trapmf, gaussmf, and gauss2mf, where two types of Neuro-fuzzy structures were used; grid partitioning and clustering structure to improve the classification accuracy and compare the two structures and determine which one is better in this case.

The neuro-fuzzy model uses 512 rules to obtain the output and the degree of belonging of a child to obesity or not. The neuro-fuzzy model was able to classify and predict obesity with high accuracy results comparing with other AI traditional models. The contribution of the research in the field of data mining is oriented to allow the construction and determination of a hybrid model that allows optimizing data classification in the field of health; it is aimed at allowing better prevention against obesity.

1.3 Overview

In this chapter of the thesis, the problem statement that is the obesity classification and prediction, our objectives are illustrated and contributed. The rest of the thesis are arranged as follows:

Chapter two will give a background about the dataset, then literature within the related research field will be discussed, also, it provides some other used techniques to predict and classify obesity. Chapter three will discuss in general the dataset, at first the collection and preprocessing steps will be shown, and then the general method procedures and the performance metrics are used to evaluate the machine learning techniques will be illustrated. In chapter four, some experiments and the results of each technique will be discussed, also, a comparison between techniques showing the best one will be provided. Chapter five presents the conclusion and future work.

Chapter 2

Background

2 Introduction

Conforming to the global health institutions, obesity is known as a disease that mainly affects a person's health, obesity is one of the main risk factors for type-2 diabetes and cardiovascular disease, particularly, in children. Obesity affects the respiratory system and increases the incidence of asthma [16].In the United States, more than 30% of children and teenagers suffer from overweight or obesity [17]. Also, more than 300,000 people suffer from obesity and asthma were related. Studies have indicated that increasing a person's obesity increases the risk of asthma [18]. Studies on obesity are rare in Palestine, that it is a very important topic which should be pointed out, due to the significant increase in the percentage of obese children from 3 to 6% within five years, while in the rest of the world the percentage has increased from 1 to 7% in 41 years [19]. In Palestine, according to the ministry of health (2018), type-2 diabetes, hypertension and metabolic syndrome associated with increased obesity account for 50% of deaths [20]. Many techniques have been developed to predict and classify obesity disease based on the patient's records. In this work, clustering techniques such as k-means and fuzzy c- means with the aim of grouping similar obesity data are used and after that different AI models and a hybrid Neuro-fuzzy model are used to classify and predict obesity.

2.1 Related Works

In [23], the authors used ANFIS to test the performance of this technique to estimate the BMI. They used five factors to predict BMI (waist circumference, fasting glucose, high-density lipoprotein, triglycerides, and homeostatic model assessment-Insulin Resistance). Also, the researchers have concluded that obesity can be avoided if changeable risk factors and modified lifestyle are controlled. So, this method was effective to predict BMI.

In the research [24], the authors made a comparison of clustering methods for obesity classification. These methods were used in that study decision tree, discriminant analysis, Fuzzy Rule-based System (FRBS), Kth-Nearest Neighbor (KNN), Neural Networks (NN), and Support Vector Machine (SVM). The researchers aimed to find out some other new method to classify obesity with no need to measure other dimensions of the body, which sometimes difficult to measure. Their result showed that FRBS is more accurate than other used techniques.

According to paper [25], obesity affects more than 500 million adults worldwide and by 2030, more than 1 billion adults are expected to suffer from obesity. It is exponentially grown in obesity and that increases the mortality levels, in this research, a review of the application of Artificial Intelligence (AI) on obesity management is presented. Different datasets selected from well-known databases are used to prove the effect of AI in solving this problem. The author's study includes; a decision support system for obesity surgery, Artificial Neural Network that relates obesity to heart disease; Artificial Neural Network to predict obesity in the future and a Neuro-Fuzzy Model to refine body mass index results. It was concluded that all AI research models have a tendency to more accurate results.

In [26], the researchers studied the accuracy of an adaptive Neuro-fuzzy computing system in prediction of the anti-obesity using the medicinal plant. The researchers compare the results of the Neuro-fuzzy model with the results of Support Vector Regression (SVR) depending on the Root-Mean-Square Error (RMSE) and the coefficient of determination (R2). The result of the Neuro-fuzzy model outperforms the result of SVR.

In [27], the researchers aimed to predict body fat percentage from some measurable data such as (age, gender, weight, height, waist Circumference, and different laboratory results). They applied some soft computing methods such as support vector machine, linear regression, and feed-forward neural networks to their data, they concluded that support vector machine is better than feed-forward neural networks and linear regression.

In [28], the authors took the obesity problem among children aged 12 years old. They used four classification techniques which are: Bayesian Network, Decision Tree, Neural Networks, and Support Vector Machine (SVM). They concluded that the Decision Tree and Support Vector Machine is the best technique for this data set and problem.

In [29], the researchers used logistic regression and neural networks to classify obesity. Their results showed that both techniques have no difference in classification. The researchers used lifestyle information and demographic variables of (age, marital status, level of education, duration of physical activity per week, and history of smoking) with their frequencies, which were obtained from each individual). Besides that, they have used anthropometric measures of (weight, height, Mid-Upper Arm Circumference (MUAC), Waist Circumference (WC), Hip Circumference (HC), triceps skinfold, and abdomen thicknesses measured).

In [30], the researchers used a fuzzy inference system due to uncertainty and vagueness data which form obesity and can be measured such as lifestyle, eating habits, and level of activity. They used many inputs and one output where the inputs are (sex, age, Gene GAD2, high-calorie food, fast food infiltration in our culture, Lack of physical activity, sociology of food, energy imbalance, larger portion sizes, diseases and drugs, negative emotions and early menarche). It is concluded that their system can be used as decision support to help experienced physicians.

In [31], the authors used multiple linear regression, Artificial Neural Networks (ANN), and Adaptive Neuro-Fuzzy Inference System (ANFIS). Their results showed that ANFIS is more beneficial than the other models in predicting BMI. They have interpreted the results that ANFIS is an appropriate model for uncertain fuzzy variables as in this case.

In [32], the authors said that obesity is the main major that forms a background for many diseases, also they considered obesity as a disease that shortens the human life and negatively affects the quality of an individual's life, in general, they aimed to make diagnoses with AI depending on medical data. It is concluded that Neuro-Fuzzy Systems results are a reliable classification system for quickly diagnosing.

The authors in [33] used two types of data mining algorithms called logistic regression and a multilayer perception (MLP) neural network to classify and follow-up obesity. The two models' results showed that the classifying performance NNs is better than logistic regression.

In [34], the authors used Recurrent Neural Networks (RNNs) depending on the time-aware architecture to select important features and detect irregular observation from clinical patient records for obesity with the goal of obesity status improvement prediction. The method can detect 77-86% accuracy of irregularities.

In paper [35], the authors made a review about childhood obesity prevention and treatment using computerized decision support and machine learning, they confirmed that obesity is linked to several chronic diseases such as cardiovascular disease, diabetes, and cancer. From their review, it is found that researches in the CDS and ML field for childhood obesity is rare, and they highly recommended using these methods in childhood obesity diagnoses and treatment. Because these methods can extract useful knowledge from the provided data and support the Specialists in their decisions, they concluded that ML is very useful and helpful for predicting childhood obesity.

In paper [36], the authors found that obesity in childhood and teenagers is rising dangerously in many countries, this poses a threat, as it is associated with several potentially fatal diseases. They recommended using ML models tools in this field to its strength in prediction and represent the complex nonlinear relationships between the parameters. In this research they reviewed and compared the ML models and deep learning methods with the traditional statistical methods, their comparison showed that ML models give better results in prediction.

As It has been reviewed, obesity is considered a very dangerous disease that leads to fatal diseases, and the researches in this field are rare. Therefore, in this research, different machine-learning methods are used to predict and classify obesity, a comparison is made between Logistic Regression, KNN, SVM, Decision Tree, Neural Network and Neuro-Fuzzy to come up with a summary of which method is better, in this case, to help the specialists to diagnose and treat the obesity early. Also, clustering methods (K-means and Fuzzy C means) are used to give more accurate results.

Chapter 3
The Proposed Method

3 Proposed Method

This chapter clarifies the proposed method which aims to perform a comparison of classification algorithms to determine the most suitable classifier, improve prediction accuracy in cases of obesity and predict if someone tends to be obese. In the beginning, it starts by selecting the dataset and describing the preprocessing steps. After that, the deployed models will be described: Regression, Decision Trees, SVM, and Neural Networks. The main method that will be investigated is clustering techniques with the aim of grouping similar obesity data. Afterward, depending on this clustering result, Neuro-Fuzzy (Neural networks and Fuzzy Logic) Model will be used to classify and predict several parameters that will speed up and increase the accuracy of the diagnostic process. Figure 3-1 shows the block diagram of the forecasting model in general.

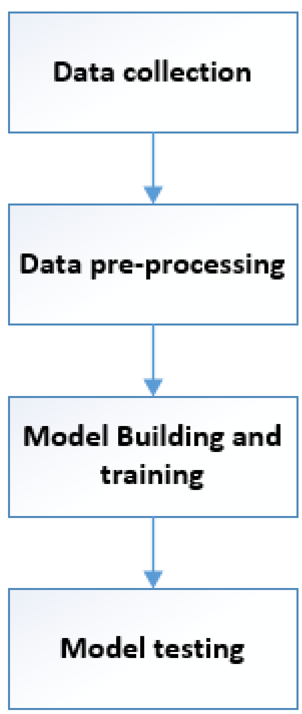

Figure 3-1: The block diagram of the forecasting model in general

3.1 Dataset Description

The Dataset is Palestinian locally collected to apply it in different models. The dataset contains 970 random records for people from four cities, each record has 12 features and depending on these features, the results were deduced. Figure 3.1-1 shows some of the data which is used and Table 1 represents the valid ranges for each feature:

Table 1 : Valid ranges for each feature

Feature Name	Minimum Value	Maximum Value
BMI	1	4
Age	8	12
WHtR	0.34	0.69
WHR	0.56	1.48
Fat mass%	3	57.9
Fat mass	0.6	52.8
Lean mass	13.9	54.3
Weight	19.4	94.2
Height	113	166
Hip Circumference	41	125
Waist Circumference	43	111
City	1	4

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	BMI	age	WHtR	WHR	Fatmass%	Fatmass	Lean mass	Weight	Height	Hip circur	Waistcircu	City	target
2	2	12	0.45	0.87	16.5	7.5	37.9	45.4	152	78	68	2	0
3	2	11	0.36	0.81	17	4.7	23	27.7	133	59	48	3	0
4	1	11	0.42	0.82	7.7	2.2	26.1	28.3	142	72	59	4	0
5	4	10	0.46	0.78	37.5	17.8	29.6	47.4	135	79	62	2	1
6	2	11	0.43	0.82	18.7	5.4	23.6	29	125	66	54	1	0
7	2	10	0.47	0.95	12.3	4.3	30.7	35	150	74	70	3	0
8	2	11	0.45	0.86	11.9	3.6	26.9	30.5	139	72	62	3	0
9	2	12	0.42	0.83	13.9	3.9	25.1	29	137	70	58	3	0
10	2	9	0.5	0.91	15	3.4	19.6	23	121	67	61	2	0
11	2	9	0.67	1.12	21.1	6.8	25.6	32.4	138	83	93	2	0
12	2	10	0.47	0.79	23.4	8.6	28.1	36.7	142	85	67	2	0
13	2	10	0.41	0.75	15.8	5.8	30.8	56.6	145	80	60	2	0
14	3	9	0.42	0.74	20.5	8.4	32.8	41.2	144	82	61	1	1
15	2	10	0.38	0.84	16.7	5.6	28	33.6	137	62	52	3	0
16	4	9	0.64	1.12	31.5	13.5	29.2	42.7	135	78	87	2	1
17	2	11	0.44	0.82	21.8	8.7	31.4	40.1	145	78	64	4	0
18	2	11	0.46	0.92	21.1	8.4	31.3	39.7	150	75	69	3	0
19	2	10	0.51	0.85	26.1	11.2	31.8	43	148	88	75	2	0
20	2	10	0.43	0.77	19.7	6.8	27.9	34.7	142	79	61	2	0
21	2	10	0.39	0.74	17	5.2	25.4	30.6	141	74	55	1	0
22	2	9	0.41	0.76	22.6	6.1	21	27.1	124	67	51	1	0
23	3	9	0.51	0.83	31.9	12.2	26.2	38.4	132	81	67	2	1
24	2	11	0.48	0.87	21.1	7	26	33	140	77	67	1	0
25	3	10	0.54	0.91	31.5	15.4	33.6	49	147	87	79	4	1
26	2	11	0.48	0.88	13.1	5	32.8	37.8	145	80	70	3	0
27	2	10	0.4	0.77	12.8	4.1	27.8	31.9	140	73	56	3	0
28	2	11	0.43	0.88	20.6	8.1	31.4	39.5	150	73	64	3	0
29	3	11	0.44	0.82	23.6	9.2	29.7	38.9	135	73	60	1	1

Figure 3.1-1 : Sample from the Dataset

The following are the attributes used to predict the results:

1. Body max index (BMI): Is the body fat index depending on the weight divided by the square of the height on Childs. BMI is the main major factor that affects the results [21]. BMI in the collected data is categorized into four categories as following:

   - 1= Underweight
   - 2= Normal weight
   - 3= Overweight
   - 4= Obesity

   By merging all of them in two categories as 1+2 = not overweight, 3+4= overweight a binary output 0 means not overweight and 1 means overweight are obtained. Figure 3.1-2 shows how the data is distributed depending on BMI:

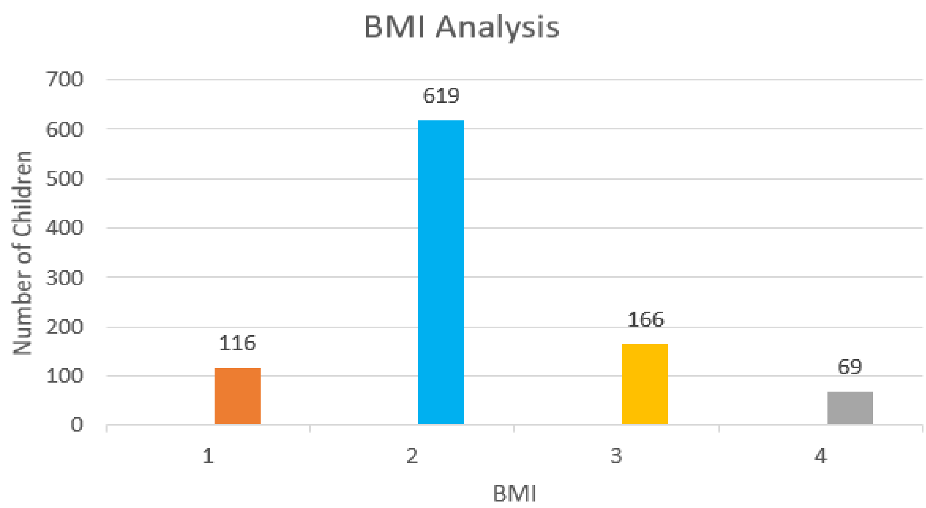

Figure 3.1-2 : The distributed data depending on BMI.

2. Age: The age of the child is the second predictor that affects the results. The records showed that children with age 8 and 12 years are more likely to be fat more than other ages as shown in Figure 3.1-3.

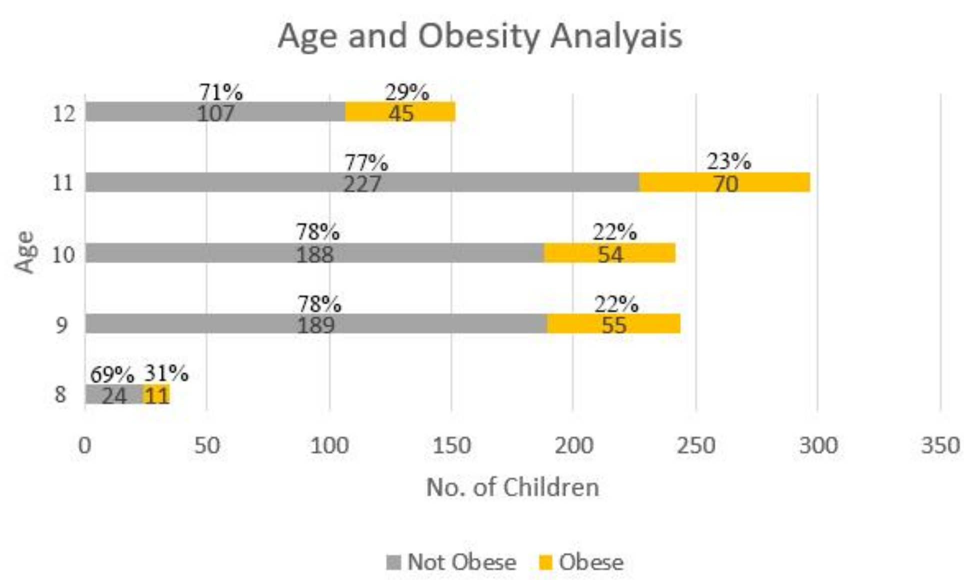

Figure 3.1-3 : Age and Obesity Analysis

3. Waist-to-Height Ratio (WHtR): was calculated as Waist Circumference (CM) divided by height (CM), the higher WHtR, the greater the risk of obesity most of the time. Figure 3.1-4 shows the distribution of the population according to WHtR.

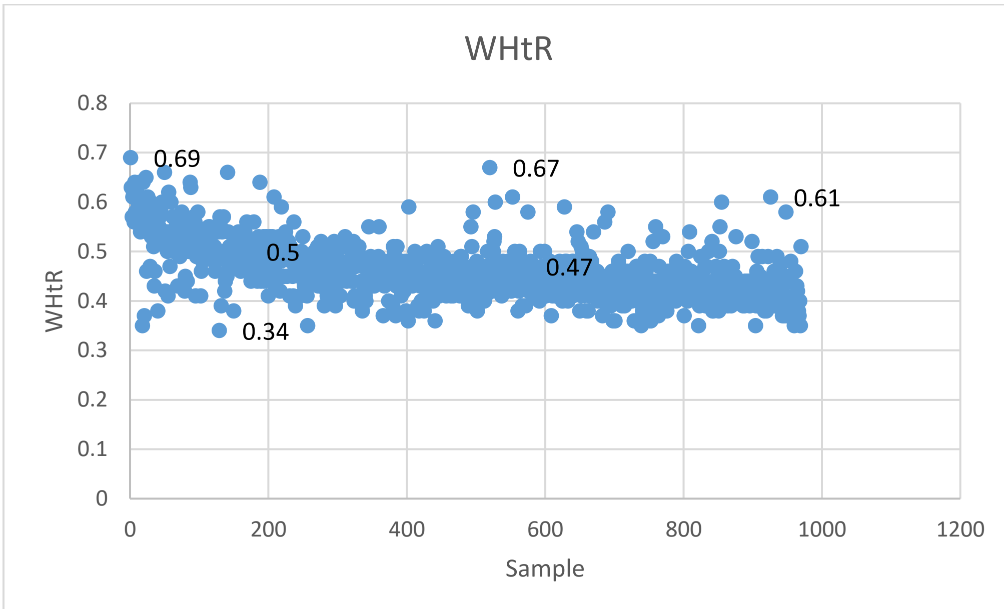

Figure 3.1-4 : The distributed population depending on WHtR

4. Waist-to-Hip Ratio (WHR): Was calculated as Waist Circumference (CM) divided by Hip Circumference (CM). Figure 3.1-5 shows the distribution of data depending on WHR.

