T.C.

BAHÇEŞEHİR ÜNİVERSİTESİ

DRUG DOSAGE PLANNING OF DIABETES DISEASE: AN APPLICATION OF THE ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM (ANFIS) IN ASSISTING DRUG THERAPY

Master Thesis

Emirhan Gülçin YILDIRIM

İSTANBUL, 2009

T.C.

BAHÇEŞEHİR ÜNİVERSİTESİ

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

COMPUTER ENGINEERING

DRUG DOSAGE PLANNING OF DIABETES DISEASE: AN APPLICATION OF THE ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM (ANFIS) IN ASSISTING DRUG THERAPY

Master Thesis

Emirhan Gülçin YILDIRIM

Supervisor: ASSOC. PROF. DR. ADEM KARAHOCA

İSTANBUL, 2009

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ
The Graduate School of Natural and Applied Sciences
Computer Engineering

Title of Master Thesis	: Drug Dosage Planning of Diabetes Disease: An Application of the Adaptive Network Based Fuzzy Inference System (ANFIS) in Assisting Drug Therapy
Name/Last Name of the Student	: Emirhan Gülçin YILDIRIM
Date of Thesis Defense	: 01 September 2009

The thesis has been approved by the Graduate School of Natural and Applied Sciences.

Signature

Prof. Dr. A. Bülent ÖZGÜLER
Director

This is to certify that we have read this thesis and that we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Science.

Examining Committee Members	Signature
Assoc. Prof. Dr. Adem KARAHOCA (Supervisor)	----------
Prof. Dr. Nizamettin AYDIN	----------
Assist. Prof. Dr Yalçın ÇEKİÇ	----------

ACKNOWLEDGEMENTS

I dedicate this thesis to my father who is my biggest idol in my life. Due to the contributions and support to my thesis thanks to Dr. Kazım YILDIRIM who is one of the best doctor in the world and he is my father.

I would like to express my gratitude to my supervisor Assoc. Prof. Dr. Adem KARAHOCA. Without his guidance and persistent help this thesis would not have been possible.

Also I thank to my mother because she always encouraged and love me.

Finally I would like to thank to my spouse for his patience.

ABSTRACT

DRUG DOSAGE PLANNING OF DIABETES DISEASE: AN APPLICATION OF THE ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM (ANFIS) IN ASSISTING DRUG THERAPY

Yıldırım, Emirhan Gülçin

M.S. Department of Computer Engineering
Supervisor: Assoc. Prof. Dr. Adem Karahoca

September 2009, 63 pages

Diabetes Mellitus is one of the world's most widespread diseases. During treatment diabetic patients should be patient and careful. For diabetics, continuous medical supervision and regularly use drugs are required to continue their lives as healthy people. Usually the patients' age, body mass index, genetic condition, blood and urine test results are used to decide to the appropriate drug doses by doctors.

The aim of this study is, drug dosage planning for diabetic patients using data mining techniques. In this study, ANFIS and Rough Set (RSES) data mining methods are used. Input parameters consisted of gender, age, body mass index, genetics, insulin, fast blood glucose, urine, creatinine, cholesterol, triglyceride, HDL cholesterol, LDL cholesterol, uric acid, glucose in urine, acetone, diabetes type and dose of drugs. Dosage planning made drugs are Glifix, Diamicron, Glucophage, Humilin_M, Novorapid and Insulin Lantus.

As a result of this study, ANFIS give better results than RSES is observed. ANFIS is the most effective data mining technique for dosage planning.

Key Words: ANFIS, Diabetes Mellitus, dosage planning, rough set, data mining.

ÖZET

ŞEKER HASTALIĞI İLAÇ DOZAJ PLANLAMA: İLAÇ TEDAVİSİNE YARDIMCI UYARLAMALI NÖRO BULANIK ÇIKARSAMA SİSTEMİ (ANFIS) UYGULAMASI

Yıldırım, Emirhan Gülçin

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Adem Karahoca

Eylül 2009, 63 sayfa

Şeker hastalığı dünyadaki en yaygın hastalıklardan birisidir. Şeker hastaları tedavileri süresince sabırlı ve dikkatli olmalılar. Bu hastaların hayatlarını sağlıklı insanlar gibi devam ettirebilmeleri için sürekli doktor kontrolünde düzenli ilaç kullanmaları gerekmektedir. Doktorlar genelde hastanın yaşına, boy kilo endeksine, genetik durumuna, kan ve idrar tahlili sonuçlarına bakarak uygun ilaç dozuna karar vermektedirler.

Bu çalışmanın amacı, veri madenciliği tekniklerini kullanarak diyabet hastaları için ilaç dozu planlamaktır. Bu çalışmada ANFIS ve Rough Set (RSES) veri madenciliği yöntemleri kullanılmıştır. Kullanılan giriş parametreleri cinsiyet, yaş, boy kilo endeksi, genetik, insulin, açlık kan şekeri, üre, kreatinin, kolesterol, trigliserit, HDL kolesterol, LDL kolesterol, ürik asit, idrarda bulunan glukoz, aseton, diyabet tipi ve ilaç dozudur. Dozaj planlaması yapılan ilaçlar Glifix, Diamicron, Glucophage, Humilin_M, Novorapid ve Insulin Lantus'tur.

Bu çalışma sonucunda ANFIS'in RSES'den daha iyi sonuçlar verdiği gözlenmiştir. Dozaj planlama yaparken, ANFIS en etkili veri madenciliği yöntemidir.

Anahtar Kelimeler: ANFIS, şeker hastalığı, dozaj planlama, rough set, veri madenciliği.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Adaptive Neuro Fuzzy Inference System	:	ANFIS
Rough Set Exploration System	:	RSES
Root Mean Square Error	:	RMSE
Body Mass Index	:	BMI
High Density Lipoprotein	:	HDL
Low Density Lipoprotein	:	LDL
Fast Blood Sugar	:	FBS
Glucose in Urine	:	GU
Diabetes Mellitus	:	DM
Acetone	:	As

1. INTRODUCTION

1.1 PROBLEM DEFINITION

Human body needs energy for activation. The carbohydrates are broken down to glucose. Glucose is the primarily importance energy source for the human body cells. Insulin is needed for glucose transport to body cells (Williams & Wilkins 2007 pg: 15). The human body needs steady stead blood glucose. These supply with insulin and glucagon hormones which produced by pancreas. Insulin hormones produced by the beta cells of the islets of langerhans and glucagon hormones are produced by the alpha cells of the islets of langerhans in the pancreas. When the blood glucose increases, beta cells are stimulated and insulin given to the blood. Insulin enables blood glucose to get in to the cells and this glucose used for energy. So blood glucose kept in a narrow range (Porte, Robert, Sherwin, & Baron 2003).

Diabetes Mellitus is a disorder caused by decreased production of insulin or by decreased ability to use insulin, for this reason glucose levels in the blood increases. Diabetes increases the risks of developing heart disease, kidney disease, blindness, nerve damage and blood vessel damage (An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease (Polat & Gunes 2006). Diabetes Mellitus is a widespread disease in the world and researchers attach importance for the diagnosing and predicting of this disease. Diabetes Mellitus is divided into two types: Type1 diabetes and Type2 diabetes. Type1 diabetes (insulin dependent diabetes, IDDM) occurs when the pancreas no longer produces any or very little insulin as a result of autoimmune destruction of beta cells. It usually develops in childhood or adolescence and affects 10percent of people with diabetes. Different from Type1, Type2 diabetes (non insulin dependent diabetes, NIDDM) occurs when the pancreas does not produce enough insulin to meet the body's needs or the insulin is not metabolized effectively. Type2 diabetes usually occurs later in life and affects 90percent of people with diabetes (Su, YANG, Hsu, & CHIU 2005).