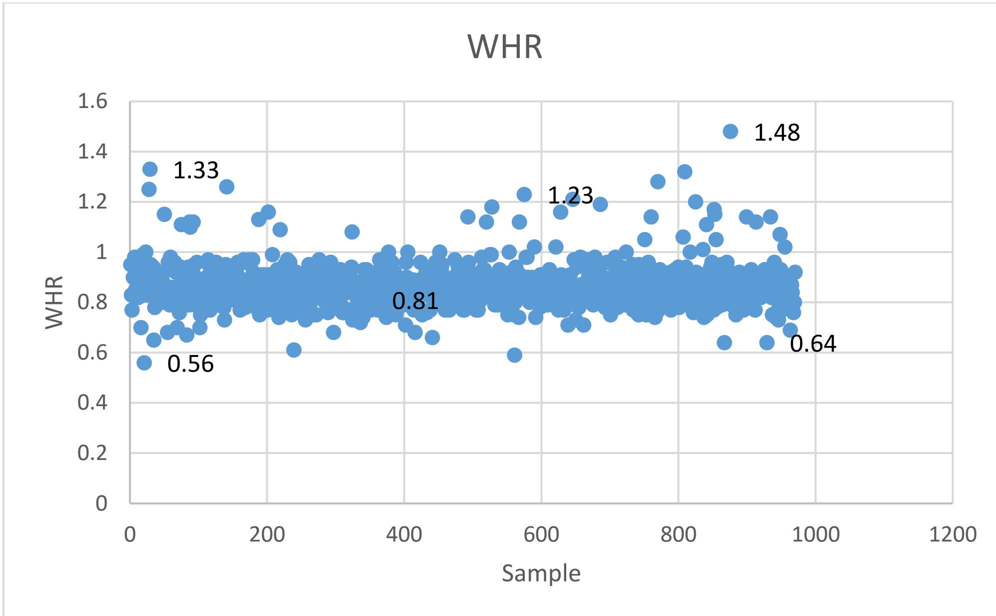

Figure 3.1-5 : The distributed data depending on WHR

5. Fat mass percentage: This is a measurement of body composition aiming to know how much the weight of your body is fat, also the total mass of fat divided by the total body mass which is multiplied by 100, the higher fat mass%, the higher risk of obesity (to get more accurate results, many other parameters should be taken into consideration) as shown in following figure where the y-axis represents values 0 not obese and 1 is obese:

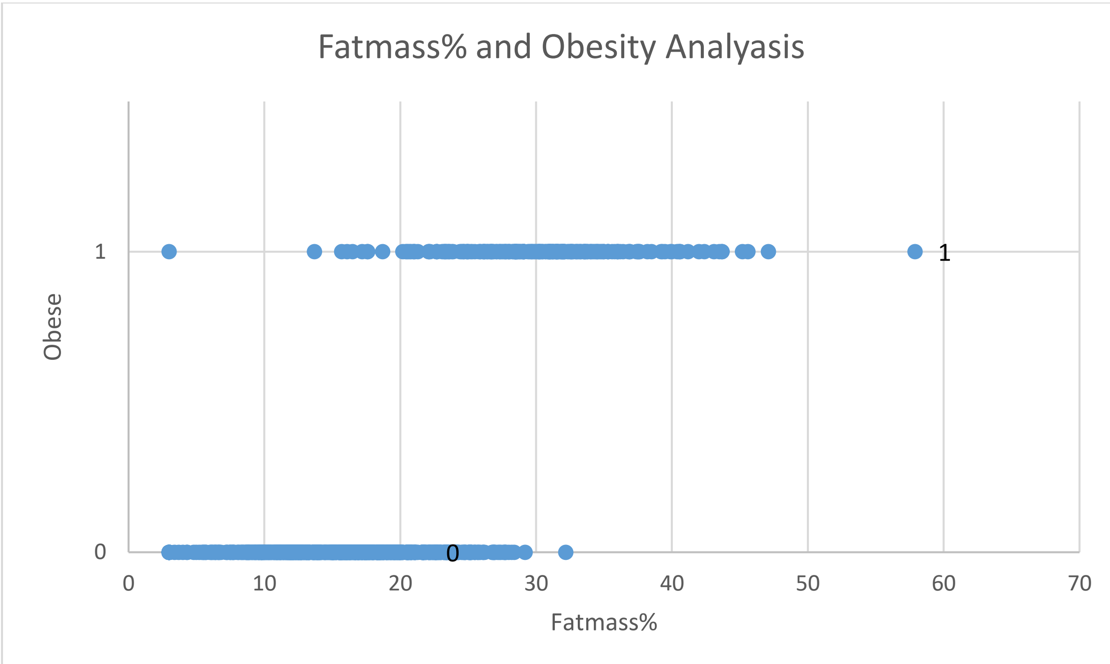

Figure 3.1-6 : Fat mass% and Obesity Analysis

6. Fat mass (kg): That portion of the human body is composed strictly of fat (Total body fat). It can be measured with dual-energy absorptiometry or bioelectrical impedance techniques.
7. Lean mass (kg): Which means that the mass of all organs except body fat, in another word lean mass is the difference between total body weight and body fat weight.

Lean mass = weight – fat mass

8. Weight (kg): Weight is one of the main reasons to have obesity.
9. Height (cm): Height is related to a person's weight (as the weight of a person may be appropriate for his height), while the same weight is not suitable for another person with a different height who might be classified with obesity. For example, in our data, a child with a weight of 53.9KG and a height of 165cm is

categorized not obese because he is considered tall; another child with the same weight and height of 147cm is considered as an obese child.

10. Hip Circumference (CM): The distance around the human body at the level of the maximum posterior extension of the buttocks.
11. Waist Circumference (CM): This feature sometimes was used as an indicator tool for abdominal obesity [22].
12. City: The data was calculated from four different cities, for each city, numerical representation was used as follows:
    - 1= Jenin.
    - 2= Nablus.
    - 3= Hebron.
    - 4= Tulkarem.

3.2 Data Collection

The performed data were collected from four cities in Palestine. Figure 3.2-1shows how the population is distributed according to the city. In the population, ages range between 8 to 12 years old.

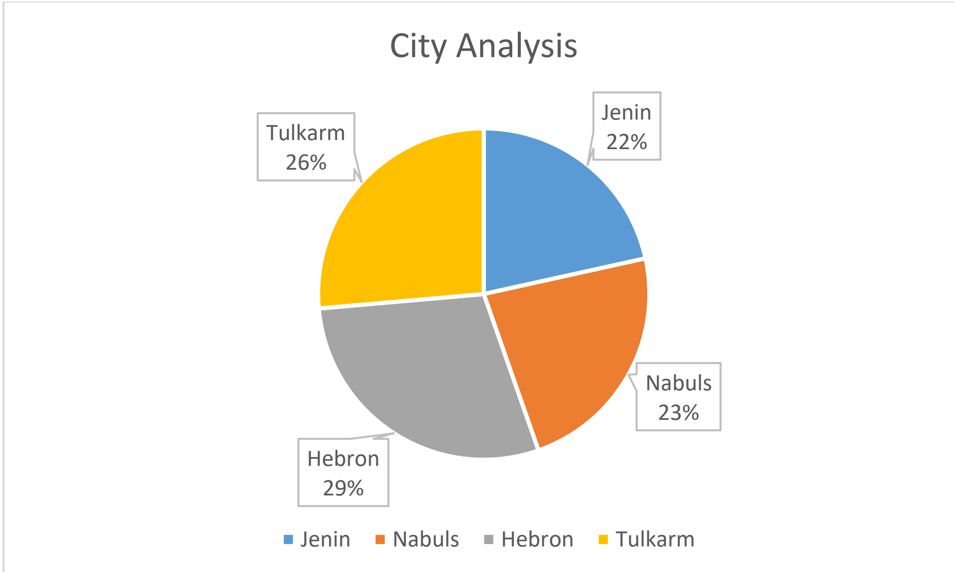

Figure 3.2-1: Distribution of samples based on the city

In this research, children were categorized according to BMI. When the BMI values equal 1 and 2, it indicates that the child is not obese, in this case, the output equals 0. On the other hand, when the BMI values equal 3 and 4, it indicates that the child suffers from obesity, in this case, the output equals 1. Therefore, BMI cannot be used as input in the experiments that the output directly associated with it.

3.3 Data Preprocessing Phase

3.3.1 Clustering

In this section, two clustering methods were applied and a comparison between them was given. The first method is fuzzy is c means clustering with five classes; the second method is k-means clustering, also using k=five.

3.3.2 Clustering Methods

Clustering methods are used to organize similar data into groups or clusters. In this thesis, the k-means clustering and Fuzzy C-means clustering were used. Here in our thesis clustering methods will be used to group the homogeneous obesity data for persons, regions, or communities depending on the similarities in input features.

- **K-Means Clustering**

Simply, it is an algorithm to group that the data based on some features into K number of groups or clusters .Data in different clusters differs from each other and in the same cluster is similar to each other (Figure 3.3-1) [50]. Each cluster has a centroid or cluster representative [51]. So, to classify data with K clusters, the following steps to be used:

1. Initialization: Initializing cluster centers.
2. Classification: In this step, the start is classifying data by calculating the distance for each data point in the input data from each cluster center usually Euclidean distance. The data point that has the minimum distance from the cluster center is assigned to that cluster.
3. Centroid Recalculation: Update the cluster centers based on the data points assigned to that cluster using the following equation :

$$k_i = \left(1/c_i\right) \sum_{j=1}^{c_i} x_i \qquad (3.1)$$

Where $k_i$is the ith center, c_i represents the number of data points in the cluster, k_i, x is the data point and i is the number of clusters.

4. Convergence Condition: When to stop?
   - When number of iterations is reached.
   - When the data points do not change between clusters.
   - When the threshold value is achieved.

5. If not one of the above stopping conditions are satisfied, it is advised go to step two and repeat until at least one condition is satisfied.

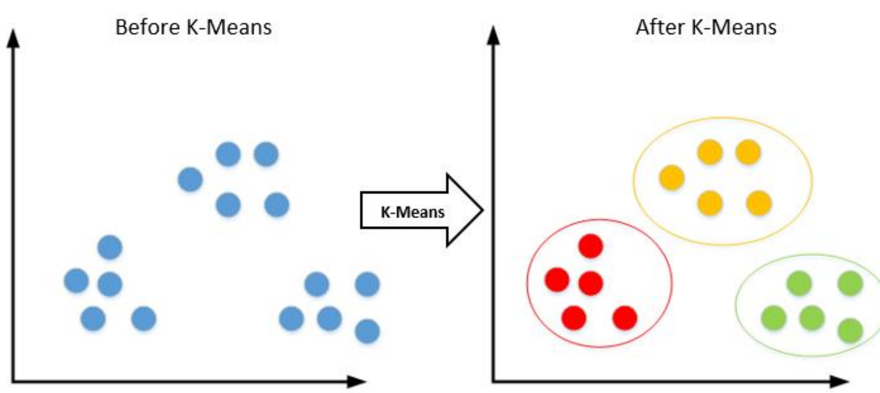

Figure 3.3-1: K-Means example

- **Fuzzy C-Means Clustering**

FCM is an extension of KM, Bezdek (1981) which developed the algorithm based on Dunn's study 1973 [52]. FCM is a soft algorithm clustering in which data point does not only belong to one cluster, it can have several more than one cluster but varying degrees of membership. The closer the points are to the cluster center, the more its membership is close to that cluster center [53]. So, data points located on boundaries but not forced to fully belong to only one cluster. Therefore, the advantage of FCM that an object may belong partially to a different cluster rather than belong to one cluster [54]. To classify some data, these steps should be followed:

1. Initialization: Selecting the number of clusters.
2. Randomly selecting cluster centers.
3. Computing the fuzzy membership using the following equation:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} (d_{ij} d_{kj})^{\frac{2}{m-1}}} \tag{3.2}$$

Where i is the cluster, j: is the object in the dataset, k,c: centers of the clusters, m: fuzziness index that any r, d: the Euclidean distance between the object and the cluster, μ: the degree of membership for each data point in each cluster.

4. Compute the fuzzy center's c_i using the following equation:

$$c_i = \frac{\sum_{j=1}^{n} \mu_{ij} \cdot x_j}{\sum_{j=1}^{n} \mu_{ij}{}^{m}} \qquad (3.3)$$

Where c_i is the d-dimension center of the cluster and x_j is the ith d-dimensional measured data.

5. If following equation is satisfied then stop other wise go to step 2

$$\| \mu^{i+1} - \mu^{i} \| < \varepsilon \qquad (3.4)$$

Where ε is the termination standard between 0 and 1 and i the iteration step.

The following experiment using k-means clustering and fuzzy c mean clustering applied on one of the features (fatmass%) to explore the dataset behavior. Table 2 describes how much each class contain samples for each model using fat mass% as input for each method:

Table 2 : FCM and k-means Results

Class	FCM Boundaries	Number of samples using FCM	K-means Boundaries	Number of samples using K-means
1	[3-12.9]	220	[3-10.8]	125
2	[13-19.2]	312	[10.9- 16.4]	258
3	[19.3-25.9]	231	[16.5-22.4]	269
4	[26.1-33.8]	153	[22.5-30.3]	200
5	[34-57.9]	54	[30.4-57.9]	118

Figures 3.3-2 and 3.3-3 show the fat mass% values and each value in which class belongs to:

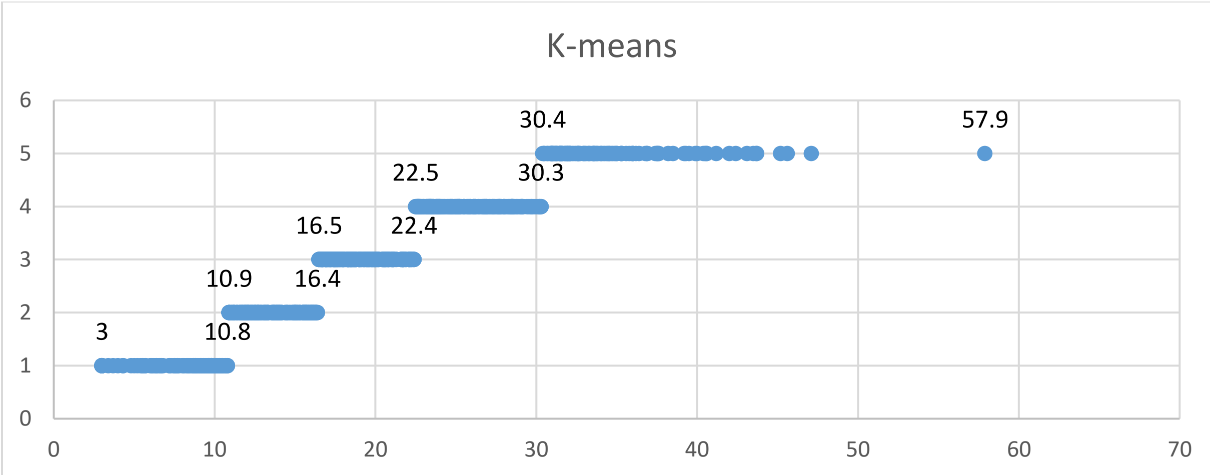

Figure 3.3-2 : K-means clustering on Fat mass%

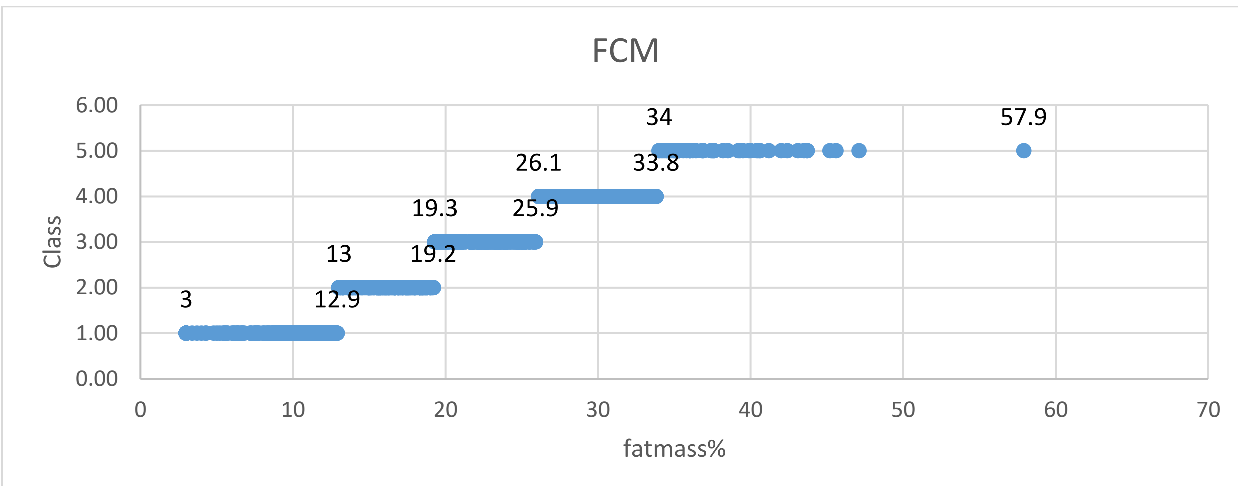

Figure 3.3-3 : FCM clustering on Fat mass%

3.4 Building Models Phase

This section shows the models, which have been applied to the study data. In the first stage, different data mining tools were applied like Logistic Regression, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Neural Network (NN). Then in the second stage, clustering algorithms like K-means or Fuzzy C-means clustering were combined with a Neuro-Fuzzy model that depends on the integration of the Neural Network with Fuzzy Logic models and after that, a comparison between this model and the other models (in the first stage) were made. MATLAB will be used to apply these models. Figure 3.4-1 shows the general phases for predicting and classify obesity for all models.

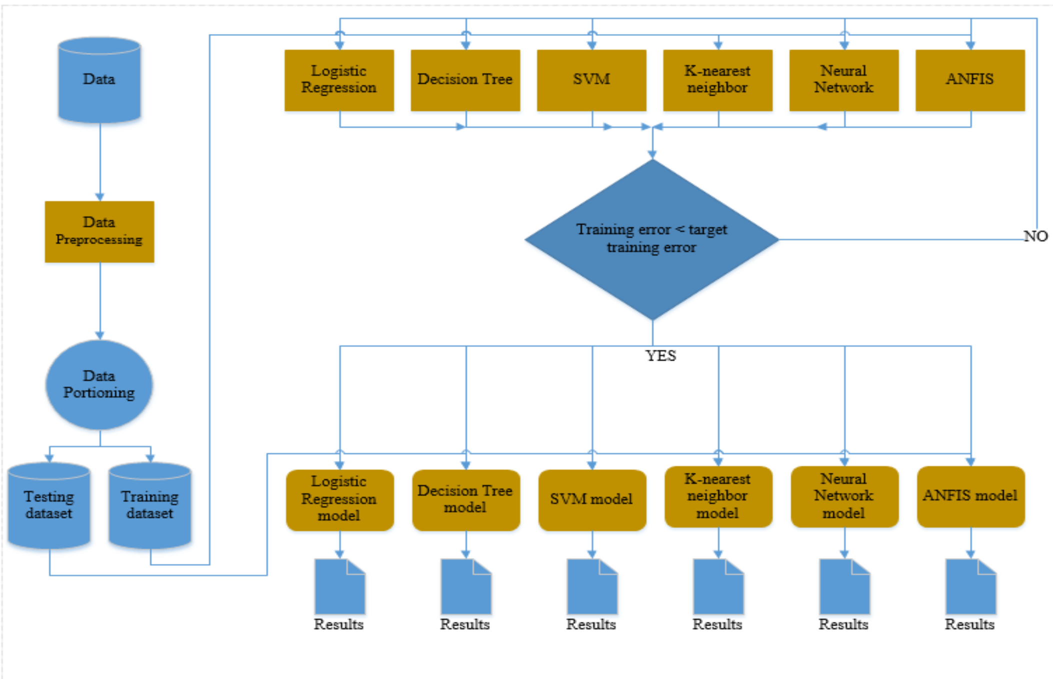

Figure 3.4-1 : General phases for predicting and classify obesity for all models

3.4.1 Logistic Regression

Regression techniques are considered to be versatile in their medical applications due to their abilities to predict outcomes, control for confounding variable effects, and measure associations [37]. For example, if we want to see how an outcome and many independent variables are associate with each other, or in another way if we want to predict an outcome from independent variables.

Many types of Regression depend on what are the objectives of the research and the variable format. In this thesis, Logistic Regression is used, which is used when the outcomes are binary represented [38], this type is different from other types in that the predictors can be any mix of continuous, discrete, and dichotomous variables [39]. The following equation is the Logistic Regression Equation:

$$\text{Probability of Outcome } (\hat{Y}) = \frac{e^{\beta 0 + \beta 1X1 + \beta 2X2 + \dots \beta iXi}}{1 + e^{\beta 0 + \beta 1X1 + \beta 2X2 + \dots \beta iXi}} \tag{3.5}$$

where the outcome $\hat{Y}$ represents the approximated probability of being in one category versus in the other, binary outcome, β_0 it's a constant value where its value considered as the point where the Regression line reaches the Y-axis and $\beta_1X_1 + \beta_2X_2 + \dots \beta_iX_i$ is the value of each independent value Xi weighted by its respective beta coefficient (β), where beta is the slope of the line or how much the resulting increase for each increase of one on the independent variable. More beta value means more strongly this independent variable affects the outcome [37].$e^{\beta 0 + \beta 1X1 + \beta 2X2 + \dots \beta iXi}$ expressed in the logit scale. The reason for using the logit scale is that Logistic Regression outcome is binary must fall in 0 and 1. The memory requirement for storing the classifier and the time complexity of the training phase are linearly proportional to the number of classes in the data set [58].

3.4.2 Decision Tree

The decision tree is a technique mostly used for data mining to create a classification system based on multiple independent variables or used to develop some prediction algorithms for an outcome variable. Data mining provides effective performance in data mining due to its accuracy and inexpensive to compute [40]. Decision tree time complexity depends on the levels of the tree it has a time complexity of $O(m \cdot n^2)$ of the standard decision-tree learning algorithm where m is the size of the training data and n is the number of attributes [59]. In general, this method is used to classify some populations into branches or levels by building a tree consists of nodes and branches showing in Figure 3.4-2. There are three types for a node, which are:

1-) The A root node (called a decision node): Represents a choice that will result in the subdivision of all records into two or more mutually exclusive subsets.

2-) Internal nodes (called chance nodes): Represents one of the possible choices that are available in the tree structure where the top edge of the node is connected to its parent node and the bottom edge is connected to its child or leaf nodes.

3-) Leaf nodes (called end nodes): Represents the combination of decisions or events in the final result.

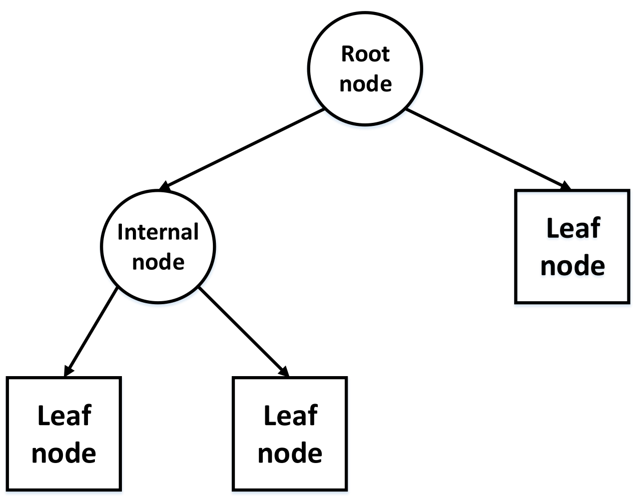

Figure 3.4-2 : Decision Tree Structure

The branches represent the probability or results which go from internal nodes and root nodes only. Each path from the root node over internal nodes until it ends on some leaf node represents a classification Decision Tree. In addition, these paths can be represented as if-then rules [40]. There are several types of Decision Trees, but each type uses a different mathematical model and selects the splitting attribute in extracting the Decision Tree rules. The most common types are Gini Index, Information Gain and Gain Ratio. [41]. The process of building a decision tree in general passes through several stages, namely:

1- Splitting: It is the process of splitting the node into one or purer sub-nodes. At first, when building some model, it is important to select the most important inputs (which most affects the result), and then it splits the records into two or more categories.

2- Stopping: The main goal of this step is to prevent the model from being very complex, but there are some rules to stop which includes minimizing: the number of records in the leaf, the depth or number of levels (number of steps from the root node to the leaf node) and the number of records before the splitting.

3- Pruning: It is the opposite of splitting, sometimes stopping ways don't work well, and another way to build a decision tree is to build a large one at first and then prune it to the best size by removing nodes that are useless or don't provide any additional information [40].

3.4.3 Support Vector Machine (SVM)

Support vector machines are used to recognize patterns. in this technique, the algorithm illustrates the data into high dimensional space and the algorithm separates the data by a hyper-plane in that space to classify the data and predict a newer one [42].In other words, the SVM goal is to make boundaries between two classes called margin as shown in Figure 3.4-3 [42]. These boundaries are created to be as far as from the closest data point from each class that the distance between boundary and closest data point in each class is the maximum, and these the closest points are called the support vectors [42]. SVM has time complexity of $O(n^3)$ where n is the numbers of the data [60].

Given training dataset:

$$(x_1, y_1), ..., (x_n, y_n), x_i \in R^d \text{ and } y_i \in (-1, +1) \quad (3.6)$$

Where x_i is the representation of the feature vector and y_i is the class label (1 or 0) of some training compound *i*.

So, to express the optimal hyper-plane, they defined the following equation:

$$W . x. T + b = 0 \quad (3.7)$$

Where *w* is the weight vector, *x* is the feature vector, T is the number of data samples divided on support vectors and *b* is the bias.

w and *b* should satisfy the differences for the elements of the training set:

$$w. xi. T + b \geq +1 \text{ if } yi = 1 \quad (3.8)$$

$$w\ xi\ T + b \leq -1 \text{ if } yi = -1 \quad (3.9)$$

The goal of training an SVM model is to calculate the appropriate *w* and *b* to separate the data with optimal hyper-plane and maximizes the margin $1/ \| w \|^2$ [43].

The Vectors For x_i that $|yi|\ (wx_i^T + b) = 1$ (will be called support vector).

In the beginning, SVM was the main goal for constructing a linear classifier. To model more than one dimensional (non-linear models), the kernel method is used and kernel function works in a way to make linear problems by adding more dimensional to the raw data [42].

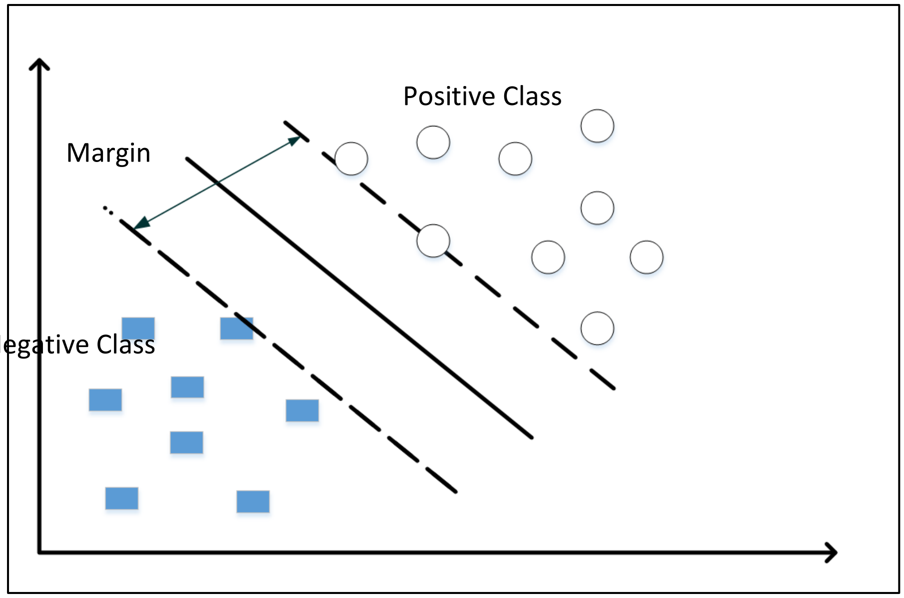

Figure 3.4-3 : How SVM separate data and showing the margin distance

3.4.4 K-Nearest Neighbor (KNN)

K-Nearest Neighbor is one of the simplest classification methods and its results are good. It is learning samples from the dataset of training data. When new data is provided, it is compared to a similar set and classified to its class and simply it assumes similar data exist in close [12]. The new data is classified based on the class that is closest to it, often there is more than one neighbor that should be taken into consideration.

The number of neighbors, which is taken into account in determining the class is defined by a variable known as “K”, K-NN called “lazy learner” that is just learned from a given data set and tested in a new data, then generalization was performed to determine the class of new data based on similarities or the distance to the data set, it is also called “instance-based learner” [12].

These algorithms work as follow: Choosing the value of K and then for the given test data calculating the distance (mostly Euclidean distance) in between all training data, then sorting the distances and choosing the closest K. Figure 3.4-4 shows the flowchart of K Nearest Neighbor classifier procedure [44].

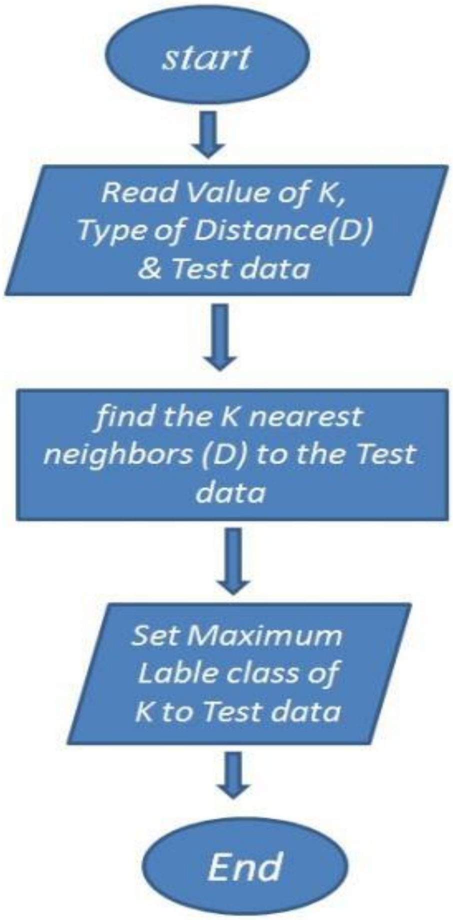

Figure 3.4-4 : The flowchart of K- Nearest Neighbor classifier procedure

Algorithm 1: The general procedure for k-Nearest Neighbor

If **x is being classified,**
A: training data, **B**: class labels
For i=1 **to** number of training data **do**
    Compute the distance (A_i, x) where d denotes the Euclidean distance
End for
Compute set I containing indices for the k smallest distances (A_i, x).
Return majority label for {Bi where $i \in I$}.

3.4.5 Artificial Neural Networks (NNs)

The Neural Networks is a compositional system that mimics the human brain or human neurological system which is represented in mathematics. This model simulates the learning capability of the human brain. The Neural Networks is one of the widely powerful methods that is used for classification due to its effective methods when the data set in terms of increasing the number of features and how to find the function that is appropriate to classify these data accurately. A Neural Networks consists of layers; each one contains a series of neurons that are organized and connected to other neurons in other layers with some weight. The first layer also called the input layer, after this layer, there are one or more hidden layers and finally, the output layer, the general structure of NNs is shown in Figure 3.4-5.

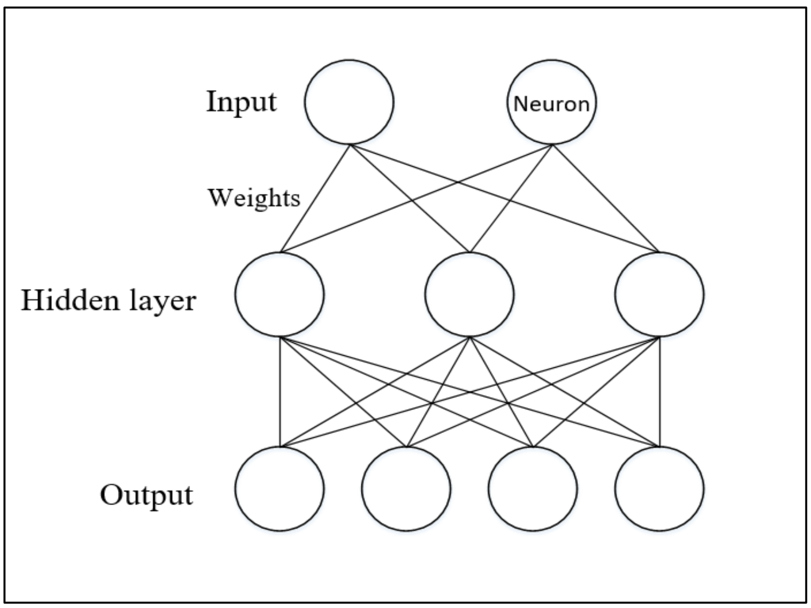

Figure 3.4-5 : The general structure of NNs

In Neural Networks, the data is divided into two sets training data and testing data. The first phase is the learning or training phase. In this phase, the weights between neurons that are associated with connections are modified until the model performs well. In the testing phase, the model is used to make a prediction using new data (testing data) [45]. In a neural network, an activation function (e.g. Sigmoid, Tanh, and Softmax) is used to introduce nonlinearity; it takes the output from the previous neuron and converts it in a way that can be an input to the next neuron. The neural network consists from three layers:

- Input layer: X_i where i=1,2, …, n. and n represents the number of the inputs.
- Hidden layer: this layer consist from j neurons, each neuron is connected to the input layer by the weight w_{ij}, where i is the input node and j is the hidden layer.
- Output layer: this layer consist of nodes that gives the output y.

The weights for j^{th} node in the hidden layer is given by:

$$y = \sum W_{ij} X_i + \theta_j \tag{3.10}$$

Where θ_j is the bias always equals to 1. To calculate the gradients efficiently in Multi layer perceptron neural networks it uses backpropagation algorithm. This algorithm requires activation function to be differentiable, usually sigmoid function is used, so the output of the j^{th} node is:

$$Y_j = X_k = \frac{1}{1+e^{-y}} \tag{3.11}$$

Here the output y_k is used as input (x_k) for the next layer.

The error signal δ_k of the output is calculated using the following equation:

$$\delta_k = \Delta_k \, Y_k \, (1 - Y_k) \tag{3.12}$$

Where Δ_k is the difference between the actual value of the node k and its target value.

To find the change in weight in some two nodes node j and node k we can use the following formula:

$$\Delta w_{jk} = l\,\delta_k\, X_k \tag{3.13}$$

Where l is the learning rate. Using the following equation we can update the weight w_{jk} between the two nodes:

$$w_{jk} = w_{jk} + \Delta w_{jk} \tag{3.14}$$

The backpropagation algorithm repeats until the error on the output node is minimized. To calculate the error E:

$$E = \frac{1}{2}\sum\sum(t_k - Y_k)^2 \tag{3.15}$$

There are many methods to adjust the network parameters. In this thesis, the back-propagation method (Scaled Conjugate Gradient Descent algorithm “SCGD”) was applied.

This method uses a gradient descent algorithm to train a multilayer feed-forward neural network many times. To start the training process in the Neural Network, the values of initial biases and weights are given. Then at the steepest descent direction of the function, the weights and inputs are updated [46]. The following equation is used to update weights and biases using gradient descent:

$$\mathrm{x}_{\mathrm{n}+1} = \mathrm{x}_{\mathrm{n}} - \alpha_{\mathrm{n}}\,\Delta \mathrm{g}_{\mathrm{n}} \tag{3.16}$$

Where $\mathrm{x}_{\mathrm{n}+1}$ represents the vector of the updated weights and biases, $\Delta\mathrm{g}_{\mathrm{n}}$ the current gradient of the function and α_{n} is the learning rate.

The Scaled Conjugate Gradient Descent algorithm is better than other types because it finds the minimum performance of the function and low time searching for the steepest descent to update the weights compared to other types [47]. There are many ways to estimate the step size. In this thesis, we used the Levenberg- Marquardt method is used. Figure 3.4 -6 shows the basic structure of the Artificial Neural Network-MLP.

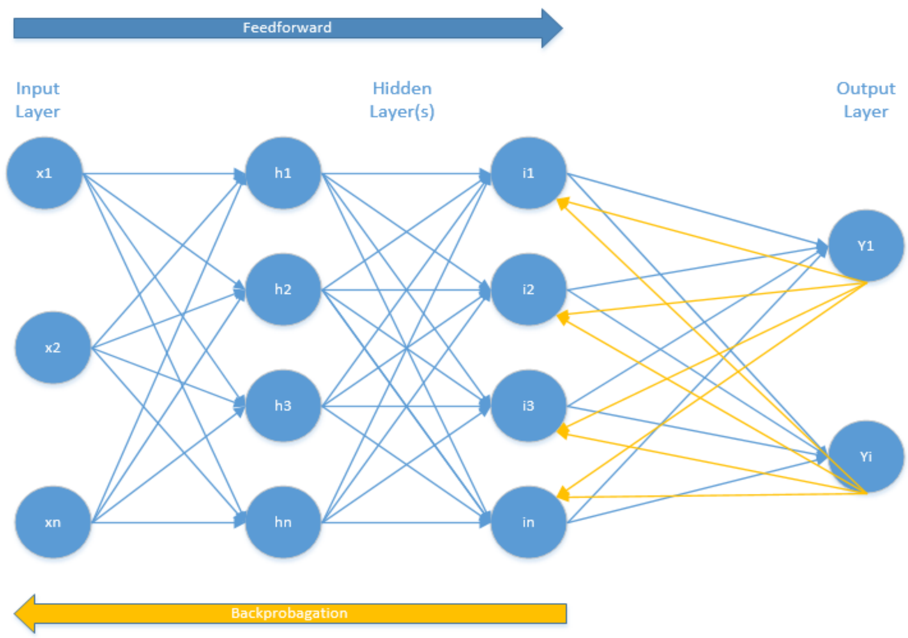

Figure 3.4-6 : The basic structure of the Artificial Neural Network-MLP

3.4.6 Adaptive Neuro Fuzzy Inference System (ANFIS)

Neuro-fuzzy systems is a hybrid technique which combines artificial Neural Network and Fuzzy Logic to make the human knowledge represented as fuzzy rules (if … then) with the best membership function to generate the best input/output pairs [48]. So, it takes the advantage of the neural network as a very good technique in classification and pattern recognition, also Fuzzy Inference System is good at explaining decisions when there are inaccurate reasons. ANFIS technique was presented for the first time by Jang in 1993[48]. ANFIS passes through two phases; the first phase feed-forward to determine the consequent parameters and the backward phase to update the parameters of fuzzy inference by an algorithm of backpropagation [49].

There are two types of ANFIS:

- Grid partitioning: This type calculates all combinations of input membership functions which is an exponential relation between the number of membership functions and the number of rules which generate a large number of rules. For instance, if there are five inputs; each one has four membership functions, the number of rules will be $4^5 = 1024$. Figure 3.4-7 shows the ANFIS-Grid partitioning structure.

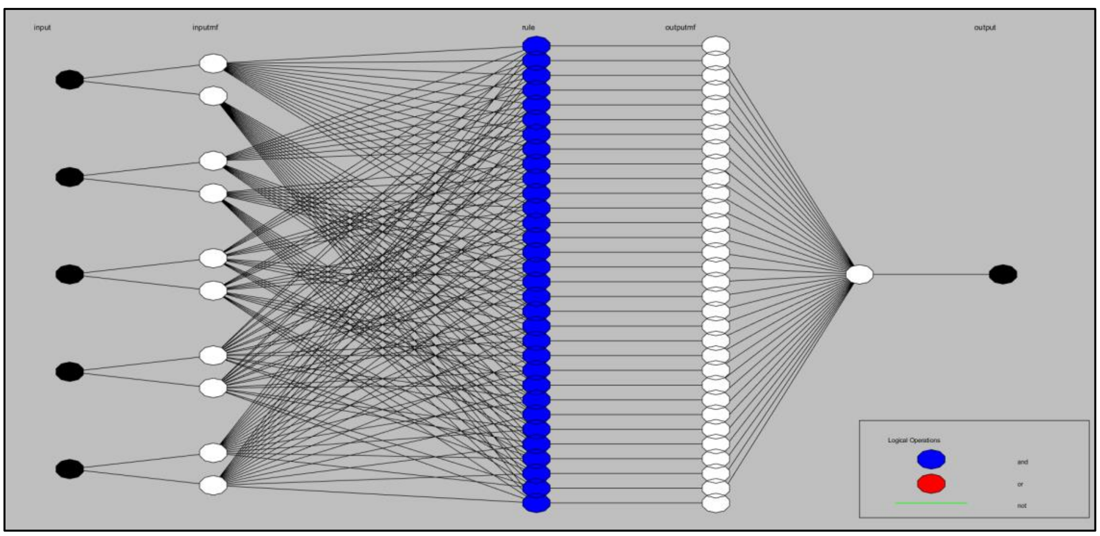

Figure 3.4-7 : ANFIS-Grid partitioning structure

Clustering: It is also called subtractive clustering. This type has a small number of rules which each rule indicates just one cluster. Figure 3.4-8 shows ANFIS-sub. Clustering structure.

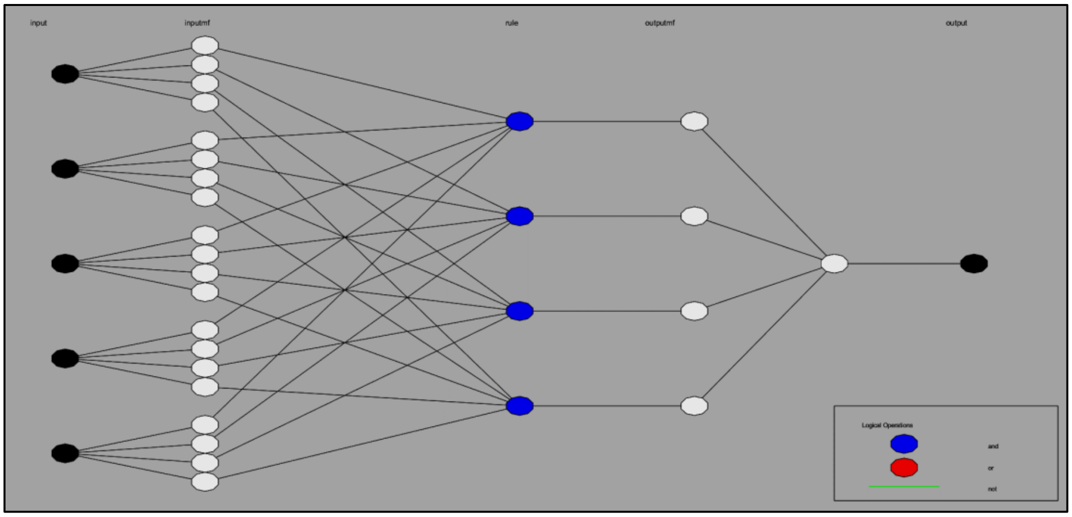

Figure 3.4-8 : ANFIS-sub. Clustering Structure

The structure of the ANFIS consists of five layers and each layer is associated with a certain phase in the model as shown in Figure3.4-9. The following are the steps and layers of ANFIS:

- First layer: the input layer, the neurons in layer represents the input variables that send the crisp value to the second layer as follow:

$$y_i = x_i \tag{3.17}$$

  Where y_i is the output of the input x_i.

- Second layer: Also called rule layer or the fuzzification layer which contains fixed nodes and it is responsible for generating strong rules. In this layer, a fuzzification neuron receives the outcome of the previous layer and specify to which neuron's fuzzy set belongs using membership function or activation function. The outcome of this layer is the fuzzy membership grade of the inputs. It has many functions to calculate membership value using many membership functions for example:
  - Gaussian membership function as in equation :

$$f(x,c,\sigma) = e^{\frac{-(x-c)^2}{2\sigma^2}} \tag{3.18}$$

where {c,**σ**} is the parameter set

- Triangular membership function as in equation:

$$f(x,a,b,c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \tag{3.19}$$

where {a,b,c} are the parameter set.

- Third layer: this layer is called the fuzzy rule layer in which each node it has represents a single fuzzy rule. The outcome of this layer is the multiplication of the incoming signals and it sends the product out to the next layer. As represented below, the output of the neuron i:

$$y_i = x_{1i} \times x_{2i} \times \ldots \times x_{ki} \tag{3.20}$$

$$\mu_{Ri} = y_{Ri} = \mu_{Ai} \times \mu_{Bi} \tag{3.21}$$

Where R_i means the ith neuron in the third layer. For example, as shown in the figure the neuron R_1 inputs are from A_1 and B_1

- Fourth layer: this layer is called the output membership layer. Each neuron or node in this layer represents fuzzy sets and combines all of its inputs using the fuzzy operation union. This value can be implemented by taking the probability OR as follow:

$$y_i = x_{1i} \oplus x_{2i} \oplus \ldots \oplus x_{li} \tag{3.22}$$

$$\mu_{Ci} = y_{Ri} = \mu_{R3} \oplus \mu_{R6} \tag{3.23}$$

Where C_i means the ith neuron in the fourth layer. For example, as shown figure the neuron C_1 inputs are μ_{R3} , μ_{R6}

- Fifth layer: the last layer is the output layer or defuzzification layer. In this layer, the overall output is computed by taking the result of fuzzy sets and transforming them into a single fuzzy set. There are many methods to apply defuzzification. For example, the sum-product composition method.