1.1.1 Diagnosis of Diabetes Type

This research analysis has two essential types. Nowadays, diabetes patients two kind of symptoms appear. These are Type1 and Type2 Diabetes Mellitus. For separation of Type1 and Type2 diabetes mellitus checked on fast blood sugar, insulin, c_peptide, triglyceride and acetone levels. At Type1 diabetes; fast blood sugar level is over 180 mg/dl , insulin level is under 2 mcg/dl, the blood has not include c_peptide and urine has include acetone. However, Type2 diabetes; fast blood sugar level is over 125 mg/dl, insulin level is over 20 mcg/dl, the blood includes c_peptide and urine has not include acetone. Type2 diabetes patients have high triglyceride level. This is not valid for Type1 diabetes patients because at Type1 diabetes patients usually have normal triglyceride level.

1.1.2 Prediction of Medicine Each Diabetes Type

Nowadays, insulin injection is the only treatment method for type1 diabetes. Daily insulin requirement varies depending on the patients' body mass index, age, food consumption and activity level. A treatment is a factor that affects the Type2 diabetes mellitus is diet, activity, short acting insulin injection and medical such as metformin, sulphonylurea. Treatment method has various types which are changes with patients' blood glucose level, age, body mass index, genetic, food consumption and activity level.

1.1.3 Medicine Dosage Planning

Oral medicine treatment applied for Type2 diabetes mellitus. Beside this treatment, some of different treatments are proved. These are gets on a diet and make some exercises. Drug treatment starts with metformin (Glucophage) or gliclazide (Diamicron) included medicines. If the patients are not overweight or obese, patients will begin to use gliclazide, otherwise begin to use metformin. However, the blood glucose level cannot decrease metformin and gliclazide used together. After using drugs still blood glucose level cannot decrease start short acting insulin injection.

1.1.4 Insulin Injection Dosage Planning

Insulin dosage adjustments based on blood glucose levels. Nowadays, doctors generally using two type of insulin injection; regular insulin and NPH (Neutral Protamine Hagedorn) insulin. Regular insulin is rapid or short acting insulin. Rapid acting insulin starts being absorbed in 30-60 minutes, and has their peak action within 2 to 3 hours of injection in most individuals. Regular insulin is often used before eating to control the large rise of blood glucose that often occurs after a meal.

NPH insulin is slow or long acting insulin. NPH insulin being absorbed 3-4 hours after injection and have their peak action after 7-9 hours. The best act is injection at bedtime to control the morning glucose of the next day (http://www.diabetesnet.com/diabetes_treatments/ 2009).

1.2 BACKGROUND

Data mining techniques have been applying for existed comprehensive diabetic records for decades. Data mining, (knowledge discovery in databases (KDD)) is the process of analyzing data and summarizing the useful information which can be use to make predictions for future experiments. The aim of this study is diabetes patients' drug dosage planning by using ANFIS (adaptive neuro fuzzy inference system). Besides, I will notice some researches like this study.

1.2.1 Literature Survey

There have been a lot of studies about using data mining techniques with diabetes mellitus data. One of them is Data mining for the Diagnosis of Type 2 Diabetes from three Dimensional Body Surface Anthropometrical Scanning Data. This study is about investigating what the risk factors for anthropometrical data of Type 2 diabetes. Neural network, logistic regression, decision tree and rough set data mining approaches are used to predict diabetes. This study is reported that BMI (body mass index) and WHR (waist hip ratio) are massive for diabetes disease (SU, YANG, HSU, & CHIU 2005).

Another study of these is an expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. This study is about improving the diagnostic accuracy of diabetes disease combining PCA and ANFIS. There were two classes as healthy and patient. The obtained sensitivity, specificity and MSE values by PCA-ANFIS for diabetes are 85.71percent, 92.0 percent and 0.262 (Polat & Gunes 2006).

A different study is feature selection and classification model construction on Type 2 diabetic patient's data. This study defines significant factors influencing diabetes control, by applying feature selection to a working patient management system to assist with ranking, classification and knowledge discovery. There is 2064 Type 2 diabetic patients' information. To improve the computational efficiency, they used to rank the attributes. Age, diagnosis duration, insulin treatment, random blood glucose, diet treatment are the most important factors influencing blood glucose control. After that, they applied Naïve Bayes, IB1 and C4.5 classification techniques. The models provided a best predictive accuracy of 95 percent and sensitivity of 98 percent. In this study the obtained results are, IB1 is the best classification, Naïve Bayes process the data fastest and C4.5 is the most stable classifier with the highest precision and the best balance between sensitivity and specificity (Sigurdardottir, Jonsdottir & Benediktsson 2007).

Another one is a comparative study on diabetes disease diagnosis using neural networks. In this study a comparative Pima Indian diabetes disease diagnosis was realized. Pima Indian diabetes dataset is used. There are 768 samples. Multilayer neural network structure and a probabilistic neural network structure were used. They obtained 82.37 percent classification accuracy for multilayer neural network and 66.78 percent for ANFIS (Temurtas, Yumusak & Temurtas 2008).

A further study is outcomes of educational interventions in Type 2 diabetes: WEKA data mining analysis. The aim of this study is to enhance diabetes related self-care that contributes to good metabolic control which minimizes the occurrence of both acute and chronic complications. Data were analyzed with WEKA. Data mining educational content and intensity of education did not predict changes in HbA1c levels (Sigurdardottir, Jonsdottir & Benediktsson 2007).

An altered study is Data Mining Diabetic Databases: Are Rough Sets a Useful Addition? This study is about analyzing a diabetic dataset with rough sets. Rough set is applied to Pima Indian Diabetic Database. Rosetta software is used. 392 data is selected randomly. It divided into training (300) and testing (92). In this study discretization method is used best with Johnson algorithm. 82.6 percent accuracy rate is observed (Breault).

Besides all these, another study is developing an expert-system for diabetics by supporting with ANFIS. The aim of this study is arranging an expert system to help dosage planning. It used data mining with a knowledge based on diabetic patients. In this study classification and association rules data mining techniques and ANFIS applied. 390 patients' records are used as 300 for training and 90 are used for checking (Kara 2008).

The last study is prediction of cyclosporine A blood levels: an application of the adaptive-network-based fuzzy inference system (ANFIS) in assisting drug therapy. In this study therapeutic drug monitoring is observed. The aim of the study is predicting the concentration of cyclosporine A level in blood. ANFIS is used for predicting cyclosporine A level in blood. 654 TDM assays are collected from 138 patients. 473 records are used for training and 181 records are used for checking (Goren, Karahoca, Onat, & Goren 2008).

2. MATERIAL & METHODS

In this study, diabetics' data is collected from hospitals in Turkey. After the collecting data, information extracting is realized by using data segmentation process. The collected data are arranged for use in data mining techniques.

2.1 PREPARING DIABETES DATA SET

Diabetes Mellitus assays carried out in 2008 and 2009 on diabetes patients who made laboratory tests in State Hospitals were included in this study. In this research, collected blood and urine samples were analyzed in State Hospital Laboratories.

Data on the total of 318 diabetes assays were collected from 89 patients (44 Type1 DM, 45 Type 2 DM). The data collected for each assays were gender, age of patient (years), body mass index of patient, genetic (parents of patients were diabetic or not), blood insulin (mcg/dl), c_peptide, fast blood sugar (FBS)(mg/dl), urine in blood (mg/dl), creatinine (mg/dl), total cholesterol (mg/dl), triglyceride (mg/dl), high density lipoprotein (HDL)(mg/dl), low density lipoprotein (LDL)(mg/dl), uric acid (mg/dl), acetone in urine (mmol/l) and glucose in urine (mmol/l). Diabetic patients use drugs and injections to reduce blood insulin level in a narrow range.

In this research, drug dosage planning made. These drugs are Glifix, Diamicron, Glucophage, Humilin_M, Novorapid and Insulin Lantus. ANFIS algorithm is help to planning to degree of opponent medicine for both types of diabetes patients. Some basis population statistics for the training and checking sets of used medicines are shown in Table 2.1.