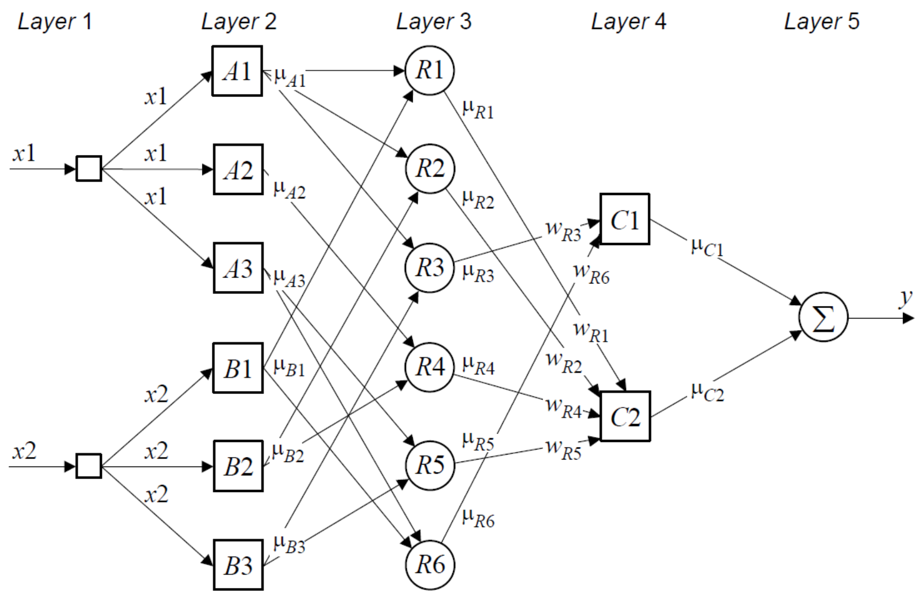

Figure 3.4-9 : Five layers of the ANFIS

In this thesis, the impeded tool in MTALB ANFIS is used to build our model.

Algorithm 2: The general procedure which was used in building the ANFIS model.

Data was divided into two parts: A training set and a testing set.

1- Importing the whole data.
2- Partition the data.
3- Coding the data.
4- Using the training set and training the ANFIS model using Grid/Clustering.
5- Using the testing set and testing the ANFIS model.
6- Getting the FIS model from the ANFIS model results.
7- Indicating if the prediction model accuracy is good or not.
8- Trying to enhance the ANFIS model.

Chapter 4

Experiments and Results

4 Experiments and Results

In the previous chapter, a brief discussion about the algorithms which were used in the comparison with ANFIS to forecast obesity was explained. This chapter will show the found results from our experiments using MATLAB. To evaluate, the model's accuracy will be taken into account, where the accuracy means how many data points are correctly predicted in a ratio to all samples. As mentioned previously, the whole data set is divided into two sets; training set and testing set. Forecasting models take the testing set and predict the outcome using the model, then checking how many data points are correctly predicted. Building the ANFIS model was done by using MATLAB tool called Neuro-Fuzzy Designer, neural pattern recognition tool was used for Neural Network and other classification models, classification learner tool was used, and finally, finding cluster tool was used for clustering.

This chapter is divided into three sections; the first section shows the results of some machine learning methods that are D-tree, KNN, SVM, and logistic regression. The second section presents the results of the neural network and the last section shows the results of the ANFIS model.

4.1 Computing Environment

In the recent work, two different computers were used, the first device with fine specifications was very slow in doing experiments and gives out of memory sometimes. So, different computers with higher performance were used; the two devices are:

1. MSI gt72s 6qe dominator pro g: I7-6820HK 2.70GHz, RAM: 64GB, HD: 2 TB SSD with windows 10 pro. figure 4.1-1 shows the performance of this device while running the MATLAB ANFIS tool.

2. Asus AR5B 125: i5-4200u 2.30GHz, RAM: 6GB, HD: 500GB HDD with windows 10 pro.

For building and doing the experiments MATLAP_R2019b was used.

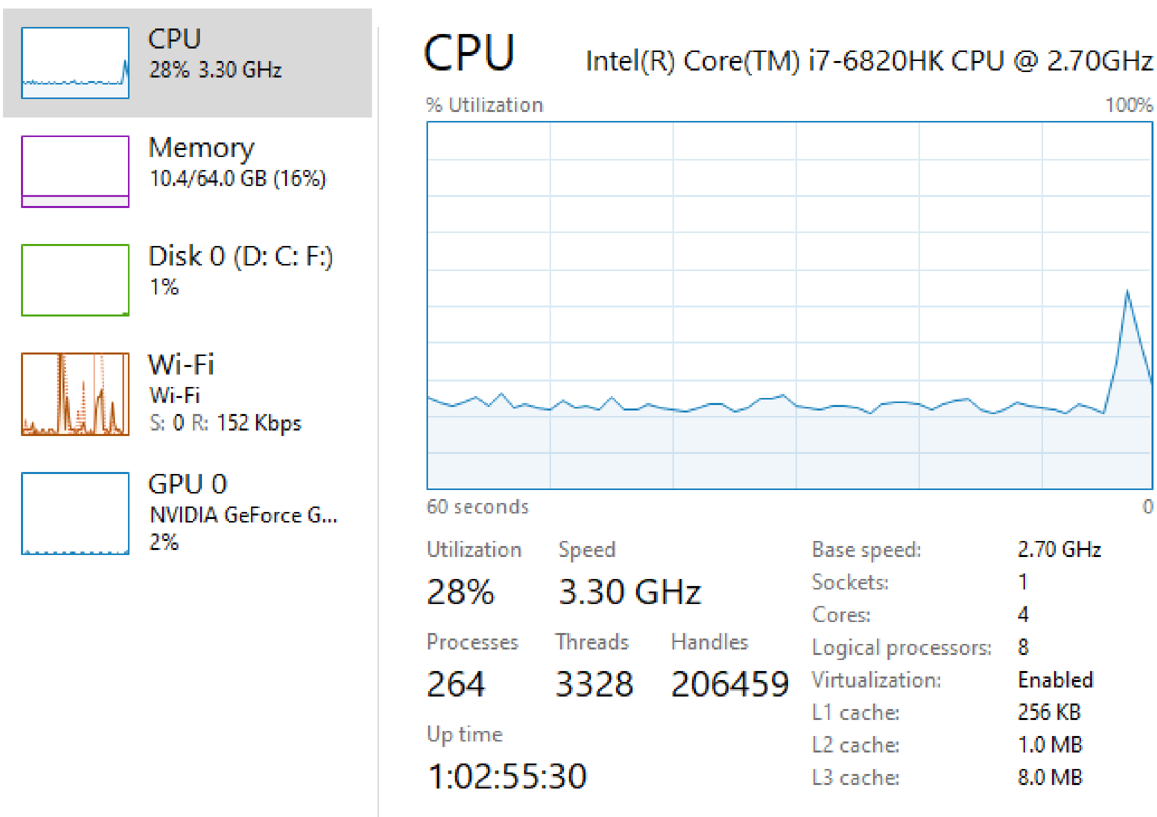

Figure 4.1-1 : The performance of MSI device while running MATLAB ANFIS too

4.2 Machine Learning Results

Supervised machine learning methods were used, which means prior knowledge about the sample dataset to predict a new dataset [55]. Some main measurements like; the accuracy, confusion matrix, and area under curve "AUC" were shown for all algorithms D-tree, KNN, SVM, and logistic regression. Before showing the results, the measurements must be clarified as follow :

- **Accuracy:** This means how many data points are correctly predicted in a ratio to all samples.
- **Confusion matrix:** it is a two-dimensional matrix; the columns represent the actual values of the target variable and the rows show the predicted values of the target variable Figure 4.2-1shows the confusion matrix.

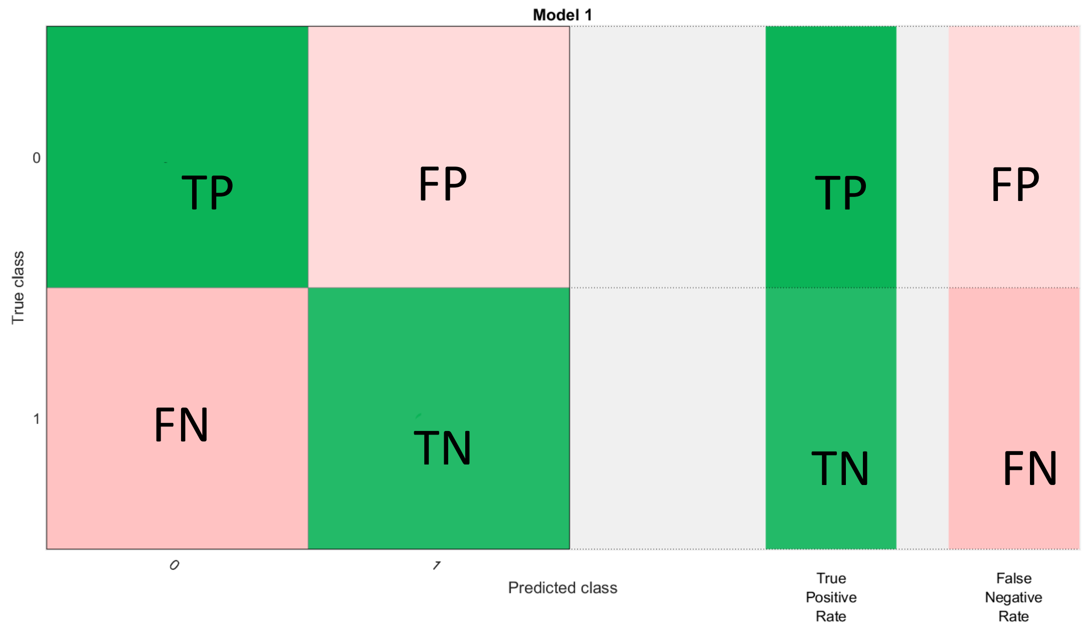

Figure 4.2-1 : Confusion matrix structure

  - **True Positive TP:** Number of data points that are in positive class and classified as the positive class.

- **True Negative TN:** Number of data points that are in negative class and classified as negative class.
- **False-Positive FP:** Number of data points that are not in the positive class and classified as the positive class.
- **False-Negative FN:** Number of data points that are not in negative class and classified as negative class.

- **Receiver Operating Characteristic Curve "ROC":** Shows the performance of classification at all classification thresholds. The curve is plotted True Positive Rate "TPR" on the y-axis and False Positive Rate "FPR" on x-axis fig where:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{4.1}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}} \tag{4.2}$$

- **Sensitivity or Recall:** measures the positives data that are correctly identified from the whole set. Using the following equation to measure:

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{4.3}$$

- **Specificity:** measures the negatives data that are correctly identified from the whole set. Using the following equation to measure:

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN+FP}} \tag{4.4}$$

The Area Under Curve "AUC": This represents the areas under the ROC curve and represents the ability of that model to classify classes. Figure 4.2-2 shows AUC and ROC.

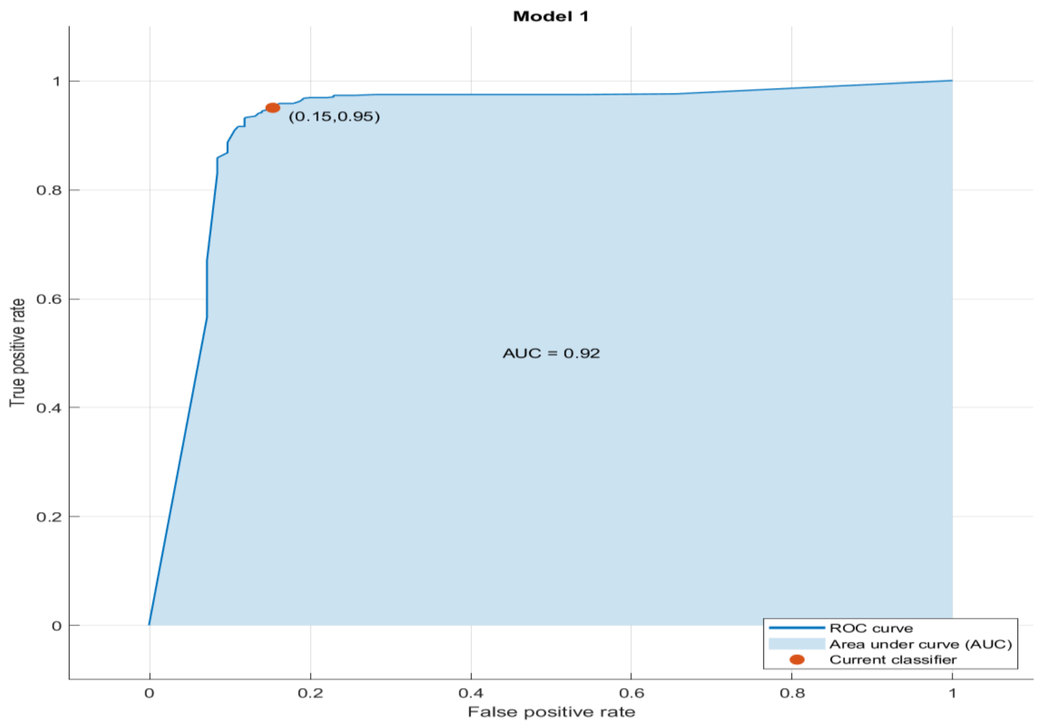

Figure 4.2-2 : AUC and ROC

In all these models, the k-fold cross-validation methodology was used to divide data into training and testing sets to evaluate the classification of model performance, where the data is divided into equally sizes k fold randomly, one of the k folds will be used for testing and the rest for training. In the recent experiments, for example, if 5-fold cross validation is used, 970 records each fold 194 records are obtained, 1 fold for testing randomly, and 4 folds for training. Figure 4.2-3 shows the Folds cross validation example.

194	194	194	194	194

Figure 4.2-3 : 5-Folds cross validation example

To build these models, the classification learner tool is used in MATLAB.

Test 1: This experiment was applied directly using the built-in MATLAB tool, 3-fold cross-validation is used and 9 inputs one output. The experiments were done at all cities together then for each city for a case, Table 3 shows the results:

Table 3 **:** Test 1 Results

Data	**Model Type**	**Validation**	**Accuracy**	**Sensitivity**	**Specificity**
All cities	D-tree	3-fold	92.50%	86.30%	94.40%
	KNN	3-fold	92.80%	81.30%	95.30%
	SVM	3-fold	96.60%	92.40%	97.80%
	Logistic Reg.	3-fold	**97.40%**	**94.20%**	**98.90%**
City 1	D-tree	3-fold	90.90%	77.80%	93.50%
	KNN	3-fold	90.40%	73.20%	94.20%
	SVM	3-fold	93.30%	76.50%	**97.20%**
	Logistic Reg.	3-fold	**94.30%**	**87.20%**	95.50%
City 2	D-tree	3-fold	91.10%	83.30%	94.10%
	KNN	3-fold	90.60%	80.00%	95.00%
	SVM	3-fold	94.20%	86.70%	**97.70%**
	Logistic Reg.	3-fold	**95.10%**	**92.30%**	95.80%
City 3	D-tree	3-fold	89.70%	84.40%	91.20%
	KNN	3-fold	91.10%	76.30%	95.80%
	SVM	3-fold	96.1%	91.50%	**97.80%**
	Logistic Reg.	3-fold	**96.40%**	**94.10%**	96.90%
City 4	D-tree	3-fold	89.80%	79.80%	95.80%
	KNN	3-fold	91.40%	84.80%	94.30%
	SVM	3-fold	94.20%	**94.20%**	**97.90%**
	Logistic Reg	3-fold	**95.70%**	92.30%	96.80%

It was concluded that logistic regression is better in this case and it is close to SVM results; here 3-fold cross-validation was used, which (it) means that 33.33% testing data and 66.67% for training and the best accuracy obtained here is 97.40%. Also, the best sensitivity is 94.20% and specificity is 98.90%.

Test 2: This experiment was applied directly using the built-in MATLAB tool,5-fold cross-validation is used and 9 inputs one output. The experiments were done at all cities together then for each city for a case, Table 4 shows the results:

Table 4 : Test 2 results

Data	**Model Type**	**Validation**	**Accuracy**	**Sensitivity**	**Specificity**
All cities	D-tree	5-fold	92.60%	86.30%	94.40%
	KNN	5-fold	93.10%	85.70%	95.40%
	SVM	5-fold	97.20%	93.30%	98.90%
	Logistic Reg.	5-fold	**98.20%**	**96.10%**	**98.90%**
City 1	D-tree	5-fold	93.30%	80.10%	96.20%
	KNN	5-fold	91.40%	73.40%	95.50%
	SVM	5-fold	**94.30%**	80.30%	**97.40%**
	Logistic Reg.	5-fold	92.80%	**86.30%**	94.30%
City 2	D-tree	5-fold	93.80%	**89.70%**	95.60%
	KNN	5-fold	92.90%	86.40%	95.50%
	SVM	5-fold	**94.60%**	86.70%	**97.70%**
	Logistic Reg.	5-fold	93.80%	86.40%	95.50%
City 3	D-tree	5-fold	94.30%	83.70%	97.50%
	KNN	5-fold	92.90%	80.90%	97.40%
	SVM	5-fold	**96.4%**	**91.50%**	**97.80%**
	Logistic Reg.	5-fold	96.10%	**91.50%**	**97.80%**
City 4	D-tree	5-fold	91.80%	84.80%	94.30%
	KNN	5-fold	94.90%	90.60%	96.70%
	SVM	5-fold	**97.30%**	**96.07%**	**97.90%**
	Logistic Reg.	5-fold	96.50%	96.03%	96.90%

It was concluded that logistic regression for the big data (which may be different from each other) is better than the other three models, but for small data and similar data from each city in a case, SVM was the best model. When 5-folds cross-validation is used, which (it) means 20% testing data and 80% for training, and the best accuracy obtained here is 98.20%.

Test 3: This experiment was applied directly using the built-in MATLAB tool, 10-fold cross-validation is used and 9 inputs one output. The experiments were done on all cities together then for each city for a case, Table 5 shows the results

Table 5 : Test 3 results

Data	Model type	Validation	Accuracy	Sensitivity	Specificity
All cities	D-tree	10-fold	92.30%	84.80%	94.30%
	KNN	10-fold	93.60%	85.80%	96.50%
	SVM	10-fold	97.10%	92.50%	**98.90%**
	Logistic reg.	10-fold	**97.50%**	**94.20%**	**98.90%**
City 1	D-tree	10-fold	92.80%	85.50%	94.30%
	KNN	10-fold	90.40%	73.20%	94.20%
	SVM	10-fold	93.30%	78.40%	**97.30%**
	Logistic reg.	10-fold	**95.20%**	**87.30%**	96.30%
City 2	D-tree	10-fold	93.30%	86.40%	95.50%
	KNN	10-fold	91.10%	83.90%	93.10%
	SVM	10-fold	93.30%	85.90%	**97.60%**
	Logistic reg.	10-fold	**95.10%**	**92.30%**	95.80%
City 3	D-tree	10-fold	91.50%	79.30%	94.90%
	KNN	10-fold	92.90%	81.50%	96.20%
	SVM	10-fold	96.40%	92.40%	97.80%
	Logistic reg.	10-fold	**97.90%**	**95.10%**	**98.90%**
City 4	D-tree	10-fold	91.80%	84.80%	94.30%
	KNN	10-fold	94.10%	92.20%	94.80%
	SVM	10-fold	**97.30%**	**96.07%**	**97.90%**
	Logistic reg	10-fold	**97.30%**	**96.07%**	**97.90%**

It was concluded that logistic regression is better in all cases when the training data is bigger, also the accuracy for all other models is better in most cases when 10 folds cross-validation is used, which (it) means 10% of the data used for testing and 90% for training and the best accuracy we got here is 97.90%.

The following table 6 shows the comparison between the three tests, the best case from each experiment was chosen:

Table 6 : Folds Comparison

Data	k-fold	Model type	Accuracy	Sensitivity	Specificity
All cities	3-fold	Logistic regression	97.40%	94.20%	98.90%
	5-fold	**Logistic regression**	**98.20%**	**96.10%**	**98.90%**
	10-fold	Logistic regression	97.50%	94.20%	98.90%
City1	3-fold	Logistic regression	94.30%	87.20%	95.50%
	5-fold	SVM	94.30%	80.30%	**97.40%**
	10-fold	**Logistic regression**	**95.20%**	**87.30%**	96.30%
City2	**3-fold**	**Logistic regression**	**95.10%**	**92.30%**	95.80%
	5-fold	SVM	94.60%	86.70%	**97.70%**
	10-fold	**Logistic regression**	**95.10%**	**92.30%**	95.80%
City3	3-fold	Logistic regression	96.40%	94.10%	96.90%
	5-fold	SVM	96.4%	91.50%	97.80%
	10-fold	**Logistic regression**	**97.90%**	**95.10%**	**98.90%**
City4	3-fold	Logistic regression	95.70%	92.30%	96.80%
	5-fold	**SVM**	**97.30%**	**96.07%**	**97.90%**
	10-fold	**SVM or logistic**	**97.30%**	**96.07%**	**97.90%**

From the three tests, it is shown that whenever k-folds were used, the logistic regression will mostly give the best accuracy, sensitivity, and specificity when large data e.g. all cities data, especially when 5-folds cross-validation is used, the best accuracy is obtained using logistic regression 98.20%.

In general, using 10-fold cross-validation gives the best accuracy and sensitivity in all cases except when using all cities data 5-fold cross-validation, it gives a little bit better. Also, in many cases 5-fold cross-validation, it gives better results when classifying negative class correctly (Specificity). It is decided to choose 10-fold cross-validation to be used in the comparison in section 4.4. The following figures from 4.2-4 to 4.2-8 represent the table comparison above:

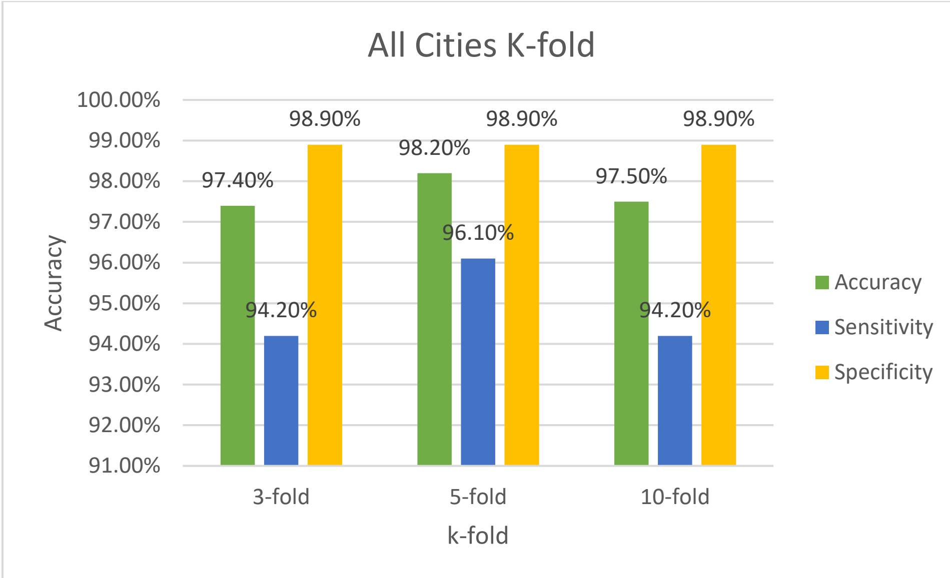

Figure 4.2-4 : All Cities K-fold comparison

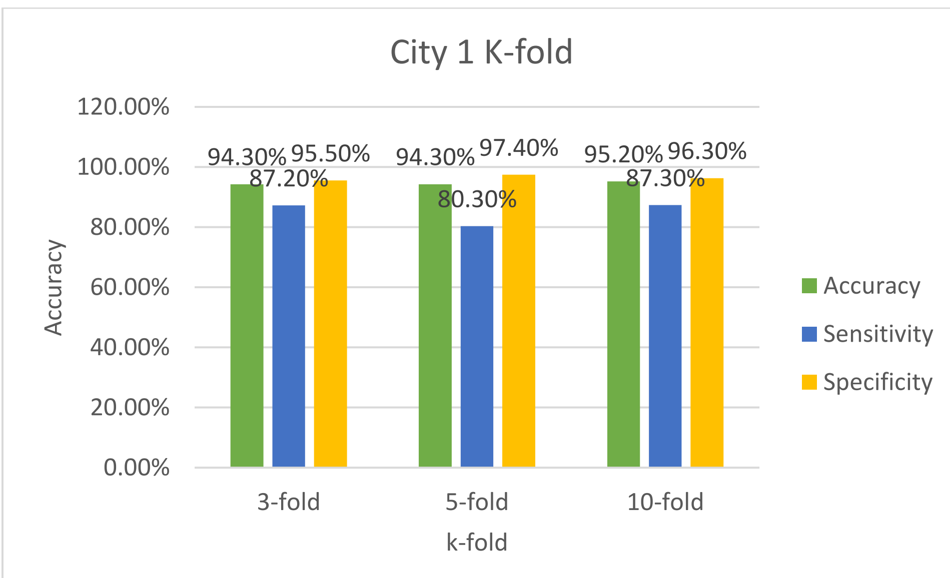

Figure 4.2-5 : City 1 K-fold

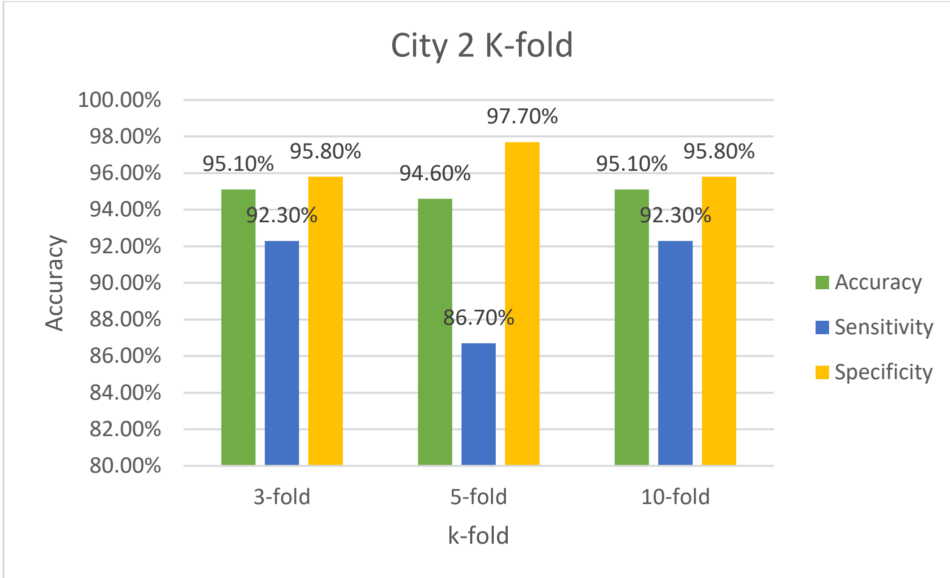

Figure 4.2-6 : City 2 K-fold

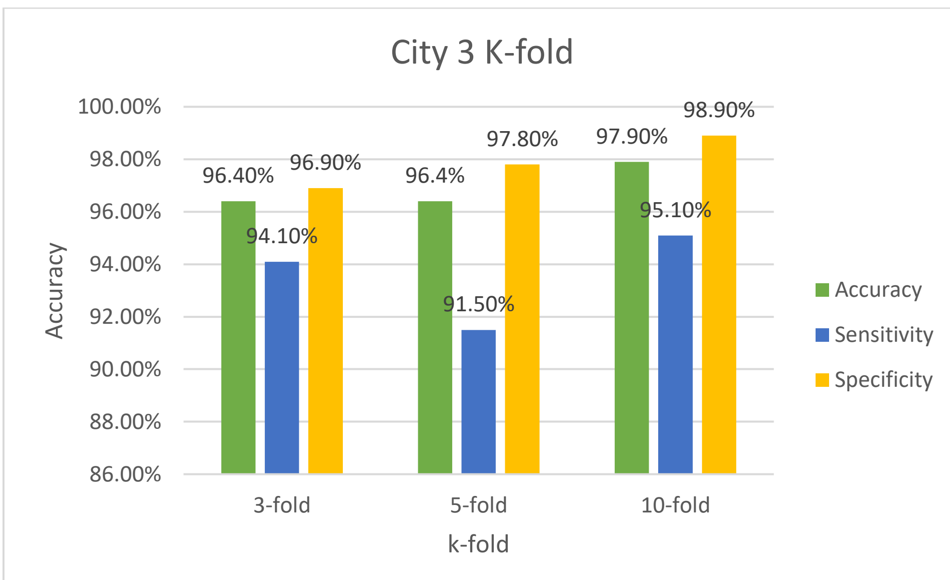

Figure 4.2-7 : City 3 K-fold

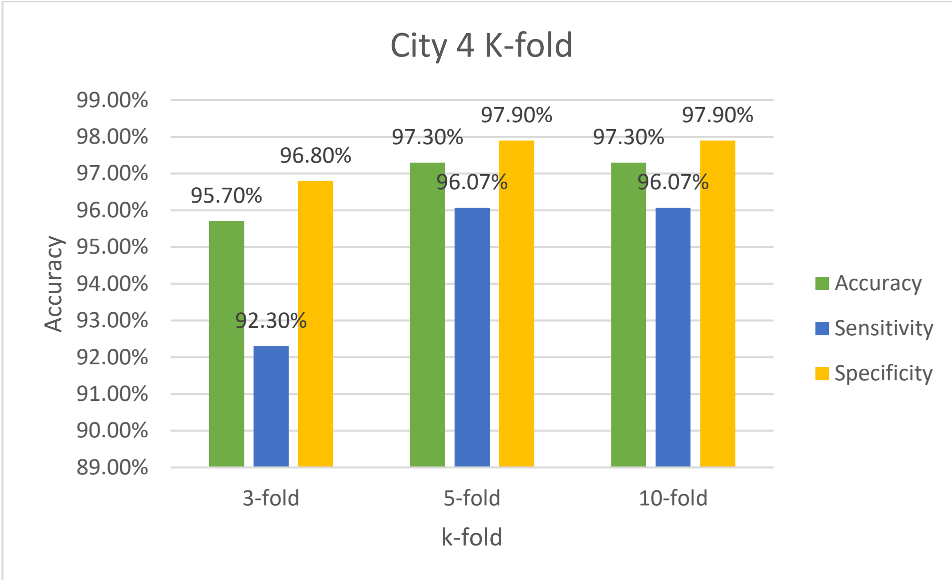

Figure 4.2-8 : City 4 K-fold

4.3 Neural Network Results

The Neural Network is one of the strongest classification methods. Many tests were done to get the best results. For each test of 2, 4, 6, 8, and 10, neurons were used, and the data was divided in all the tests as following:

- 15% of the data for testing.
- 15% of the data for validation.
- 70% for training.

To build these models, the nprtool tool is used in MATLAB.

Table 7 shows the results when using 2, 4, 6,8,10 neurons, using different data as input.

Table 7 : Neural Network Results

Data	Number of Neurons	Accuracy
All cities	2	97.8%
	4	97.7%
	6	97.4%
	8	97.8%
	10	**98.4%**
City 1	2	93.3%
	4	**97.1%**
	6	95.2%
	8	95.2%
	10	94.7%
City 2	2	**96.9%**
	4	95.5%
	6	**96.9%**
	8	93.8%
	10	96.0%
City 3	2	96.8%
	4	94.0%
	6	96.8%
	8	95.7%
	10	**97.9%**
City 4	2	**97.7%**
	4	**97.7%**
	6	96.1%
	8	96.1%
	10	96.5%

The best accuracy obtained here was when the whole data (970 samples) from the whole cities together was used (98.4%) using 10 neurons and it is better than the previous models in section 4.1. However, when the number of records is small relatively, using 10 neurons, it doesn't give the best accuracy, for example, the experiment using City 1 dataset (209 samples) when 4 neurons are used it gives the best accuracy, using City 3 dataset (281 samples) larger from City 1 dataset also using 10 neurons gives the best accuracy. The figures from 4.3-1 to 4.3-5 represent each dataset the accuracy comparison in each case.

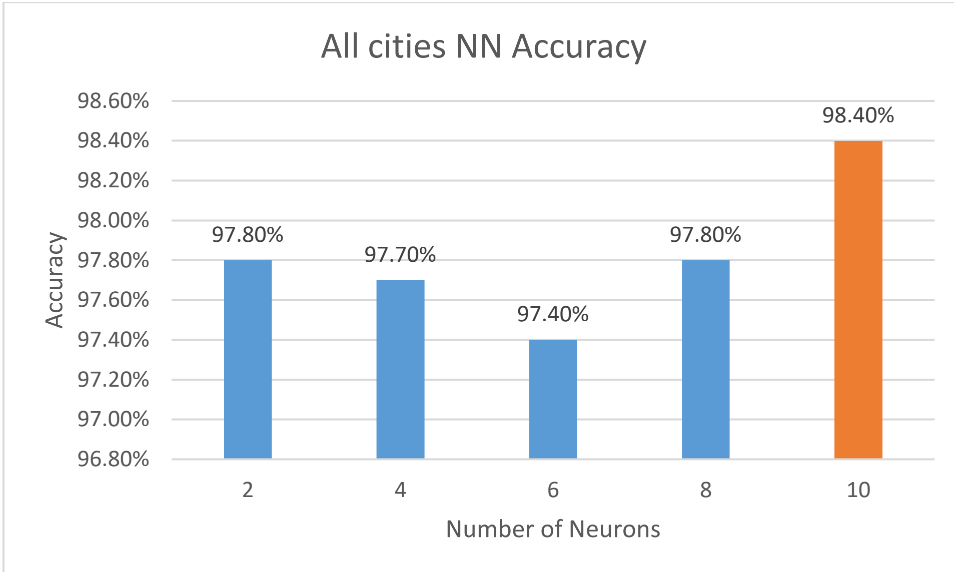

Figure 4.3-1 : All cities NN Accuracy

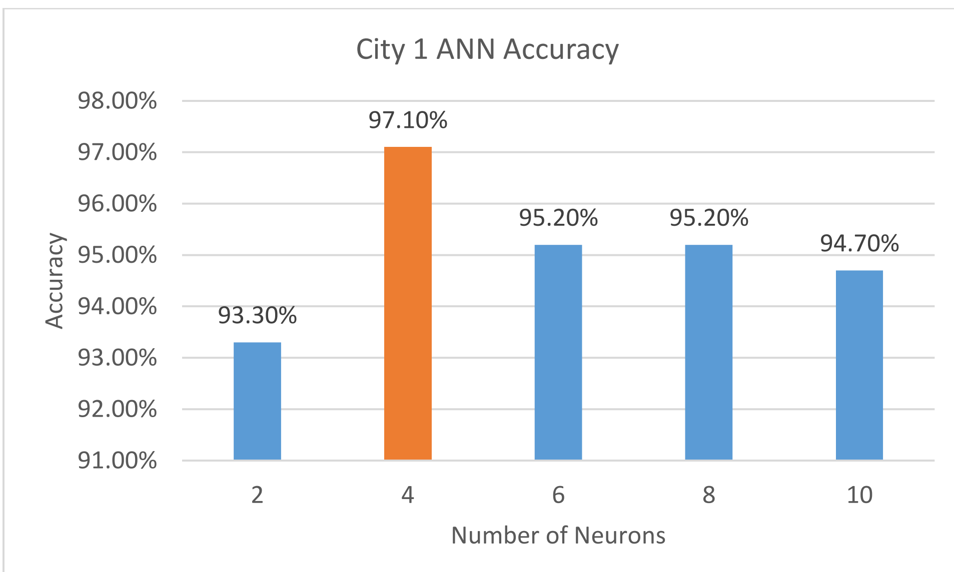

Figure 4.3-2 : City 1 NN Accuracy

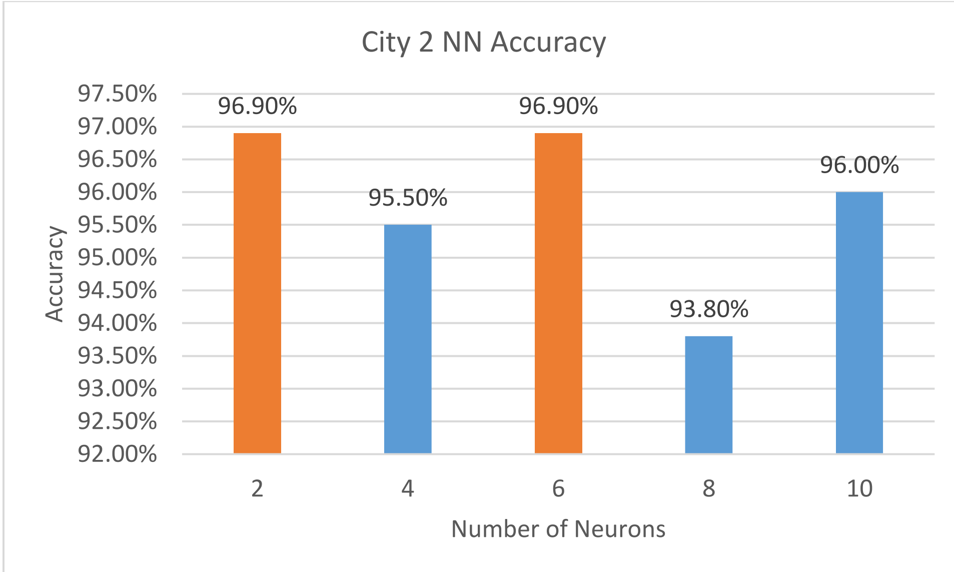

Figure 4.3-3 : City 2 NN Accuracy

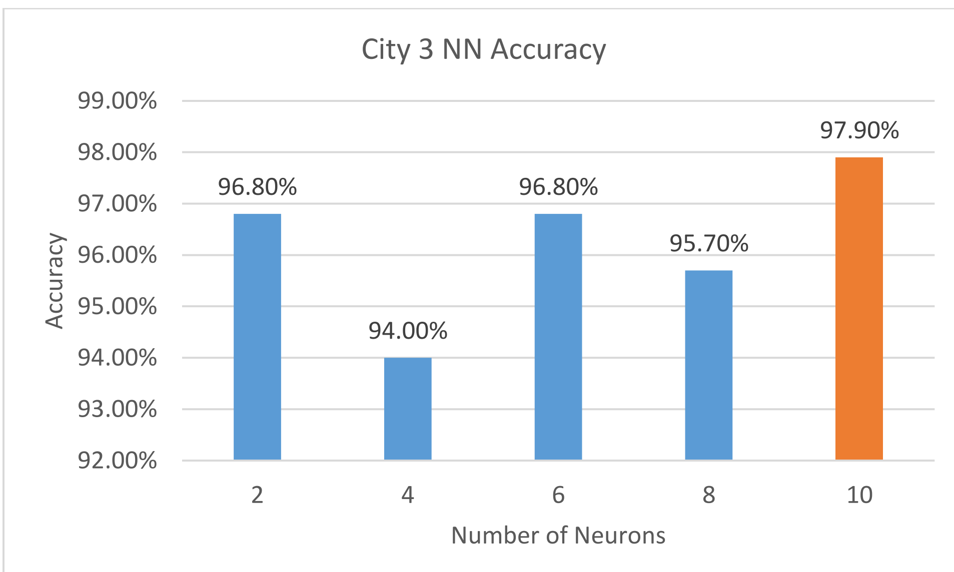

Figure 4.3-4 : City 3 NN Accuracy

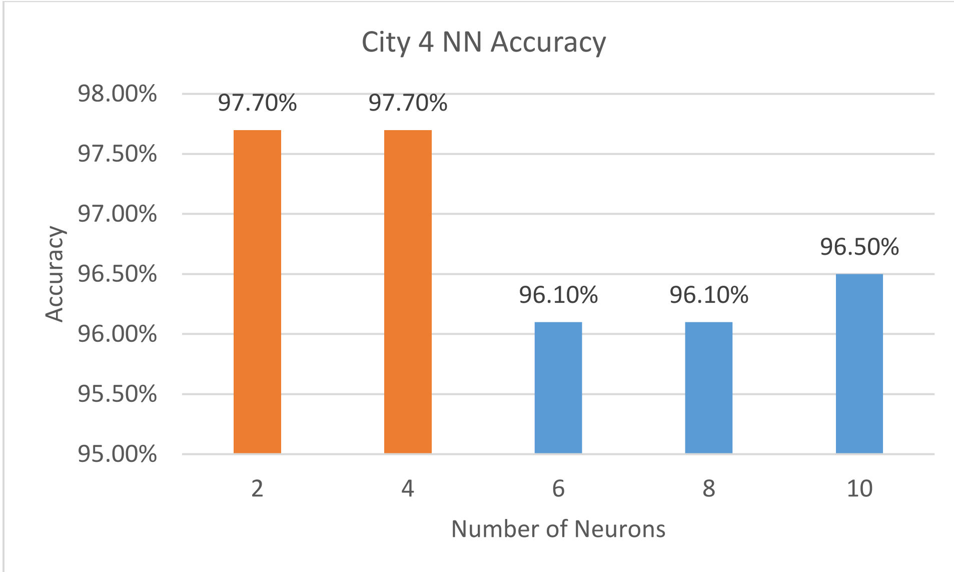

Figure 4.3-5 : City 4 NN Accuracy

4.4 Adaptive Neuro Fuzzy Inference System Results

ANFIS is the integration between neural networks and fuzzy logic. This model is suitable when solving predicting problems and having a large data set with many features.

In our experiments, ANFIS grid partition and ANFIS sub- clustering were used. Where grid partitioning produces a large number of rules because the relation between membership functions and the number of inputs is exponential where:

$$\text{Number of rules} = m^{\mathrm{i}} \tag{4.5}$$

Where *m* is the number of membership functions and i is the number of inputs. in our tests 9 inputs and 2 membership functions were used, so 2^9 = 512 rules shown in Figure 4.4-1 and 4.4-2 are obtained.