Table 2.1: Distribution of data use in the checking and training data sets by medicines

Drug	Checking Data (n)	Training Data (n)
Glifix	12	25
Diamicron	28	58
Glucophage	39	78
Humilin_M	12	25
Novorapid	59	120
Insulin Lantus	62	125

The mean, ranges of the parameters and standard deviation of parameters are calculated with drugs basis. Calculation formula of standard deviation is below formula 2.1.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x)} \qquad (2.1)$$

As the data in the table Glifix parameters, the maximum and minimum range and standard deviation and the average are given in Table 2.2.

Table 2.2: Glifix dosage planning parameters used for fuzzy modelling

Parameters	Mean	Range (minimum-maximum)	Standard Deviation
Age	53,97	45-82	9,97
BMI	27,9	23,15-37,6	4,11
Insulin (mcg/dl)	29,23	22,7-37,7	3,59
Fast Blood Sugar (mg/dl)	168,11	94-381	73,22
Urine (Blood)(mg/dl)	32,32	12-60	11,25
Creatinine (mg/dl)	1,55	0,6-2,65	0,46
Total Cholesterol (mg/dl)	163,05	87-301	42,57
Triglyceride (mg/dl)	156,24	118-208	24,27
HDL Cholesterol (mg/dl)	49	26-71	11,77
LDL Cholesterol (mg/dl)	94,49	25-204,5	42,08
Uric Acide (mg/dl)	6,15	4,2-8,9	1,28
Glucose in Urine (mmol/l)	0,62	0-4	1,04
Acetone in Urine (mmol/l)	0	0-0	0
Glifix Dosage (mg)	38,11	30-90	18,23

As the data in the table Diamicron parameters, the maximum and minimum range and standard deviation and the average is given in Table 2.3.

Table 2.3: Diamicron dosage planning parameters used for fuzzy modelling

Parameters	Mean	Range (minimum-maximum)	Standard Deviation
Age	56,7	40-82	10,78
BMI	27,65	21,6-40,5	3,14
Insulin (mcg/dl)	29,56	17,6-37,7	4,94
Fast Blood Sugar (mg/dl)	171,03	87-381	68,9
Urine (Blood)(mg/dl)	38,21	12-180	28,19
Creatinine (mg/dl)	1,66	0,6-4,2	0,67
Total Cholesterol (mg/dl)	194,16	117-480	63,61
Triglyceride (mg/dl)	170,28	107-320	39,55
HDL Cholesterol (mg/dl)	48,25	21,8-76	12,64
LDL Cholesterol (mg/dl)	118,38	32,8-399,4	61,48
Uric Acide (mg/dl)	5,98	3,7-9,9	1,48
Glucose in Urine (mmol/l)	0,58	0-4	0,93
Acetone in Urine (mmol/l)	0	0-0	0
Diamicron Dosage (mg)	95,23	60-120	16,71

As the data in the table Glucophage parameters, the maximum and minimum range and standard deviation and the average is given in Table 2.4.

Table 2.4: Glucophage dosage planning parameters used for fuzzy modelling

Parameters	Mean	Range (minimum-maximum)	Standard Deviation
Age	56,58	40-82	10,42
BMI	28,3	21,6-40,5	4,06
Insulin (mcg/dl)	29,11	17,6-37,7	5,05
Fast Blood Sugar (mg/dl)	171,16	87-394	68,7
Urine (Blood)(mg/dl)	36,47	12-180	24,75
Creatinine (mg/dl)	1,64	0,6-4,2	0,62
Total Cholesterol (mg/dl)	186,33	87-480	60,42
Triglyceride (mg/dl)	168,11	106-320	39,23
HDL Cholesterol (mg/dl)	48,08	21,8-78	12,73
LDL Cholesterol (mg/dl)	110,41	25-399,4	58,34
Uric Acide (mg/dl)	5,95	3,7-9,9	1,34
Glucose in Urine (mmol/l)	0,58	0-3	0,9
Acetone in Urine (mmol/l)	0	0-0	0
Glucophage Dosage (mg)	2235,04	850-2550	480,45

As the data in the table Humilin_M parameters, the maximum and minimum range and standard deviation and the average is given in Table 2.5.

Table 2.5: Humilin_M dosage planning parameters used for fuzzy modelling

Parameters	Mean	Range (minimum-maximum)	Standard Deviation
Age	62,92	45-83	12,64
BMI	29,15	24,93-35,1	3,04
Insulin (mcg/dl)	33,26	27,9-37,7	3,26
Fast Blood Sugar (mg/dl)	207,05	92-394	93,86
Urine (Blood)(mg/dl)	45,49	12-167	32,99
Creatinine (mg/dl)	1,9	0,6-4,15	0,91
Total Cholesterol (mg/dl)	185,81	142-248	28,01
Triglyceride (mg/dl)	176,95	111-302	38,42
HDL Cholesterol (mg/dl)	42,11	30-54,8	6,4
LDL Cholesterol (mg/dl)	109,39	67-157,2	24,7
Uric Acide (mg/dl)	6,14	3,7-9,9	1,66
Glucose in Urine (mmol/l)	1	0-3	1,15
Acetone in Urine (mmol/l)	0	0-0	0
Humilin_M Dosage (mg)	21,73	12-34	5,5

As the data in the table Novorapid parameters, the maximum and minimum range and standard deviation and the average is given in Table 2.6.

Table 2.6: Novorapid dosage planning parameters used for fuzzy modelling

Parameters	Mean	Range (minimum-maximum)	Standard Deviation
Age	14,49	7-80	8,88
BMI	18,40	13,04-35,1	2,54
Insulin (mcg/dl)	1,5	0,5-33,1	4,14
Fast Blood Sugar (mg/dl)	256,25	97-541	120,27
Urine (Blood)(mg/dl)	17,86	13-56	4,99
Creatinine (mg/dl)	0,92	0,6-2,8	0,27
Total Cholesterol (mg/dl)	133,65	68-177	24,24
Triglyceride (mg/dl)	133,51	91-256	20,54
HDL Cholesterol (mg/dl)	61,5	39-79	9,76
LDL Cholesterol (mg/dl)	46,28	2,6-97,2	23,91
Uric Acide (mg/dl)	3,89	2,1-5,5	0,59
Glucose in Urine (mmol/l)	1,63	0-4	1,42
Acetone in Urine (mmol/l)	0,7	0-4	1,31
Novorapid Dosage (mg)	16,85	10-22	2,86

As the data in the table Insulin Lantus parameters, the maximum and minimum range and standard deviation and the average is given in Table 2.7.

Table 2.7: Insulin Lantus dosage planning parameters used for fuzzy modelling

Parameters	Mean	Range (minimum-maximum)	Standard Deviation
Age	16,32	7-80	12,5
BMI	18,96	13,04-37,6	3,77
Insulin (mcg/dl)	2,67	0,5-33,4	6,92
Fast Blood Sugar (mg/dl)	252,12	97-541	119,67
Urine (Blood)(mg/dl)	18,75	13-56	6,57
Creatinine (mg/dl)	0,96	0,6-2,8	0,34
Total Cholesterol (mg/dl)	135,1	68-264	27,48
Triglyceride (mg/dl)	134,75	91-256	21,34
HDL Cholesterol (mg/dl)	61,17	39-79	10,02
LDL Cholesterol (mg/dl)	48,94	2,6-187	29,94
Uric Acide (mg/dl)	3,95	2,1-6,8	0,68
Glucose in Urine (mmol/l)	1,57	0-4	1,41
Acetone in Urine (mmol/l)	0,67	0-4	1,29
Insulin Lantus Dosage (mg)	15,98	8-24	3,63

To obtain the data, some algorithms can be used. ANFIS algorithm is one of them. To get the results Anfis model inputs must be entered. You can use the following table for generating ANFIS model is designed according to the data (Table 2.8). Gender can take two values male or female. For male patients gender takes 0 value, for female patients it takes 1 value. Age is segmented in 8 partition, the value ranges is shown in the table. Body mass index (BMI) is divided 5 parts. The formula of BMI is shown in formula 2.2.

$$BMI = \frac{\text{weight in kilograms}}{\text{height in meters}^2} \tag{2.2}$$

BMI divided in to five parts, between 10-18.5 under weight, 18.51-25 normal, 25.01-30 over weight, 30.01-40 obese and 40.01-70 over obese. Type 1 diabetic patients' body mass indexes are usually normal and underweight but type 2 patients' are overweight and obese in general. In general, Type 2 diabetes disease is inherited. Nowadays, according to the research, Type 1 diabetes disease has no relevance with heredity. If the patient's parents has diabetes mellitus genetic is 1, otherwise genetic is 0. Type 1, diabetes patients' blood insulin level is under 2 and type 2 diabetes mellitus patients' blood insulin level is over 20 mcg/dl. Between 2-20 mcg/dl insulin level is normal. Healthy persons and type 2 diabetic patients' blood has C_peptide in their blood. Only type 1 patients cannot have C_peptide in their blood. C_peptide is a hormone secreted with insulin in the blood. At type 1 diabetes insulin produced any or very little therefore Type 1 diabetes patients' blood have not got C_peptide in their blood. Fast blood sugar (FBS) is divided into 9 parts. Under 100 mg/dl is normal values. Between 100-125 mg/dl can be a prediabetes, more than 125 mg/dl is diabetic patient. At type1 diabetes, value of FBS is more than 180 mg/dl. The urine level is normal among 10-50 mg/dl. For overage people the acceptable range is 51-75 mg/dl, over 75 mg/dl is an attention for kidney disease. Also high level creatinine causes kidney disease. Over 2 mg/dl for creatinine is acceptable risk of kidney disease. Diabetes triggers high cholesterol and heart disease consequently cholesterol level have to be kept in a narrow range. Over 240 mg/dl is high cholesterol and a risk of heart disease. Until 200 mg/dl is normal level for cholesterol. The patient must be careful for cholesterol level between 200-240 mg/dl because this level is near high level. If the cells could not use glucose to produce energy, so burn fat cells. Then fat cells mobilise and passing through the blood. In this case, triglyceride level increases. As burning fat cells, the body needs just bit insulin. Therefore, Type2 diabetes patients have high triglyceride level. This condition is not valid for Type1 diabetes patients because at Type1 diabetes pancreas cannot produce insulin. Normal degree of Triglyceride level is under 150 mg/dl normal, between 150-199 mg/dl is near high level and over 200 mg/dl is considered high level. Over 60 mg/dl for HDL level is high and optimal condition. It considered protective against heart disease. Between HDL and LDL levels have inverse ratio. High LDL level is increased risk of heart disease. The optimal LDL level is under 100 mg/dl. The near optimal LDL level is 100-129 mg/dl, between 130-159 mg/dl is borderline high LDL level, 160-190

mg/dl is high LDL level and over 190 mg/dl is very high LDL level. High Uric acid level is associated with type 2 diabetes. Over 8 mg/dl is accepted high level for uric acid. If the insulin level is approximate to zero, acetone will appears in urine. At type 1, in the urine of patients has included acetone. High level fast blood sugar (FSB) is caused glucose in urine.

Table 2.8: Parameter Values

Parameters	Values
Gender	Male=1, Female=0
Age	0-7=1, 8-14=2, 15-25=3, 26-35=4, 36-45=5, 46-60=6, 61-70=7, >70=8
BMI	10-18.5=1, 18.51-25=2, 25.01-30=3, 30.01-40=4, 40.01-70=5
Genetic	Yes=1, No=0
Insulin (mcg/dl)	<2=1, 2-20=2, >20=3
C_peptide	Yes=1, No=0
FBS (mg/dl)	0-99=1, 100-125=2, 126-180=3, 181-220=4, 221-300=5, 301-400=6, 401-500=7, 501-600=8, >600=9
Urine (mg/dl)	10-50=1, 51-75=2, >75=3
Creatinine (mg/dl)	0,6-1,2=1, 1,21_2=2, 2,01-3=3, >3=4
Total Cholesterol (mg/dl)	0-200=1, 201-240=2, >240=3
Triglyceride (mg/dl)	0-149=1, 150-199=2, 200-500=3, >500=4
HDL Cholesterol (mg/dl)	0-40=1, 41-60=2, >60=3
LDL Cholesterol (mg/dl)	0-100=1, 101-129=2, 130-159=3, 160-190=4, >190=5
Uric Acid (mg/dl)	2-5=1, 5.01-8=2, >8=3
Acetone (in urine) (mmol/l)	Yes=1, No=0
Glucose (in urine) (mmol/l)	Yes=1, No=0
Diagnosis	Type1=0, Type2=1

Attribute Ranking applied to all datasets at WEKA. Ranked values equal to zero, the columns have been eliminated. These eliminated data sets are used with ANFIS.

The ranking results for Glifix dataset is shown in Table 2.9. For these results DM, C_peptide, As and Insulin columns are discarded.

Table 2.9: Glifix Dataset's Ranked Attributes

Ranked Value	Parameter
0.34699	LDL
0.2993	BMI
0.27814	HDL
0.26717	Age
0.24116	Creatinine
0.23466	FBS
0.21232	Gender
0.08972	Total Cholesterol
0.08298	Uric Acid
0.06462	Genetic
0.02571	Urine
0.0156	Triglyceride
0.0023	GU
0	DM
0	C_peptide
0	As
0	Insulin

The ranking results for Diamicron dataset is shown in Table 2.10. For these results DM, As and C_peptide columns are discarded.

Table2.10: Diamicron Dataset's Ranked Attributes

Ranked Value	Parameter
0.32363	Age
0.22956	Uric Acid
0.16727	FBS
0.13014	HDL
0.09346	GU
0.09167	Creatinine
0.08285	LDL
0.06835	BMI
0.04813	Cholesterol
0.04245	Urine
0.03274	Gender
0.02539	Triglyceride
0.01361	Insulin
0.00681	Genetic
0	DM
0	As
0	C_peptide

The ranking results for Glucophage dataset is shown in Table 2.11. For these results DM, C_peptide and As columns are discarded.

Table2.11: Glucophage Dataset's Ranked Attributes

Ranked Value	Parameter
0.12581	Age
0.09017	Uric Acid
0.089	BMI
0.06656	Creatinine
0.06335	HDL
0.05926	LDL
0.04228	Insulin
0.03856	FBS
0.02684	Cholesterol
0.02522	Urine
0.0249	Triglyceride
0.02245	GU
0.01017	Gender
0.00113	Genetic
0	DM
0	C_peptide
0	As

The ranking results for Humilin_M dataset is shown in Table 2.12. For these results As, DM, Insulin and C_peptide columns are discarded.

Table2.12: Humilin_M Dataset's Ranked Attributes

Ranked Value	Parameter
0.78	Creatinine
0.737	FBS
0.691	LDL
0.653	GU
0.589	Age
0.539	Cholesterol
0.53	Urine
0.434	Uric Acid
0.386	BMI
0.352	Gender
0.34	Triglyceride
0.248	HDL
0.104	Genetic
0	As
0	DM
0	Insulin
0	C_peptide

The ranking results for Novorapid dataset is shown in Table 2.13. For these results Cholesterol and LDL columns are discarded.

Table2.13: Novorapid Dataset's Ranked Attributes

Ranked Value	Parameter
0.4018	FBS
0.3174	GU
0.3064	As
0.1193	Creatinine
0.0774	Triglyceride
0.058	Age
0.0498	HDL
0.0426	Genetic
0.0425	BMI
0.0231	Uric Acid
0.0202	Insulin
0.0202	DM
0.0202	C_peptide
0.0179	Gender
0.0114	Urine
0	Cholesterol
0	LDL

The ranking results for Insulin Lantus dataset is shown in Table 2.14. For these results Urine column is discarded.