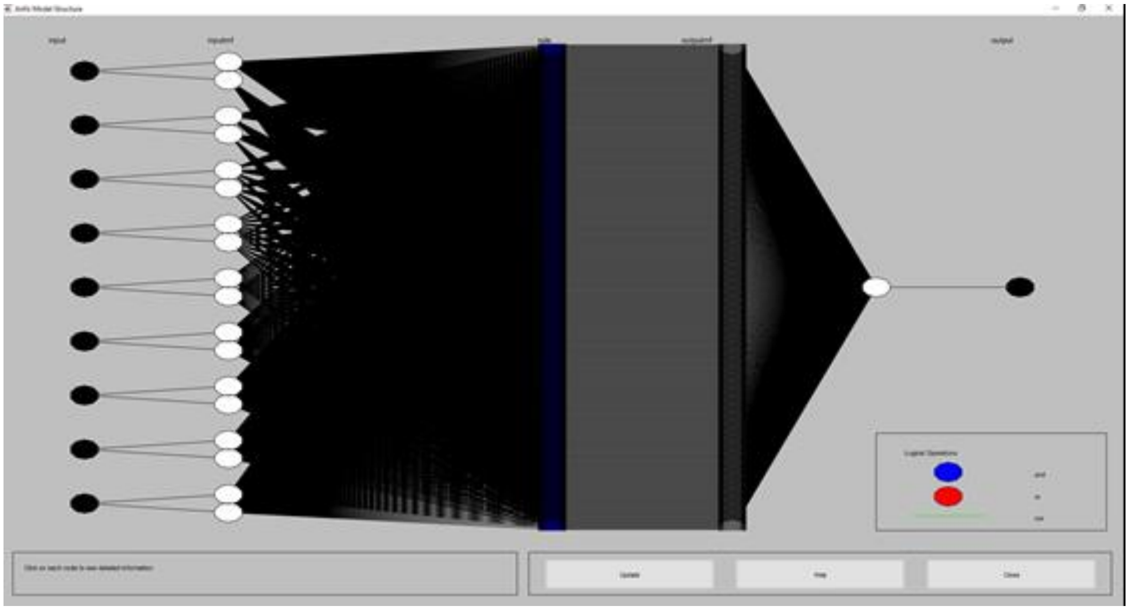

Figure 4.4-1 : Our tests ANFIS-grid structure

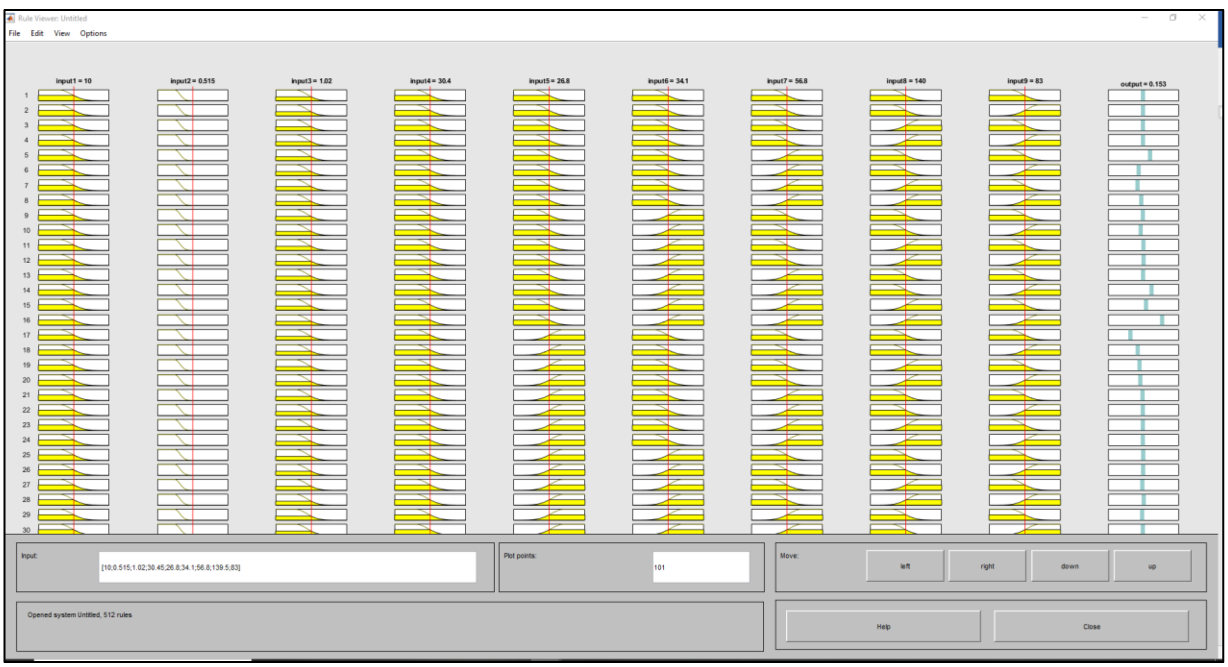

Figure 4.4-2 : ANFIS-Grid rules viewer

The second type is sub- clustering that uses scattering partitioning by subtractive clustering. Here the number of rules is small, and each cluster is one rule. See Figures 4.4-3 and 4.4-4.

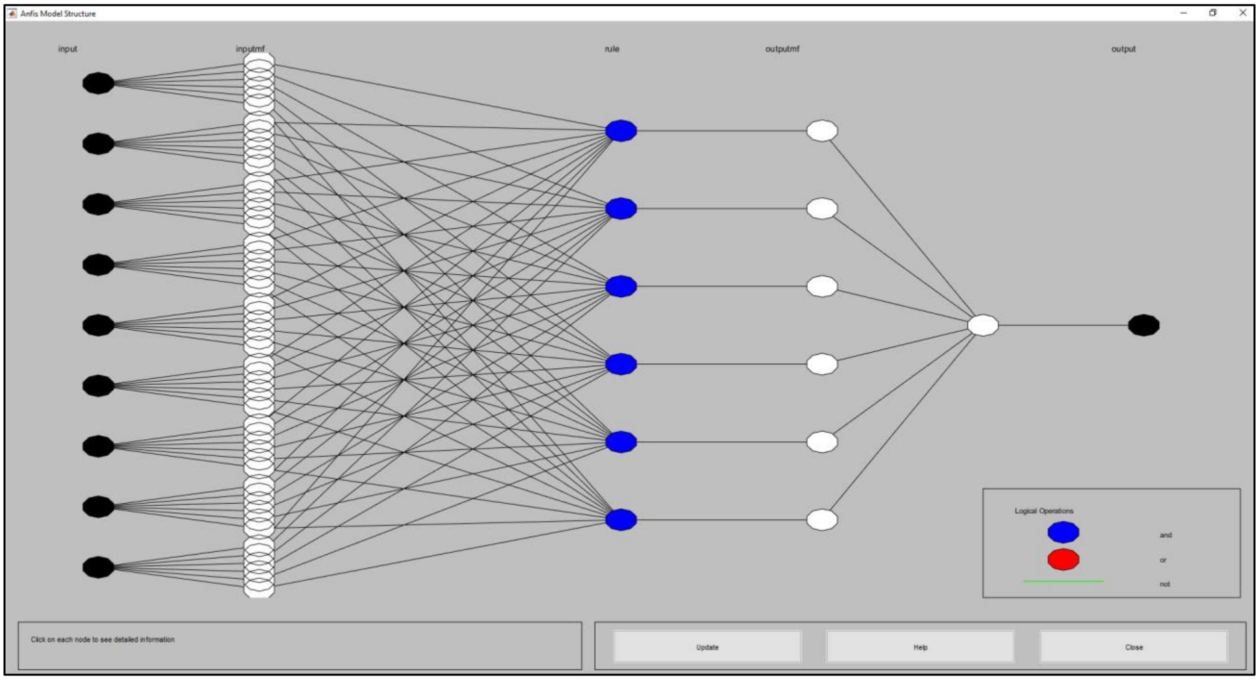

Figure 4.4-3 : Our tests ANFIS-sub cluster structure

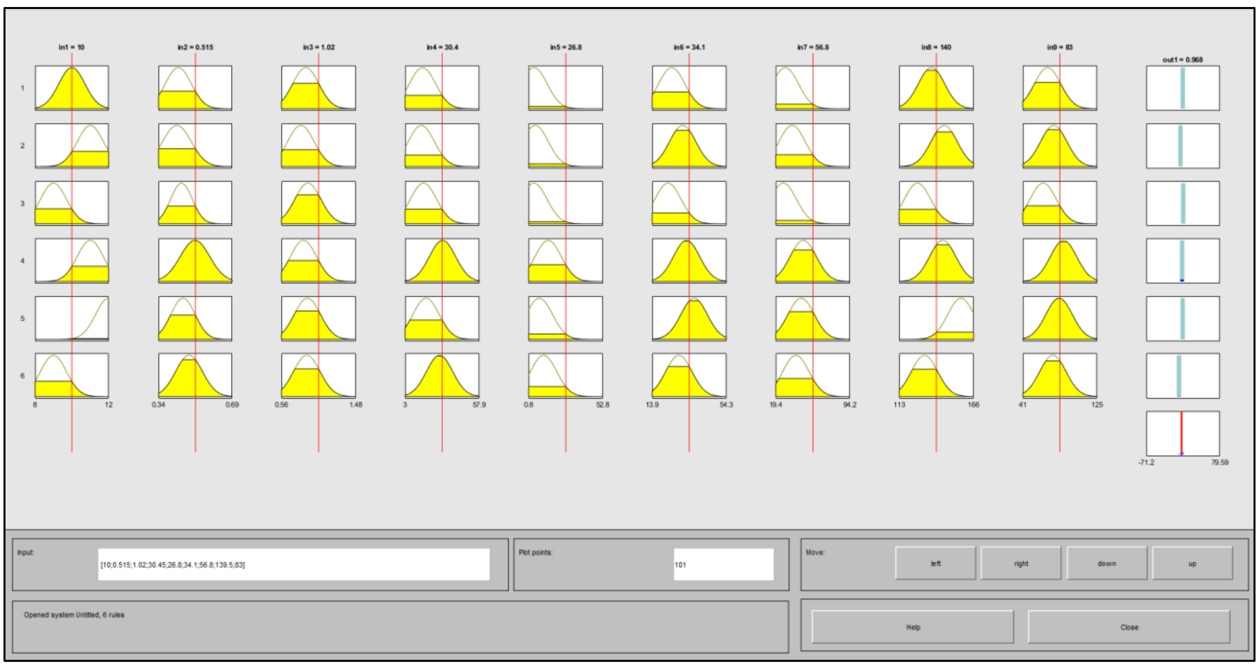

Figure 4.4-4 : ANFIS-sub cluster rules viewer

All data was divided into two sets; 70% of the data is for training and 30% used for testing. For evaluation of each model, Root Mean Square Error (RMSE) and accuracy were used.

<u>Test 1:</u> The data of all cities here is used (970 records), 70% used for training (679 records) the models, and 30% for testing (291 records). Firstly, ANFIS- Grid was applied using many membership functions, which are: triangular, trapezoidal, Gaussian, and Gaussian combination. Then ANFIS-Cluster was applied. Table 8 shows the results for both models and Figure 4.4-5 illustrates a comparison between the membership functions.

Table 8 : ANFIS Test 1 results

FIS	# of MFs	Membership Function (MF)	ERRPR TOLERANCE	# of EPOCHS	RMSE	Accuracy
Grid partition	2	trimf	0.001	15	0.144	97.92%
Grid partition	2	trapmf	0.001	15	0.14	98.04%
Grid partition	2	gaussmf	0.001	15	0.139	98.06%
Grid partition	2	gauss2mf	0.001	15	**0.129**	**98.33%**
Sub-Clustering	Range of influence	0.5	0.001	15	0.186	96.54%
	Squash factor	1.25				
	Accepted ratio	0.5				
	Rejected ratio	0.15				

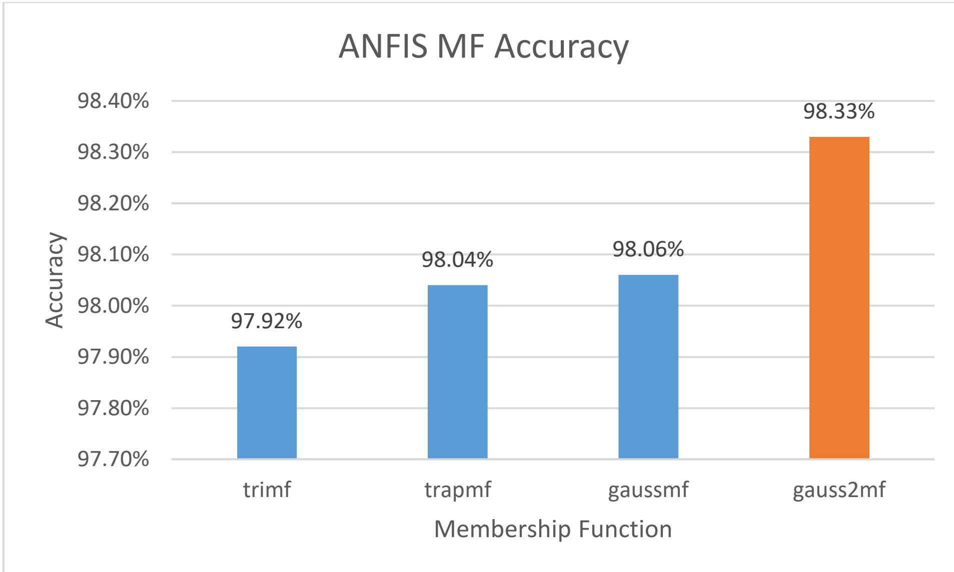

Figure 4.4-5 : Comparison between the membership functions

From the table above, it is shown that the best membership function, in this case, is Gaussian combination since it gives the best accuracy and RMSE, so it was selected to compare it with the ANFIS-Cluster model; six random samples were taken and entered in each model. The results are shown in Table 9 and Figure 4.4-6 are a comparison between model results and actual results.

Table 9 : ANFIS-Grid vs ANFIS-Cluster Test 1

Number of Records	ANFIS-Grid (gauss2mf)	ANFIS-Cluster	Actual Output
1	0.292	0.263	0
2	0.0816	-0.101	0
3	-0.239	-0.254	0
4	0.71	0.937	1
5	1.48	1.46	1
6	1.06	0.995	1

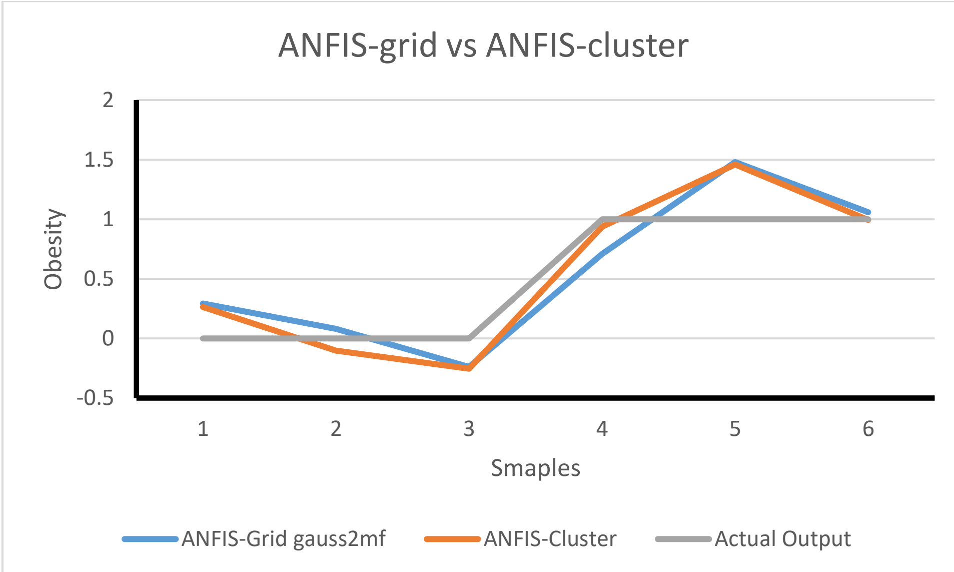

Figure 4.4-6 : Comparison between models results and actual results

From the curves above, it is clear that ANFIS-Cluster is closer to the actual output more than ANFIS-Grid.

<u>**Test 2:**</u> In the first test, the data was large, here in this case each city was tested alone.

- **City 1:** Sample of 209, the same membership functions were used and ANFIS-Cluster. Table 10 shows the results for all models, their Accuracy, and RMSE, Figure 4.4-7 shows a comparison between the membership functions.

Table 10 : ANFIS Test 2 City 1 Results

FIS	# of MFs	Membership Function (MF)	ERRPR TOLERANCE	# of EPOCHS	RMSE	Accuracy
Grid partition	2	trimf	0.001	15	0.0405	99.83
Grid partition	2	trapmf	0.001	15	0.0152	99.96
Grid partition	2	gaussmf	0.001	15	**0.0197**	**99.97**
Grid partition	2	gauss2mf	0.001	15	0.0385	99.85
Sub-Clustering	Range of influence	0.5	0.001	15	0.119	98.58
	Squash factor	1.25				

	Accepted ratio	0.5				
	Rejected ratio	0.15				

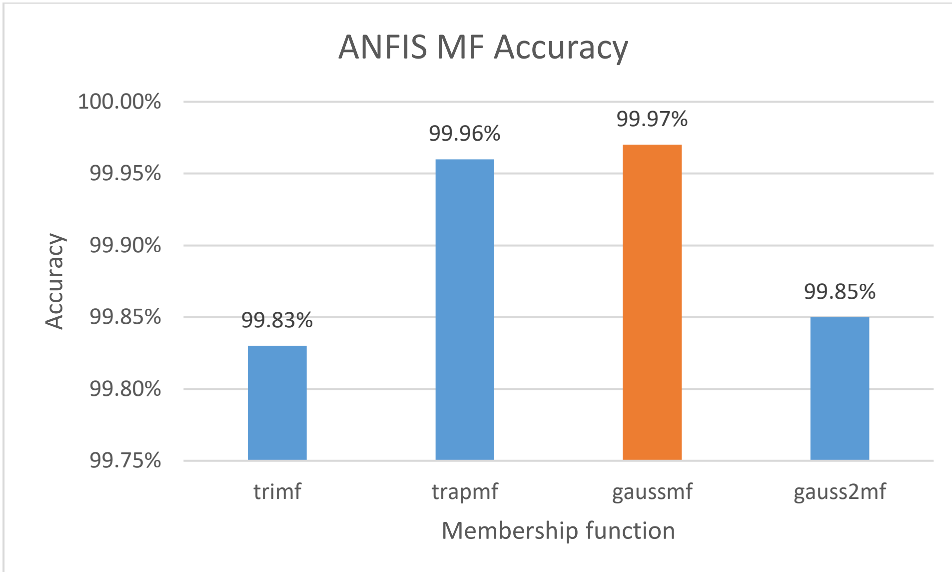

Figure 4.4-7 : Comparison between the membership functions

From Table 8 above, it is shown that the best membership function, in this case, is Gaussian membership function since it gives the best accuracy and RMSE, so it was selected to compare it with the ANFIS-Cluster model; six random samples were taken from that city and entered in each model. The results are shown in Table 11 and Figure 4.4-8 shows a comparison between model results and actual results.

Table 11 : ANFIS-Grid vs ANFIS-Cluster test 2 city 1

Number of Records	ANFIS-Grid (gaussmf)	ANFIS-Cluster	Actual Output
1	0.158	0.0355	0
2	-0.0115	-0.306	0
3	0.101	0.0173	0
4	0.999	0.963	1
5	0.138	0.322	1
6	**0.208**	**0.751**	**1**

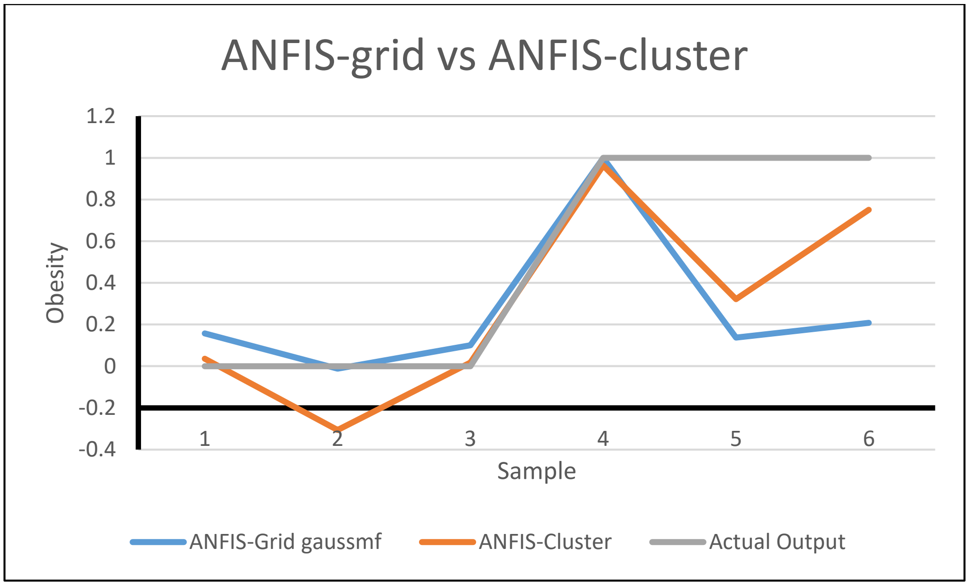

Figure 4.4-8 : Comparison between model results and actual results

- **City 2:** A sample of 224 with the same membership functions and ANFIS-Cluster were used, Table 12 shows the results for all models and its Accuracy and RMSE, and the Figure 4.4-9shows each membership function accuracy.

Table 12 : ANFIS Test 2 City 2 Results

FIS	# of MFs	Membership Function (MF)	ERRPR TOLERANCE	# of EPOCHS	RMSE	Accuracy
Grid partition	2	trimf	0.001	15	0.073	99.46
Grid partition	2	trapmf	0.001	15	**0.017**	**99.97**
Grid partition	2	gaussmf	0.001	15	0.043	99.80
grid partition	2	gauss2mf	0.001	15	0.194	96.20
Sub. Clustering	Range of influence	0.5	0.001	15	**0.04307**	**99.81**
	Squash factor	1.25				
	Accept ratio	0.5				
	Reject ratio	0.15				

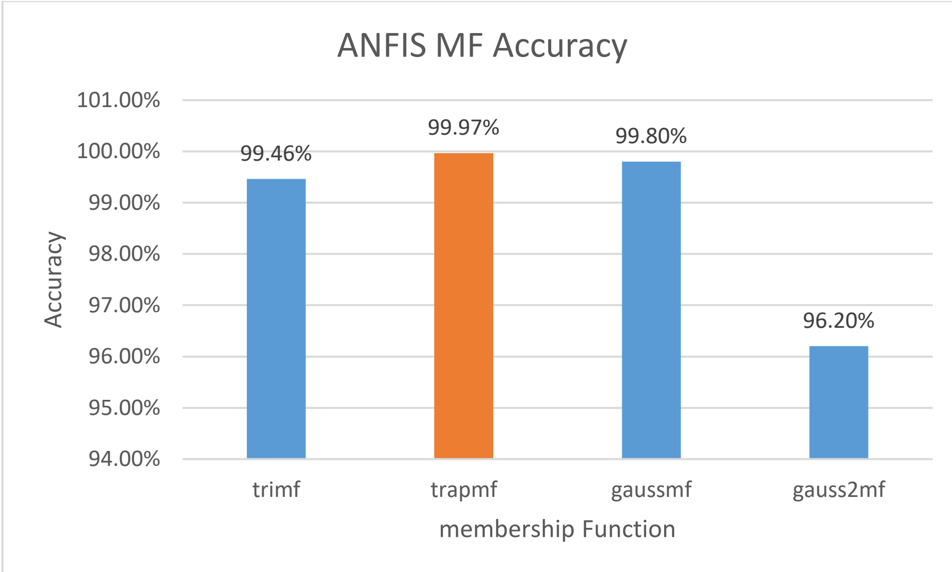

Figure 4.4-9 : Comparison between the membership functions

In this test, the dataset size was 224, using ANFIS-Grid a comparison between four membership function was made and the accuracy results of trimf, trapmf, and gaussmf was very close and very high 99%. The highest accuracy was obtained using the trapezoidal membership function so it was chosen to compare with ANFIS-Cluster, six random samples were selected from the dataset and entered in each model. The results are shown in Table 13 and Figure 4.4-10 shows a comparison between model results and actual results.