Table2.14: Insulin Lantus Dataset's Ranked Attributes

Ranked Value	Parameter
0.3205	FBS
0.2425	GU
0.242	BMI
0.236	Age
0.2019	Creatinine
0.1455	As
0.1453	C_peptide
0.1453	DM
0.1453	Insulin
0.143	Triglyceride
0.0888	LDL
0.0805	HDL
0.0701	Genetic
0.0659	Uric Acid
0.0551	Cholesterol
0.0445	Gender
0	Urine

2.2 ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS)

In this study, Fuzzy Inference System is used to evaluate the diabetes data set for planning the dosages of the medicines.

Fuzzy inference system is mapping a given input to an output using fuzzy logic. The fuzzy inference systems are used in fields such as automatic control, data classification, decision analysis, expert systems and computer vision (Guopeng & Levin 2006). The Fuzzy logic which is based on the linguistic expression is an artificial intelligence technique. The fuzzy logic approach is published by Zadeh to define the complicated systems. Adaptive Neuro Fuzzy Inference System (ANFIS) is the combination of ANN and the fuzzy logic. ANFIS is a multilayer feed forward network which uses ANN learning algorithms and fuzzy reasoning to characterize an input space to an output space (Firat & Gungor 2006). Takagi and Sugeno proposed the first systematically

fuzzy modelling. The fuzzy inference system's process consists of several components which are membership functions, fuzzy logic operators, and if-then rules (Zadeh 1965).

A fuzzy set is similar with the classical set. X is the universe and its elements are signed by x and so a fuzzy set A in X is defined as below;

$$A = \{x, \mu_{A(x)} \parallel x \in X\}$$

$\mu_{A(x)}$ is called the membership function of x in A.

The fuzzy operators are the fuzzy intersection or conjunction (AND), fuzzy union or disjunction (OR) and fuzzy complement (NOT). Fuzzy sets and fuzzy operators are the subjects and verbs of fuzzy logic.

Fuzzy inference system uses *if-then* statements and the connectors present in the rule statement are 'OR' or 'AND' to make the necessary decision rules. Simple form of *if-then* rule statement is:

if x is A and y is B then z=f(x,y)

A and B are linguistic values defined by fuzzy sets, *z=f(x,y)* is a zero or first order polynomial function (Jang JSR 1993).

In this research, Takagi and Sugeno type fuzzy if-then rules are used such that the output of each rule is a linear combination of input variables plus a constant term. The final output is the weighted average of each rule's output. ANFIS is a fuzzy rule based classifier in which the rules are learnt from examples that use a standard back propagation algorithm. Anfis uses Sugeno type fuzzy system which is a linear equation (first order Sugeno inference system) or constant coefficients (zero-order Sugeno inference system) (Shafiq, Farooq & Khayam, 2008). The first order Sugeno inference system has two rules expressed as below;

Rule1: *IF x is A_1 and y is B_1 THEN* $f_1 = p_1x + q_1y + r_1$.

Rule2: *IF x is A_2 and y is B_2 THEN* $f_2 = p_2x + q_2y + r_2$

The inputs are x and y to the node i , A_i and B_i are characterized by convenient membership functions and p_i, q_i and r_i are the consequence parameters (i = 1,2,...). The structure of first order Sugeno fuzzy inference system is shown in Figure 2.1.

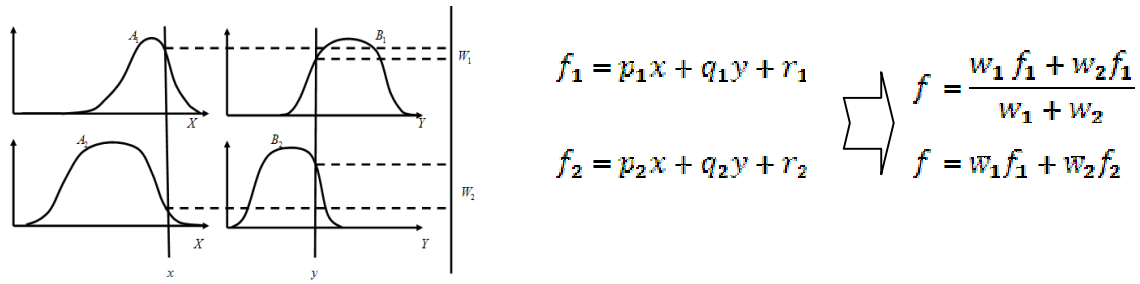

Figure 2.1: Sugeno fuzzy inference system

The ANFIS architecture is shown in Figure 2.2. The nodes of the same layer have the same functions.

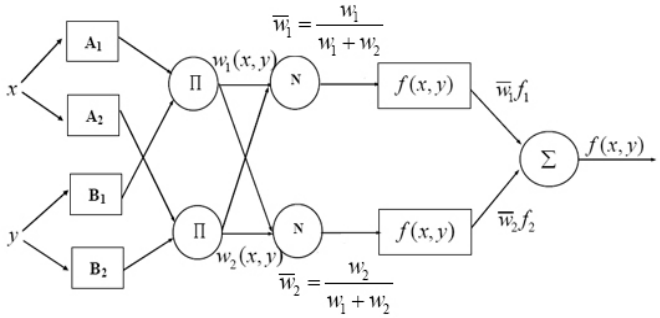

Figure 2.2: ANFIS architecture

As you can see in Figure 2.2, the ANFIS consists of five layers. The explanations of these layers are below:

Layer 0: It consists of plain input variable set.

Layer 1: In this layer every node is adaptive. *x* (or y) is the input to node i, A_i (or B_i) is linguistic label, O_i^1 is the membership function of A_i . $\{a_i, b_i, c_i\}$ is the parameter set which called premise parameters, values effects the membership function as you can see at the equations below (2.2.1):

$$O_i^1(x) = \mu A_i(x) \tag{2.3}$$

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i}\right)^2\right]^{b_i}} \tag{2.4}$$

Layer 2: Every node in this layer is a fixed node labelled Π which calculates the incoming signals and sends the product out.

$$w_i - \mu A_i(x)\mu_{B_i}(y) \tag{2.5}$$

Layer 3: Every node in this layer is a fixed node and labelled N. The i^{th} node calculates the ratio of the i^{th} rules firing strength to the sum of all rules' firing strengths. The outputs are called normalized firing strengths.

$$\overline{w}_i = \frac{w_i}{w_1 + w_2}, i = 1,2,.. \tag{2.6}$$

Layer 4: Every node in this layer is adaptive node with a node function:

$$\overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i), i = 1, 2... \tag{2.7}$$

The consequent parameter set is {p, q, r}. $\overline{w}_i$ is a normalized firing strength from the result of layer 3.

Layer 5: The single node in this layer labelled ∑ which computes the overall output as the summation of all incoming signals.

$$\sum_i \overline{w}_i f_i = \frac{\sum_i \overline{w}_i f_i}{\sum_i \overline{w}_i} \tag{2.8}$$

(Jang 1993).

2.3 ROUGH SET THEORY

The rough set theory is developed by Pawlak. It is interested in classificatory analysis of data sets. Rough Set analysis aim is to synthesize approach of concepts from the acquired data. Its principle is every object of the universe is associated some information. Because, the same information with some elements interdependent by the result (Jaafar, Jais, Hamid, Rahman, & Benaouda 2006).

From this origin, two different elements can be indiscernible in view of the available information. Information related with objects of the universe generates a relation on its elements.

Any subset of the universe is characterized by two ordinary sets. These are lower and upper approximations. The lower approximation of the target set consists of only those objects. That can positively identified as members of the set. The upper approximation includes all objects but some objects of upper approximation may not be the members of the target set. The lower approximation's objects has the probability=1 while the upper approximation's objects has the non-zero probability. (Polkowski & Skowron 1998)

The rough set occur combining of the lower and upper approximations. The Rough Set methods are applied as a component of hybrid solutions in data mining. The Rough Set data information model is collected in a table. The main purposes of analyzing data with Rough Set are data reduction, missing value handling, feature selection and feature extraction. (Magnani 2003)

3. FINDINGS

The results of this study are mentioned in this part. I applied ANFIS and Rough Set methods to the data sets. The results of ANFIS and Rough set methods are compared. There are six data sets which are Glifix, Diamicron, Glucophage, Humilin_M, Novorapid and Insulin Lantus datasets. Matlab 7.5.0 Fuzzy Toolbox is used for ANFIS method. ROSETTA software is used for Rough set algorithm. The sensitivity results of ANFIS and RSES are shown in table 3.1. The comparison of ANFIS and RSES RMSE rates are shown in table 3.2.

Table 3.1: The sensitivity results of ANFIS and RSES.

Dataset	ANFIS	RSES
Glifix	82%	80%
Diamicron	78%	70%
Glucophage	75%	67%
Humilin_M	72%	65%
Novorapid	73%	47%
Insulin Lantus	74%	60%

Table 3.2: The Comparison of ANFIS and RSES RMSE rates.

Dataset	ANFIS RMSE	RSES RMSE
Glifix	18%	50%
Diamicron	11%	10%
Glucophage	21%	61%
Humilin_M	19%	25%
Novorapid	18%	32%
Insulin Lantus	18%	50%

3.1 ANFIS RESULTS

First I will start to introduce with Glifix dataset's ANFIS results. 25 data used for training and 12 data used for checking from Glifix dataset. Training and checking data distribution are shown in figure 3.1 and 3.2. The performance of ANFIS for Glifix dataset has seen in figure 3.2. Points show actual output values and predicted output values.

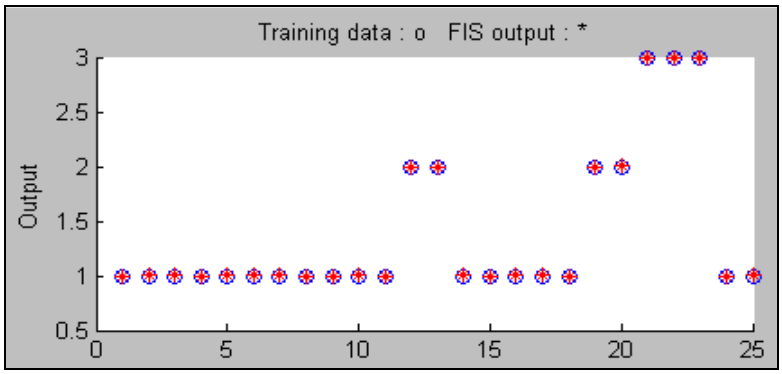

Figure 3.1: ANFIS Training data Plot

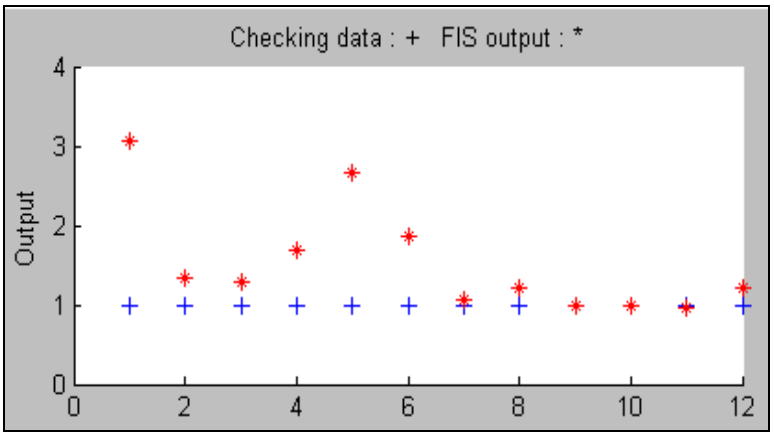

Figure 3.2: ANFIS checking data Plot

For Glifix dataset ANFIS generated 7 rules. The checking data error of Glifix is 0.1846 and the RMSE of Glifix dataset is 18 percent. The sensitivity rate of Glifix is 82 percent. The correctness of these rules is 75 percent. These rules are expressed below:

Rule 1 : [0 5 3 0 3 1 2 2 2 2 3 2 0][1]

If Gender =0 and Age=5 and Body Mass Index=3 and Genetic=0 and Fast Blood Sugar=3 and Urine=1 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acide=2 and Glucose in Urine=0 then Output is 1.

Rule 2 : [0 6 3 0 3 1 2 2 2 2 3 2 0][1]

If Gender =0 and Age=6 and Body Mass Index=3 and Genetic=0 and Fast Blood Sugar=3 and Urine=1 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acide=2 and Glucose in Urine=0 then Output is 1.

Rule 3 : [0 6 3 1 3 1 2 2 2 2 3 2 0][2]

If Gender =0 and Age=6 and Body Mass Index=3 and Genetic=1 and Fast Blood Sugar=3 and Urine=1 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acide=2 and Glucose in Urine=0 then Output is 2.

Rule 4 : [1 6 3 1 3 2 2 2 2 2 3 2 0][1]

If Gender =1 and Age=6 and Body Mass Index=3 and Genetic=1 and Fast Blood Sugar=3 and Urine=2 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acide=2 and Glucose in Urine=0 then Output is 1.

Rule 5 : [0 7 3 0 4 2 2 2 2 2 3 3 1][2]

If Gender =0 and Age=7 and Body Mass Index=3 and Genetic=0 and Fast Blood Sugar=4 and Urine=2 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acide=3 and Glucose in Urine=1 then Output is 2.

Rule 6 : [1 8 3 1 6 2 1 3 3 2 5 3 1][3]

If Gender =1 and Age=8 and Body Mass Index=3 and Genetic=1 and

Fast Blood Sugar=6 and Urine=2 and Creatinine=1 and Cholesterol=3 and Triglyceride=3 and HDL=2 and LDL=5 and Uric Acide=3 and Glucose in Urine=1 then Output is 3.

Rule 7 : [0 7 3 1 5 1 2 2 2 2 3 1 1][2]
If Gender =0 and Age=7 and Body Mass Index=3 and Genetic=1 and Fast Blood Sugar=5 and Urine=1 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acide=1 and Glucose in Urine=1 then Output is 2.

The above rules are in different situations for trained ANFIS model. Each rule is represented by a vector. It consists of input values for the system. For Glifix dataset LDL parameter's ranking ratio is 34 percent, BMI parameter's ranking ratio is 29 percent, HDL parameter's ranking ratio is 27 percent, age parameter's ranking ratio is 26 percent, creatinine parameter's ranking ratio is 24 percent, FBS parameter's ranking ratio is 23 percent, gender parameter's ranking ratio is 21 percent, cholesterol parameter's ranking ratio is 8 percent, uric acid parameter's ranking ratio is 8 percent and genetic parameter's ranking ratio is 6 percent. These parameters are more effective than the others. Also for ANFIS these parameters are used more effective than others. Descriptions of the rules which are generated by ANFIS for Glifix dataset are below.

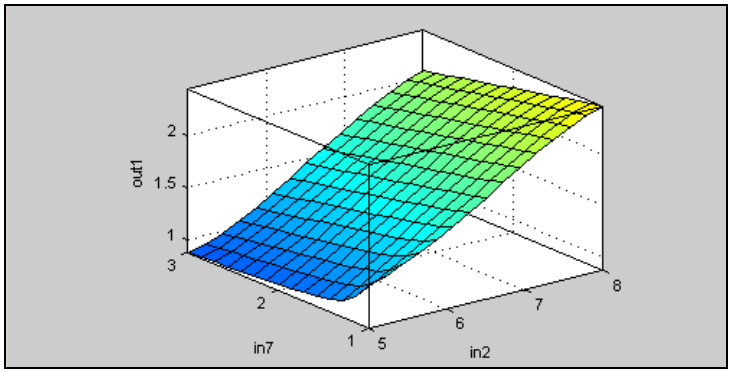

Figure 3.3: Surface Plot of FBS and Age versus Output.