Table 13 : ANFIS-Grid vs ANFIS-Cluster test 2 city 2

Number of Records	ANFIS-Grid (trapmf)	ANFIS-Cluster	Actual Output
1	0.013	0.013	0
2	**0.771**	**0.414**	**0**
3	-0.128	-0.048	0
4	0.996	1	1
5	0.939	0.97	1
6	**0.123**	**0.937**	**1**

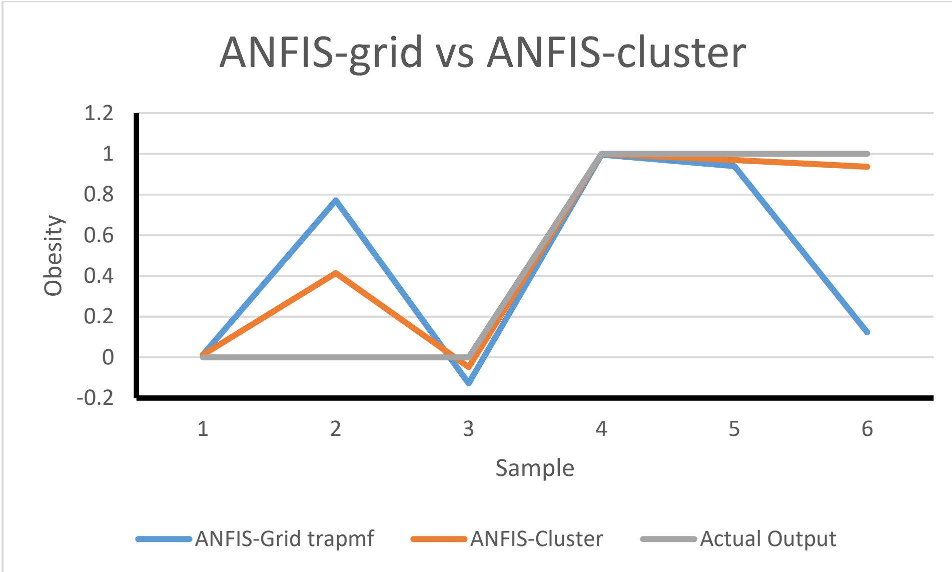

Figure 4.4-10 : Comparison between model results and actual results

.

- **City 3:** sample of 224 with the same membership functions and ANFIS-Cluster were used, Table 14 shows the results for all models and its Accuracy and RMSE, Figure 4.4-11shows comparison between all membership functions.

Table 14 : ANFIS Test 2 city 3 results

FIS	# of MFs	Membership Function (MF)	ERRPR TOLERANCE	# of EPOCHS	RMSE	Accuracy
Grid partition	2	trimf	0.001	15	0.062	99.60
Grid partition	2	trapmf	0.001	15	0.067	99.54
Grid partition	2	gaussmf	0.001	15	0.036	99.86
Grid partition	2	gauss2mf	0.001	15	**0.018**	**99.96**
Sub. Clustering	Range of influence	0.5	0.001	15	**0.116**	**98.65**
	Squash factor	1.25				
	Accept ratio	0.5				
	Reject ratio	0.15				

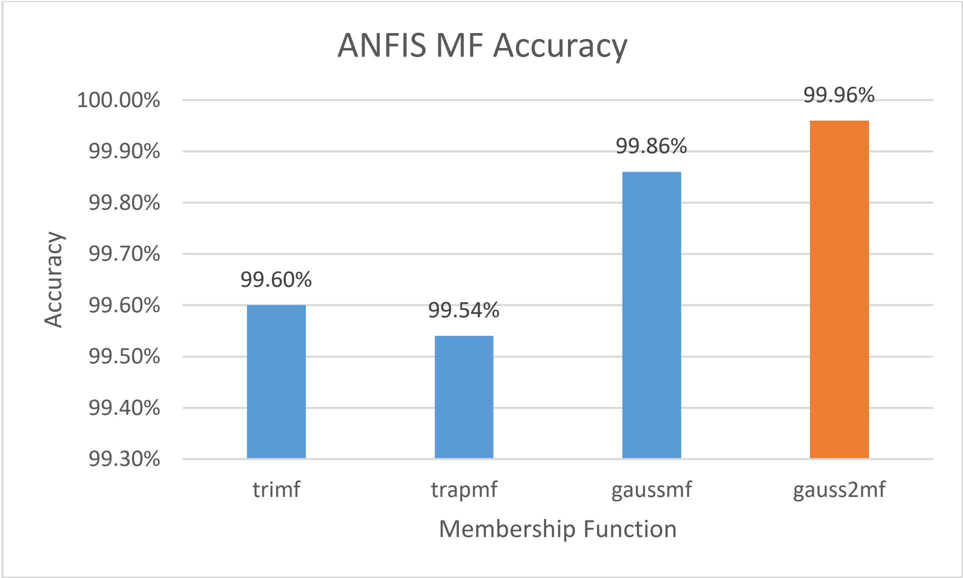

Figure 4.4-11 : Comparison between the membership functions

Testing on city 3 number of samples is 281 also gives, when using ANFIS-Grid, good results the four membership functions were very close inaccuracy. To compare with the ANFIS-Cluster model, gauss2mf was used and six random samples were chosen to make this comparison. The results are shown in Table 15 and Figure 4.4-12 shows a comparison between model results and actual results.

Table 15 : ANFIS-Grid vs ANFIS-Cluster test 2 city 3

Number of Records	ANFIS-Grid (gauss2mf)	ANFIS-Cluster	Actual Output
1	0.188	0.0125	0
2	0.116	-0.0115	0
3	0.202	0.166	0
4	0.88	1.63	1
5	0.972	1.36	1
6	0.544	0.382	1

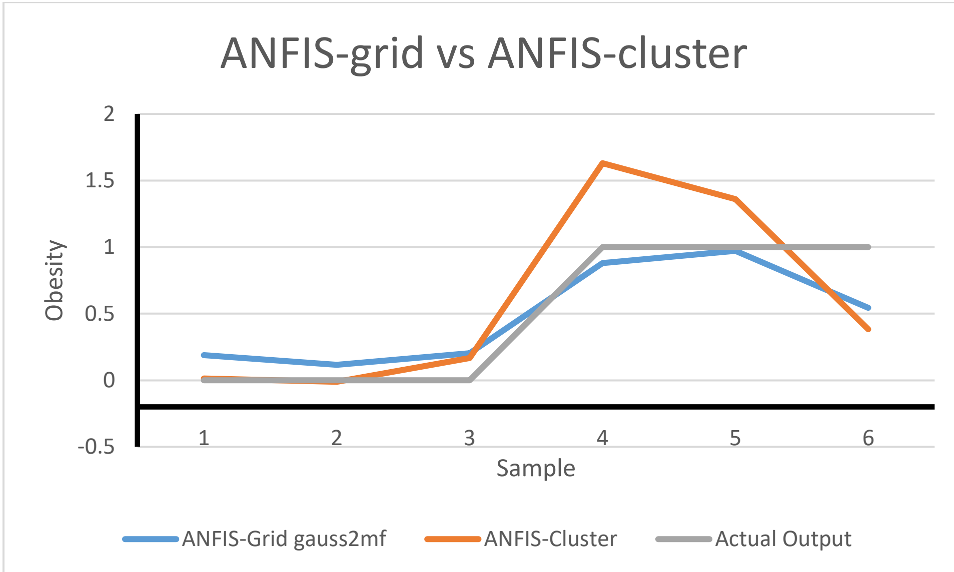

Figure 4.4-12 : Comparison between model results and actual results

- **City 4:** sample of 224 with the same membership functions and ANFIS-Grid were used, Table 16 shows the results for all models and its Accuracy and RMSE. Figure 4.4-13 shows the comparison between all membership functions.

Table 16 : ANFIS Test 2 city 4 results

FIS	# of MFs	Membership Function (MF)	ERRPR TOLERANCE	# of EPOCH S	RMSE	Accuracy
Grid partition	2	trimf	0.001	15	0.080	99.35
Grid partition	2	trapmf	0.001	15	0.068	99.53
Grid partition	2	gaussmf	0.001	15	**0.0587**	**99.65**
Grid partition	2	gauss2mf	0.001	15	0.062	99.61
Sub. Clustering	Range of influence	0.5	0.001	15	**0.075**	**99.42**
	Squash factor	1.25				
	Accept ratio	0.5				
	Reject ratio	0.15				

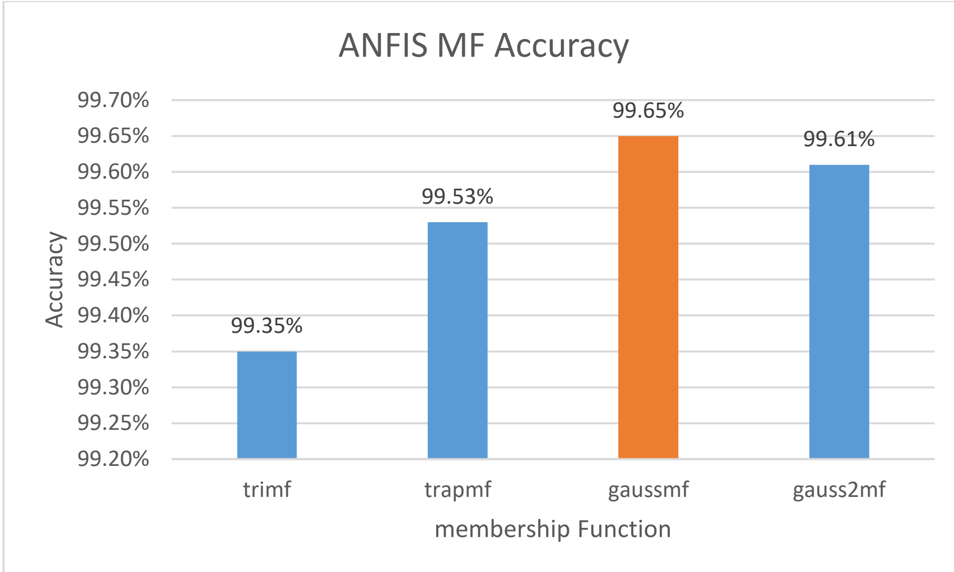

Figure 4.4-13 : Comparison between the membership functions

Table 14 above shows that the best membership function, in this case, is Gaussian since it gives the best accuracy and RMSE, so it was selected to compare it with the ANFIS-Cluster model and six random samples from that city were taken and entered in each model. The results are shown in Table 17 and Figure 4.4-14 shows a comparison between model results and actual results.

Table 17 : ANFIS-Grid vs ANFIS-Cluster test 2 city 4

Number of Records	ANFIS-Grid (gaussmf)	ANFIS-Cluster	Actual Output
1	-0.0156	0.0679	0
2	0.0201	0.136	0
3	0.0529	-0.07711	0
4	0.938	0.885	1
5	1.46	0.224	1
6	2.11	0.821	1

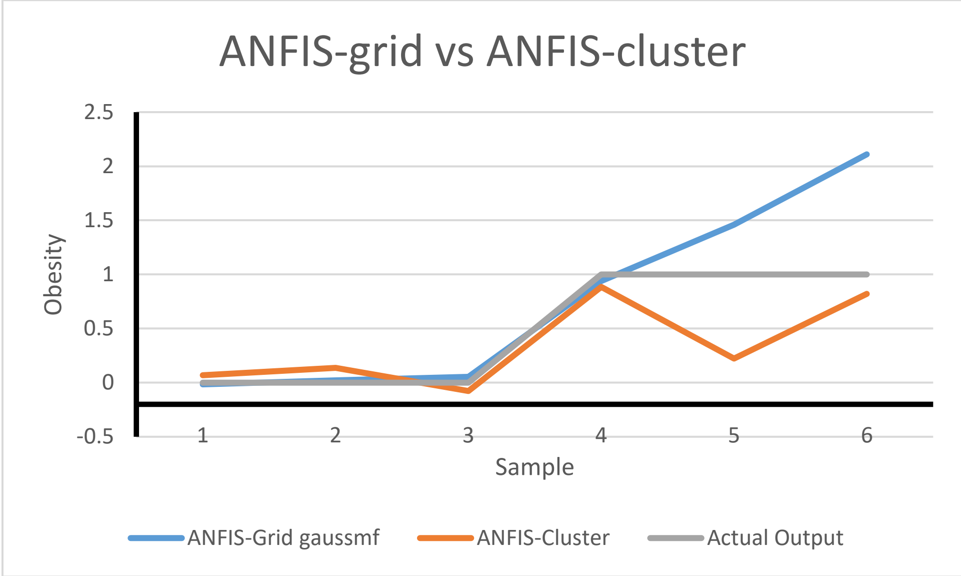

Figure 4.4-14 : Comparison between model results and actual results

4.5 Results Comparison and Discussion

Many tests were done in the previous sections, so in this section, a comparison was done to judge between all algorithms and decided which one is the best algorithm and suitable in predicting obesity disease cases. Based on our experiments, all city data use the ANFIS-grid and neural network is very close inaccurate. In general, the below four algorithms give good results when a large data is used, Table 19 shows the accuracy of each algorithm when the input is all cities data with figures from 4.5-1 to 4.5-5:

Table 18 : Best four models accuracy comparison using all cities data

Data	Algorithm	Accuracy
All Cities	**ANFIS-Grid**	**98.33%**
	ANIFS-Cluster	96.54%
	Logistic Regression	97.50%
	Neural network	**98.40%**

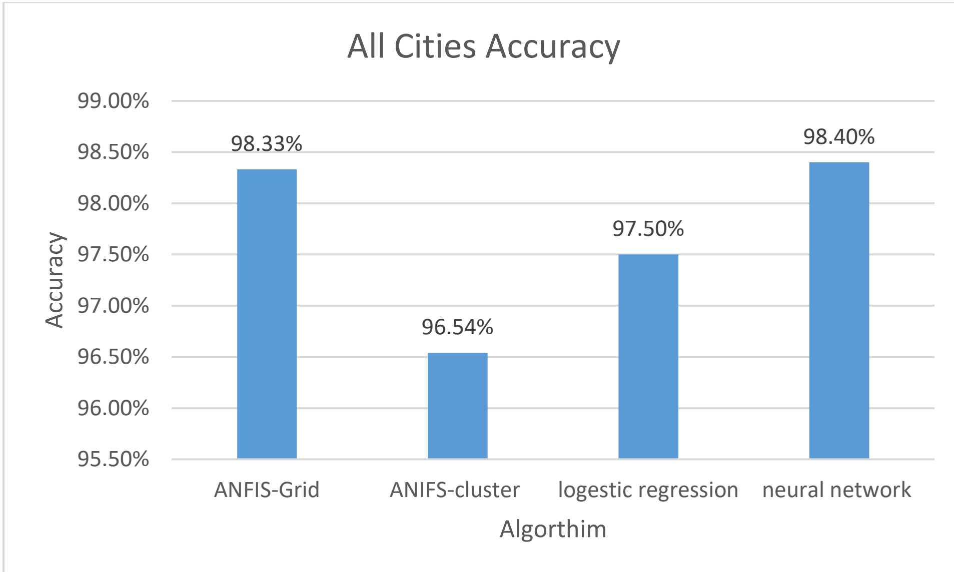

Figure 4.5-1 : All Cities Accuracy

For each city in a case, the ANFIS Grid or Cluster is always better than other algorithms data sets, the following Tables 20 to 23 show a comparison between all techniques:

Table 19 : Best four models accuracy comparison using city 1 data

Data	Algorithm	Accuracy
city 1	**ANFIS-Grid**	**99.97%**
	ANIFS-Cluster	**98.58%**
	Logistic Regression	95.20%
	Neural Network	97.40%

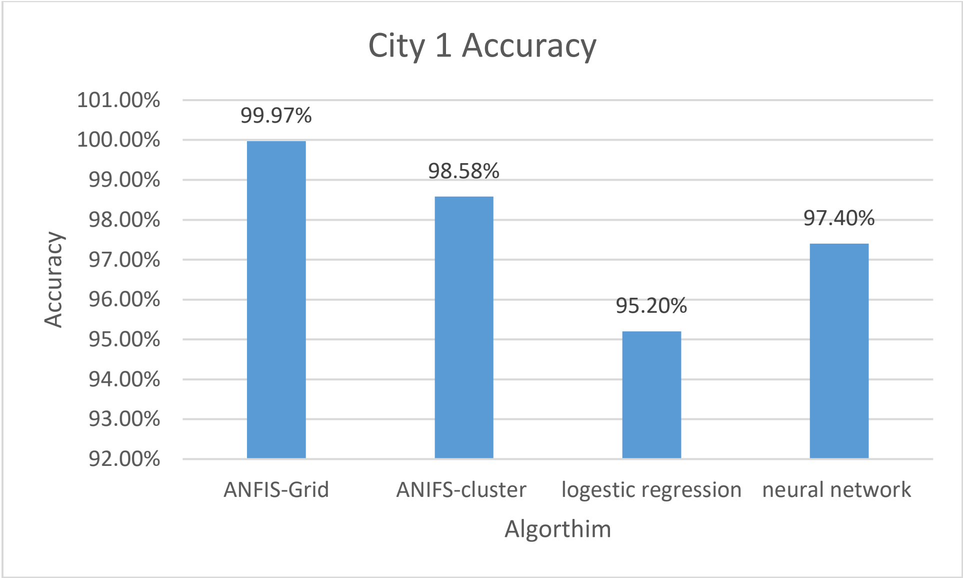

Figure 4.5-2 : City 1 Accuracy

Table 20 : Best four models accuracy comparison using city 2 data

Data	Algorithm	Accuracy
city 2	**ANFIS-Grid**	**99.97%**
	ANIFS-Cluster	**99.81%**
	Logistic Regression	95.10%
	Neural Network	96.90%

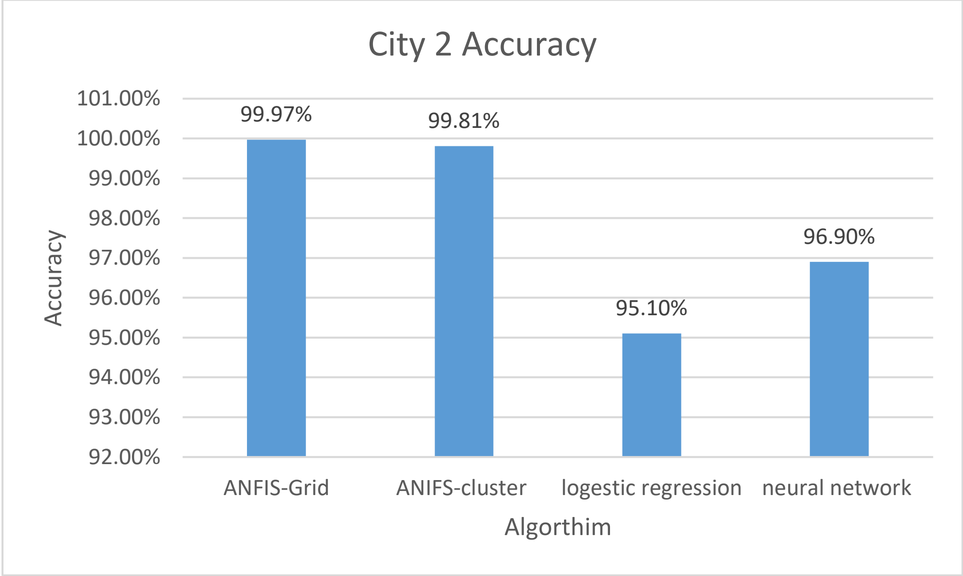

Figure 4.5-3 : City 2 Accuracy

Table 21 : Best four models accuracy comparison using city 3 data

Data	Algorithm	Accuracy
city 3	**ANFIS-Grid**	**99.96%**
	ANIFS-Cluster	**98.65%**