In rule 1, the patient is overweight, fast blood glucose level is between 126 and 180. It has boarder line high LDL level. HDL has a critical level and triglyceride is near high level. According to these values, this patient is observed a Type 2 diabetes patient. FBS is the most decisive parameter for determining diabetes. According to the range of FBS values, the patient is Type 2 diabetes patient. Type 2 patients has high ratio of triglyceride. For this parameters patient has Type 2 diabetes. The output is 1 which means that Glifix dosage level is 30 mg. Age, genetic, gender, urine and uric acid are strongly affects the output parameter. The other parameters creatinine and cholesterol are also affects the output.

Rule 2 is similar to rule 1. Difference from first rule, this patient is over middle age. Age is one of the parameter which strongly affects the output. The other parameters are in the same class with rule 1. Only the age parameter's class changed to 6. This variation is not change the output. Output is 1 so Glifix dosage is 30 mg.

Rule 3 is similar with rule 2. The only difference is genetic changes. This patient's parents have diabetes disease. The farther parameters are in the same classes with rule 2. Genetic is one of the parameters that affect the output. In this rule genetic is positive and output is 2 so Glifix dosage is 60 mg/dl.

In Rule 4, this patient is over middle age, overweight, urine is near high level, FBS level is between 126 and 180, LDL level is borderline high level, HDL and creatinine levels are acceptable normal and triglyceride level is near high level. Genetic is positive. The patient's FBS level is not very high so drug treatment can be started at low doses according to these parameter levels. Patient has Type 2 diabetes. ANFIS generated the output 1 which means Glifix dosage is 30 mg.

In rule 5, FBS level takes place between 181- 220 mg/dl and patient's urine has glucose. Age level is over aged, genetic is negative and gender is female. According to these values patient has Type 2 diabetes. The other parameters are similar with rule 4. According to these parameters our system generates the output 2. That patient starts to use 60 mg Glifix.

In rule 6, LDL level is very high, age parameter is over aged, triglyceride and cholesterol levels are very high, gender is positive and uric acid level is high. This

patient is over aged and FBS level is very high. This patient has Type 2 diabetes. Therefore, high dose drug treatment should be started. Our system generates the output 3 that means 90 mg Glifix should be started.

In rule 7, urine, creatinine, cholesterol, triglyceride, HDL, LDL and uric acid levels can be accepted normal. Age level and BMI levels are over normal level. Genetic is positive. FBS level is high, so generated output is 2. According to output patient have to use 60 mg Glifix.

Roc curve of Glifix dataset is shown in figure 3.4:

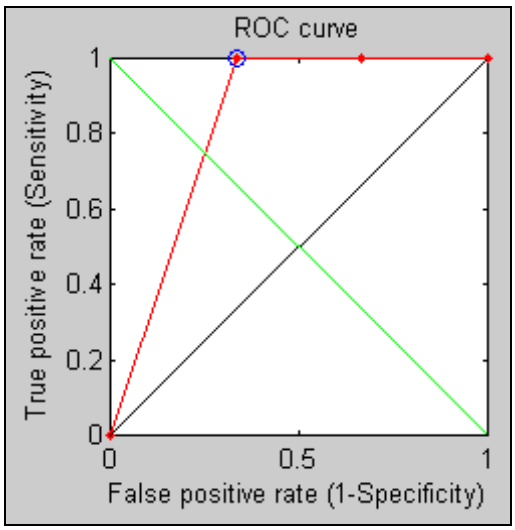

Figure 3.4: Roc curve of Glifix dataset.

ANFIS model structure of Glifix dataset is shown in figure 3.5. There are 13 inputs and one output.

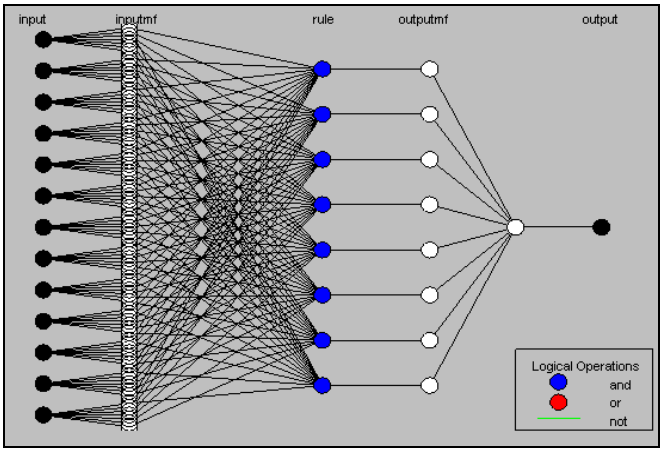

Figure 3.5: ANFIS model structure of Glifix dataset.

The other dataset is Diamicron which results I will introduce. 58 data used for training and 28 data used for checking from Diamicron dataset. Training and checking data distribution are shown in figure 3.6 and 3.7. The performance of ANFIS for Diamicron dataset has seen in figure 3.7. Points show actual output values and predicted output values.

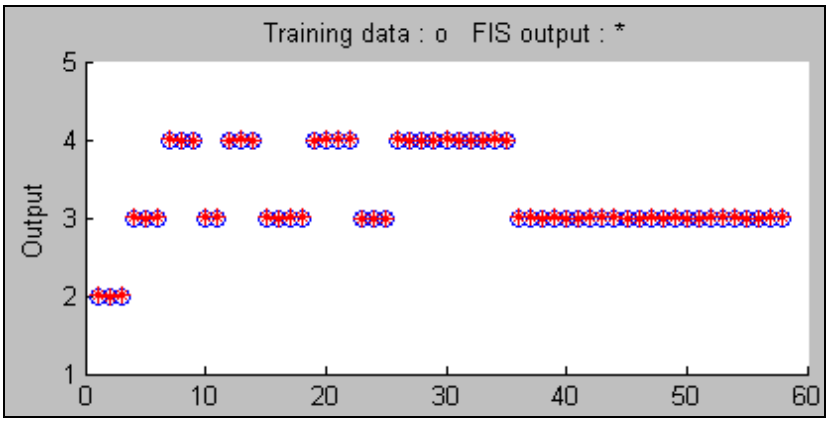

Figure 3.6: ANFIS Training data Plot

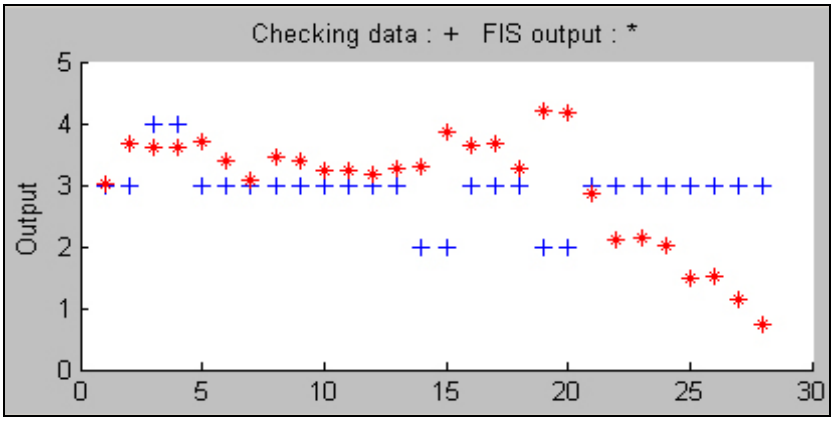

Figure 3.7: ANFIS checking data Plot

For Diamicron dataset ANFIS generates 7 rules. The checking error of Diamicron dataset is 0.1118 and the RMSE of Diamicron dataset is 11 percent. The sensitivity rate of Diamicron is 78 percent. The correctness of these rules is 71 percent. These rules are expressed below:

Rule 1 : [0 6 3 0 2 3 2 2 2 2 2 3 2 2][3]

If Gender=0 and Age=6 and Body Mass Index=3 and Genetic=0 and Insulin=2 and Fast Blood Sugar=3 and Urine=2 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=2 then Output is 3.