	Logistic Regression	97.90%
	Neural Network	97.90%

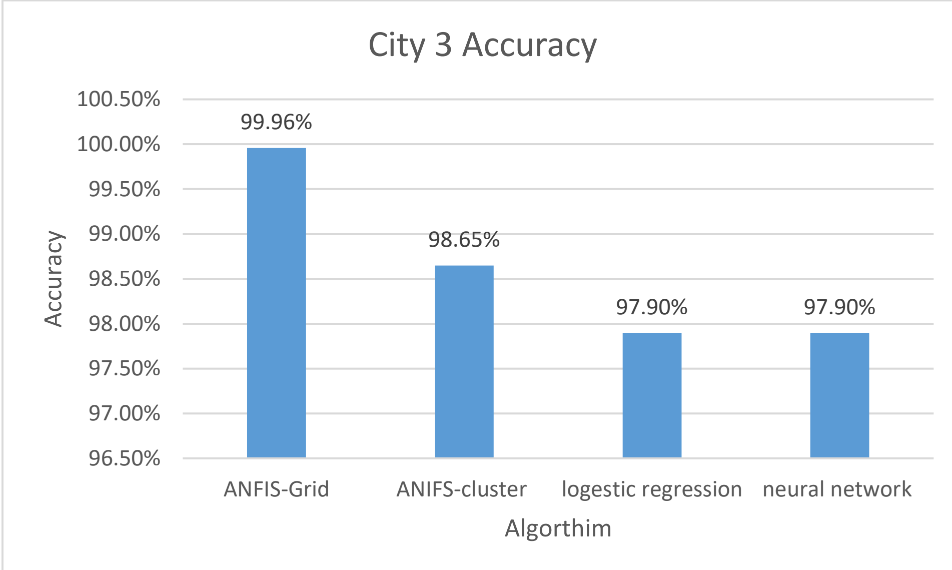

Figure 4.5-4 : City 3 Accuracy

Table 22 : Best four models accuracy comparison using city 4 data

Data	Algorithm	Accuracy
city 4	**ANFIS-Grid**	**99.65%**
	ANIFS-Cluster	**99.43%**
	SVM and LR	97.90%
	Neural Network	97.70%

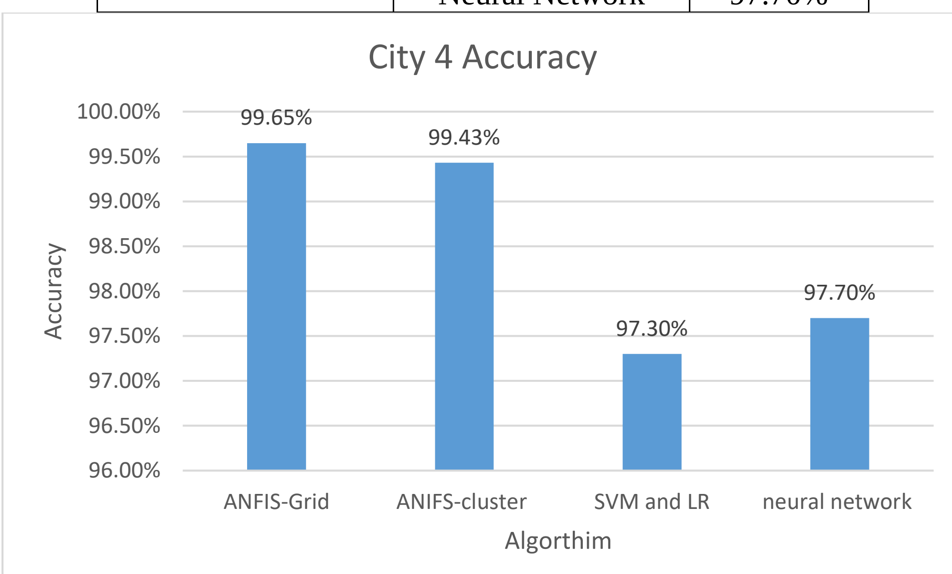

Figure 4.5-5 : City 4 Accuracy

4.6 Research Limitations

This study has potential limitations. The major challenge which was faced is that even when a computer with high specifications was used, the training took a long time, and in ANFIS-Grid, two membership functions were only used for each input if more than two membership functions are demanded, then very high specifications are needed. For collecting data in our society, the inputs of our model are considered private matters (it is not easily disclosed or obtained).

Better results could be obtained if a computer with higher specifications is used. Also, to generalize this research all over Palestine, data from many other cities is needed to be collected and the larger population.

Chapter 5

Conclusions and Future Work

5 Conclusions and Future Works

The medical field is considered the most important one in human life and diagnosing any disease in early-stage makes the treatment process easier than in the middle or final ones. Obesity is considered the main major of other diseases such as heart disease, diabetes, hypertension, stroke, pulmonary emboli, certain cancers, osteoarthritis, gallbladder disease, and respiratory abnormalities [56]. Also, COVID-19 severity increased when people are obese [57]. However, early and accurate diagnosis can save lives and gives the appropriate treatment, which helps the public health managing. Expert systems are considered very strong decision support systems based on the decisions and experiences of people who have been in that field. Also, over time, these systems gained more experience in analyzing and giving more accurate results.

In this thesis, an intelligent medical diagnosis used for the diagnosis of obesity disease using hybrid systems is the Neuro-Fuzzy model. In general, ANFIS is better than other models that have been compared in this since this model gives us results with 99.9% accuracy. ANFIS technique takes the benefits from the two machine learning algorithms fuzzy logic decision-making mechanism, learning ability, and relational structure of the ANNs. Through the results obtained, it is concluded that machine learning techniques especially the hybrid Neuro-fuzzy model technique showed to be a more real sorter and can contribute to predicting the risk of excess weight in children. The results using the whole dataset suggest the use of a hybrid Neuro-fuzzy model to identify excess obesity in children with a degree of acceptable accuracy, so the model achieved a prediction accuracy of 98.33% using grid partition and using neural networks which achieved an accuracy of 98.40%, while the other models got accuracy as follows: Logistic Regression 97.50%, d-tree 92.30%, KNN 93.60%, and SVM

97.10%. Also another results were obtained when the same techniques were applied for each city alone, Jenin city results (ANFIS-Grid 99.97% , ANFIS-Cluster 98.58%, Logistic Regression 95.20% and Neural Network 97.40%), Nablus city results (ANFIS-Grid 99.97% , ANFIS-Cluster 99.81%, Logistic Regression 95.10% and Neural Network 96.90%), Hebron city results (ANFIS-Grid 99.96% , ANFIS-Cluster 98.65%, Logistic Regression 97.90% and Neural Network 97.90%) and Tulkarem city results (ANFIS-Grid 99.65% , ANFIS-Cluster 99.43%, SVM and LR 97.90% and Neural Network 97.70%).

Both types in ANFIS give very good results but in general in this case using ANFIS-grid is better from ANFIS-cluster. In ANFIS-grid there is many membership functions, in this thesis four membership functions were used: triangular, trapezoidal, Gaussian, and Gaussian combination, overall the best membership functions were the Gaussian, and Gaussian combination that gave the better results.

For future studies, it is suggested to use the hybrid model for the prediction and classification of obesity on international datasets and using other classifiers of excess obesity. It is proposed to improve the prediction and classification accuracy by using evolutionary algorithms like genetic algorithms and particle swarm optimization methods. It is suggested to carry out future comparative studies with new multi-objective algorithms to verify the accuracy improvement, with other features on the dataset. Also, many other features like how many hours are spent on TV or playing video games. Also, some other features deepening in the living activity can be added to the model. Finally, to use this model by specialists as a friendly user interface, it should be built and developed and specialists will manage the application with a shared database that the application can learn from many other cases faster and produces more accurate results. Also, these days it can easy to benefit from smartphones and smart-watches in which these devices collect a huge number of data that could be used and sent tips for each user depending on its situation.

Bibliography

[1] Bray, G. A., Frühbeck, G., Ryan, D. H., & Wilding, J. P. (2016). Management of obesity. ***The Lancet***, 387(10031), 1947-1956.

[2] Bray, G. A., Kim, K. K., Wilding, J. P. H., & World Obesity Federation. (2017). Obesity: a chronic relapsing progressive disease process. A position statement of the World Obesity Federation. ***Obesity Reviews***, 18(7), 715-723.

[3] Janesick, A. S., Shioda, T., & Blumberg, B. (2014). Transgenerational inheritance of prenatal obesogen exposure. ***Molecular and cellular endocrinology***, 398(1-2), 31-35.

[4] Marmett, B., Carvalho, R. B., Fortes, M. S., & Cazella, S. C. (2018). Artificial Intelligence technologies to manage obesity. VITTALLE-Revista de Ciências da Saúde, 30(2), 73-79.

[5] Amisha, Malik, P., Pathania, M., & Rathaur, V. (2019, July). ***Overview of artificial intelligence in medicine***. Retrieved January 02, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6691444/

[6] Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. ***Procedia computer science***, 132, 1578-1585.

[7] Tharwat, A. (2020). Classification assessment methods. Applied Computing and Informatics.

[8] Stoltzfus, J. C. (2011). Logistic regression: a brief primer. ***Academic Emergency Medicine***, 18(10), 1099-1104.

[9] DeMaris, A. (1995). A tutorial in logistic regression. ***Journal of Marriage and the Family***, 956-968.

[10] Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. ***Shanghai archives of psychiatry***, 27(2), 130.

[11] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. ***Neural processing letters***, 9(3), 293-300.

[12] Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. ***Applied Computing and Informatics***, 12(1), 90-108.

[13] Disse, E., Ledoux, S., Bétry, C., Caussy, C., Maitrepierre, C., Coupaye, M., ... & Simon, C. (2018). An artificial neural network to predict resting energy expenditure in obesity. ***Clinical Nutrition***, 37(5), 1661-1669.

[14] Baxt, W. G. (1995). Application of artificial neural networks to clinical medicine. ***The lancet***, 346(8983), 1135-1138.

[15] Karaboga, D., & Kaya, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. ***Artificial Intelligence Review***, 52(4), 2263-2293.

[16] Forno, E., & Celedón, J. C. (2017). The effect of obesity, weight gain, and weight loss on asthma inception and control. ***Current opinion in allergy and clinical immunology***, 17(2), 123.

[17] Vanderwall, C., Clark, R. R., Eickhoff, J., & Carrel, A. L. (2017). BMI is a poor predictor of adiposity in young overweight and obese children. ***BMC pediatrics***, 17(1), 1-6.

[18] Dixon, A. E., & Peters, U. (2018). The effect of obesity on lung function. ***Expert review of respiratory medicine***, 12(9), 755-767.

[19] Al-Lahham, S., Jaradat, N., Altamimi, M., Anabtawi, O., Irshid, A., AlQub, M., ... & Al Zabadi, H. (2019). Prevalence of underweight, overweight and obesity among Palestinian school-age children and the associated risk factors: a cross sectional study. ***BMC pediatrics***, 19(1), 1-13.

[20] Ejheisheh, M. A., Correa-Rodríguez, M., Fernández-Aparicio, Á., Batran, A., Membrive-Jiménez, M. J., & Schmidt-RioValle, J. (2020). Obesity-Related Parameters Are Associated With Blood Pressure in Palestinian Children. ***Biological Research For Nursing***, 1099800420941074.

[21] Iqbal, R. K., Nasrullah, F. M., & Ikram, S. (2019). ***How Obesity Affects Our Health***.

[22] Ahmad, N., Adam, S. I. M., Nawi, A. M., Hassan, M. R., & Ghazi, H. F. (2016). Abdominal obesity indicators: Waist circumference or waist-to-hip ratio in Malaysian adults population. ***International journal of preventive medicine***, 7.

[23] Sancar, N., & Tabrizi, S. S. (2017, January). ***Body Mass Index Estimation by Using an Adaptive Neuro Fuzzy Inference System***. In ICCS (pp. 2501-2506).

[24] Ahn, S. H., Wang, C., Shin, G. W., Park, D., Kang, Y. H., Joibi, J. C., & Yun, M. H. (2018, December). ***Comparison of Clustering Methods for Obesity Classification***. In 2018 IEEE

International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 1821-1825). IEEE.

[25] Marmett, B., Carvalho, R. B., Fortes, M. S., & Cazella, S. C. (2018). Artificial Intelligence technologies to manage obesity. ***VITTALLE-Revista de Ciências da Saúde***, 30(2), 73-79.

[26] Kazemipoor, M., Hajifaraji, M., Shamshirband, S., Petković, D., & Kiah, M. L. M. (2015). Appraisal of adaptive neuro-fuzzy computing technique for estimating anti-obesity properties of a medicinal plant. ***Computer methods and programs in biomedicine***, 118(1), 69-76.

[27] Ferenci, T., & Kovacs, L. (2018). Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks. ***Applied Soft Computing***, 67, 834-839.

[28] Abdullah, F. S., Abd Manan, N. S., Ahmad, A., Wafa, S. W., Shahril, M. R., Zulaily, N., ... & Ahmed, A. (2016, August). ***Data mining techniques for classification of childhood obesity among year 6 school children***. In International Conference on Soft Computing and Data Mining (pp. 465-474). Springer, Cham.

[29] Heydari, S. T., Ayatollahi, S. M. T., & Zare, N. (2012). Comparison of artificial neural networks with logistic regression for detection of obesity. ***Journal of medical systems***, 36(4), 2449-2454.

[30] Bouharati, S., Bounechada, M., Djoudi, A., Harzallah, D., Alleg, F., & Benamrani, H. (2012). Prevention of obesity using artificial intelligence techniques. ***International Journal of Science and Engineering Investigations***, 1(9), 146-150.

[31] Tabrizi, S. S., & Sancar, N. (2017). Prediction of Body Mass Index: A comparative study of multiple linear regression, ANN and ANFIS models. ***Procedia computer science***, 120, 394-401.

[32] Hardalaç, F., Ozan, A. T., Barişçi, N., Ergün, U., Serhatlioğlu, S., & Güler, İ. (2004). The examination of the effects of obesity on a number of arteries and body mass index by using expert systems. ***Journal of medical systems***, 28(2), 129-142.

[33] Ergün, U. (2009). The classification of obesity disease in logistic regression and neural network methods. ***Journal of medical systems***, 33(1), 67.

[34] Xue, Q., Wang, X., Meehan, S., Kuang, J., Gao, J. A., & Chuah, M. C. (2018, December). ***Recurrent Neural Networks based Obesity Status Prediction Using Activity Data***. In 2018 17th

IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 865-870). IEEE.

[35] Triantafyllidis, A., Polychronidou, E., Alexiadis, A., Rocha, C. L., Oliveira, D. N., da Silva, A. S., ... & Tzovaras, D. (2020). Computerized decision support and machine learning applications for the prevention and treatment of childhood obesity: A systematic review of the literature. ***Artificial Intelligence in Medicine***, 104, 101844.

[36] Colmenarejo, G. (2020). Machine Learning Models to Predict Childhood and Adolescent Obesity: A Review. ***Nutrients***, 12(8), 2466.

[37] Stoltzfus, J. C. (2011). Logistic regression: a brief primer. ***Academic Emergency Medicine***, 18(10), 1099-1104.

[38] Sperandei, S. (2014). Understanding logistic regression analysis. Biochemia medica: ***Biochemia medica***, 24(1), 12-18.

[39] Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). ***Using multivariate statistics*** (Vol. 5, pp. 481-498). Boston, MA: Pearson.

[40] Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. ***Shanghai archives of psychiatry***, 27(2), 130.

[41] Shouman, M., Turner, T., & Stocker, R. (2011, December). Using decision tree for diagnosing heart disease patients. ***In Proceedings of the Ninth Australasian Data Mining Conference***-Volume 121 (pp. 23-30).

[42] Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. ***Cancer Genomics-Proteomics***, 15(1), 41-51.

[43] Widodo, A., & Yang, B. S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. ***Mechanical systems and signal processing***, 21(6), 2560-2574.

[44] Bazmara, M., Movahed, S. V., & Ramadhani, S. (2013). KNN Algorithm for Consulting Behavioral Disorders in Children. ***Journal of Basic and Applied Scientific Research***, 3, 12.

[45] DeGregory, K. W., Kuiper, P., DeSilvio, T., Pleuss, J. D., Miller, R., Roginski, J. W., ... & Dungan, I. (2018). A review of machine learning in obesity. ***Obesity Reviews***, 19(5), 668-685.

[46] Tiwari, S., Naresh, R., & Jha, R. (2013). Comparative study of backpropagation algorithms in neural network based identification of power system. International ***Journal of Computer Science & Information Technology***, 5(4), 93.

[47] Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. ***Neural networks***, 6(4), 525-533.

[48] Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. ***IEEE transactions on systems, man, and cybernetics***, 23(3), 665-685.

[49] Joelianto, E., & Rahmat, B. (2008). Adaptive Neuro-Fuzzy Inference System (ANFIS) with Error Backpropagation Algorithm using Mapping Function. ***International journal of artificial intelligence***, 1(A08).

[50] Tzortzis, G., & Likas, A. (2014). The MinMax k-Means clustering algorithm. ***Pattern Recognition***, 47(7), 2505-2516.

[51] Kaur, N., Sahiwal, J. K., & Kaur, N. (2012). Efficient k-means clustering algorithm using ranking method in data mining. ***International Journal of Advanced Research in Computer Engineering & Technology***, 1(3), 85-91.

[52] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. ***Computers & geosciences***, 10(2-3), 191-203.

[53] Cebeci, Z., & Yildiz, F. (2015). Comparison of k-means and fuzzy c-means algorithms on different cluster structures. ***Agrárinformatika/journal of agricultural informatics***, 6(3), 13-23.

[54] Nayak, J., Naik, B., & Behera, H. (2015). Fuzzy C-means (FCM) clustering algorithm: a decade review from 2000 to 2014. ***Computational intelligence in data mining***-volume 2, 133-149.

[55] Schrider, D. R., & Kern, A. D. (2018). Supervised machine learning for population genetics: a new paradigm. ***Trends in Genetics***, 34(4), 301-312.

[56] Poirier, P., & Eckel, R. H. (2002). Obesity and cardiovascular disease. ***Current atherosclerosis reports***, 4(6), 448-453.

[57] Akoumianakis, I., & Filippatos, T. (2020). The renin–angiotensin–aldosterone system as a link between obesity and coronavirus disease 2019 severity. ***Obesity Reviews***, 21(9), e13077.

[58] Rouhani-Kalleh, O. (2007, March). Algorithms for fast large scale data mining using logistic regression. In *2007 IEEE Symposium on Computational Intelligence and Data Mining* (pp. 155-162). IEEE.

[59] Su, J., & Zhang, H. (2006, July). A fast decision tree learning algorithm. In *AAAI* (Vol. 6, pp. 500-505).

[60] Abdiansah, A., & Wardoyo, R. (2015). Time complexity analysis of support vector machines (SVM) in LibSVM. *International journal computer and application*, *128*(3), 28-34.

الملخص

تعتبر صحة الإنسان من أهم مجالات الحياة، وحيث ان السمنة هي أحد الأسباب الرئيسية للعديد من الأمراض التي قد تسبب الوفاة مثل مرض السكري، وارتفاع ضغط الدم، والسكتة الدماغية ، فقد يتم استخدام تكنولوجيا المعلومات لتطوير العديد من جوانب الحياة، ومساعدة البشر في عملهم اليومي. وحيث يعتبر الذكاء الاصطناعي (AI) أحد المجالات المهمة والذي بدوره يقدم العديد من الطرق والخوارزميات لدعم اتخاذ القرارات السريرية، فيعد تكامل الذكاء الاصطناعي في مجال تشخيص الرعاية الصحية طريقة فعالة في عدد كبير من تطبيقات الرعاية الصحية. بالإضافة لذلك، يمكن تصنيف الإجراءات التشخيصية في مجال الرعاية الصحية على أنها مهام تنبؤية وتصنيف ذكي للبيانات من حيث إمكانية استخدام تقنيات الذكاء الاصطناعي للتنبؤ بأمراض السمنة وتصنيفها لتقديم المساعدة المناسبة للأطباء في عملية اتخاذ القرار. تُستخدم العديد من تقنيات الذكاء لتشخيص الأمراض مثل neural network, fuzzy logic, expert systems، إلخ.

ففي هذا البحث، تم استخدام تقنيات ذكاء اصطناعي مختلفة لتصنيف سمنة الطفل وتوقعاتها . ولتحقيق هدف الدراسة، تم جمع مجموعة البيانات من 4 مدن في فلسطين. حيث تمر هذه البيانات المجمعة من خلال المعالجة المسبقة للبيانات، واستخراج المتغيرات التي تؤثر بشكل كبير على سمنة الأطفال. ومن ثم تطبيق أساليب الذكاء الاصطناعي للتعرف على الأنماط المختلفة في مجموعة البيانات، فقد تم استخدام Decision-tree, k-nearest neighbor, support vector machine, logistic regression, neural network, and a hybrid adaptive neuro-fuzzy inference system الذي يجمع بين fuzzy logic and neural networks للتعرف على النمط الموجود على مجموعة البيانات وتحسين نتائج تصنيف السمنة عند الأطفال بدقة عالية. بالنسبة لهذا النموذج neuro-fuzzy hybrid ، فإن membership functions المستخدمة هي trimf و trapmf و gaussmf و gauss2mf ، حيث تم استخدام نوعين من الهياكل Neuro-fuzzy؛ grid partitioning, clustering structure. وقد وجد أن إجمالي القواعد 512 قاعدة للحصول على احتمالية اصابة الطفل بالسمنة.

وبناءً على النتائج التي تم الحصول عليها عند التطبيق باستعمال جميع البيانات لجميع المدن ، فقد حقق نظام hybrid adaptive neuro-fuzzy inference أعلى دقة تنبؤ بلغت 98.33٪ باستخدام grid partition، و تم باستعمال neural network الحصول على نسبة 98.40٪ ،

وهـي تعتبـر نتـائج جيـدة جـدًا وتسـتحق اسـتخدامها فـي تطبيـق حقيقـي لمسـاعدة المتخصصـين فـي اتخـاذ القـرارات . وفـي المقابـل، حصـلت النمـاذج الأخـرى علـى درجـات دقـة علـى النحـو التـالي: Logistic Regression 05.97٪ ، 92.30 D-tree٪، 93.60 KNN٪، 97.10 SVM٪. كمـــا تـــم الحصـــول علـــى نتـــائج أخـــرى عنـــد تطبيـــق نفـــس الأســـاليب لكـــل مدينـــة علـــى حـــدة ، نتـــائج المدينـــة 1 (99.97 ANFIS-Grid٪ ، 98.58 ANFIS-Cluster٪ ، Logistic Regression 95.20٪ Neural Network 97.40٪) ، النتائج التفصيلية الأخرى موضحة في الفصل 4.