Rule 2 : [0 7 4 0 3 4 2 2 2 2 2 3 2 2][4]

If Gender=0 and Age=7 and Body Mass Index=4 and Genetic=0 and Insulin=3 and Fast Blood Sugar=4 and Urine=2 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=2 then Output is 4.

Rule 3 : [1 5 2 0 2 3 1 1 1 2 2 3 2 0][2]

If Gender=1 and Age=5 and Body Mass Index=2 and Genetic=0 and Insulin=2 and Fast Blood Sugar=3 and Urine=1 and Creatinine=1 and Cholesterol=1 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=0 then Output is 2.

Rule 4 : [0 5 4 1 3 4 2 3 2 2 2 4 2 1][3]

If Gender=0 and Age=5 and Body Mass Index=4 and Genetic=1 and Insulin=3 and Fast Blood Sugar=4 and Urine=2 and Creatinine=3 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=4 and Uric Acid=2 and Glucose in Urine=1 then Output is 3.

Rule 5 : [1 7 3 1 3 4 2 4 2 1 2 3 2 1][3]

If Gender=1 and Age=7 and Body Mass Index=3 and Genetic=1 and Insulin=3 and Fast Blood Sugar=4 and Urine=2 and Creatinine=4 and Cholesterol=2 and Triglyceride=1 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=1 then Output is 3.

Rule 6 : [1 8 4 1 3 6 3 2 3 3 3 5 3 4][4]

If Gender=1 and Age=8 and Body Mass Index=4 and Genetic=1 and

Insulin=3 and Fast Blood Sugar=6 and Urine=3 and Creatinine=2 and Cholesterol=3 and Triglyceride=3 and HDL=3 and LDL=5 and Uric Acid=3 and Glucose in Urine=4 then Output is 4.

Rule 7 : [0 6 3 1 3 5 2 2 3 3 1 4 3 2][3]
If Gender=0 and Age=6 and Body Mass Index=3 and Genetic=1 and Insulin=3 and Fast Blood Sugar=5 and Urine=2 and Creatinine=2 and Cholesterol=3 and Triglyceride=3 and HDL=1 and LDL=4 and Uric Acid=3 and Glucose in Urine=2 then Output is 3.

Like Glifix dataset above rules are in different situations for trained ANFIS model. Each rule represented by a vector. It consists of input values for the system. For Diamicron dataset age parameter's ranking ratio is 32 percent, uric acid parameter's ranking ratio is 22 percent, FBS parameter's ranking ratio is 16 percent, HDL parameter's ranking ratio is 13 percent and glucose in urine parameter has 9 percent ranking ratio. These parameters are more effective than the others. Descriptions of the rules which are generated by ANFIS for Diamicron dataset are below.

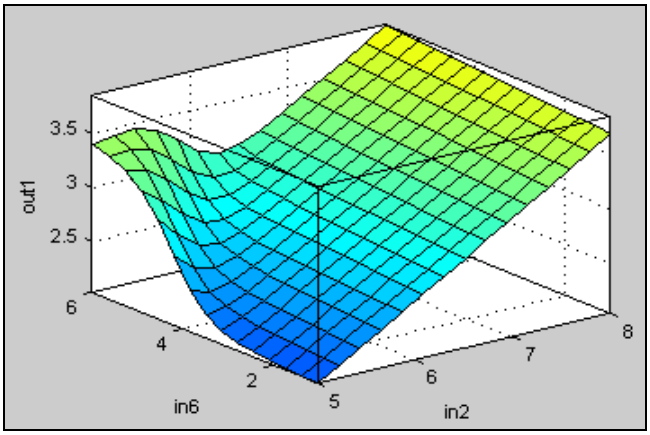

Figure 3.8: Surface Plot of FBS and Age versus Output.

Rule 1, the patient's age is later middle age. Uric acid, HDL, LDL, creatinine, cholesterol, triglyceride and urine parameters' levels can be accepted normal. Insulin level is in normal range and genetic is negative. FBS parameter's level takes place

between 126 and 180. The patient has glucose in her urine. For these values patient has Type 2 diabetes. According to these parameters, the generated output is 3. For the generated output the patient should start to use 60 mg Diamicron.

Rule 2 is similar with rule 1. The discrete parameters are Age, BMI and FBS. The patients age level states in over aged and BMI level states in obese. FBS level takes place between 181 and 220. With these parameters system generated the output 4 which means 120 mg Diamicron dosage.

Rule 3, the patient's all parameters accepted in normal ranges. Patient is middle aged and his genetic is negative. Only FBS level is high. Patient has Type 2 diabetes. ANFIS generates the output 2 according to these parameters. So Patient starts to use 60 mg Diamicron.

Rule 4, patient obese and middle aged. Her genetic is positive and her urine has glucose. Creatinine and LDL levels are high. Insulin level is more than 20 and FSB level takes place between 181 and 220. Patient has Type 2 diabetes. System generates the output 3 which is equal to 90 mg Diamicron.

Rule 5, patient is over aged and overweight. Creatinine level is high. Genetic parameter is positive. Insulin level is more than 20 mcg/dl. Fast blood glucose level is high and patient's urine has glucose. For these values patient has Type 2 diabetes. The generated output is 3 that correspond to 90 mg Diamicron.

Rule 6, urine, LDL, cholesterol, triglyceride and uric acid parameters have very high levels. The patient is over aged and his genetic parameter is positive. The patient is obese and has glucose in his urine. Insulin level is above the normal level, fast blood glucose level is very high and patient has Type 2 diabetes so generated output is 4 which corresponds 120 mg Diamicron.

Rule 7, the patient's age is later middle age and overweight. LDL, cholesterol, triglyceride and uric acid parameters have high levels. HDL level is low, genetic is positive, insulin level is above the normal level and fast blood glucose level is high. According to these values patient has Type 2 diabetes. System generates the output 3 which is equal to 90 mg Diamicron.

Roc Curve of Diamicron dataset is shown in figure 3.9:

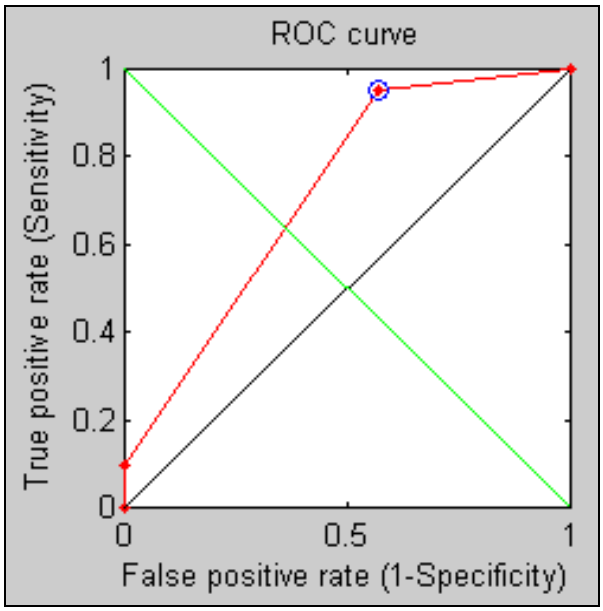

Figure 3.9: Roc curve of Diamicron dataset.

ANFIS model structure of Diamicron dataset is shown in figure 3.10. There are 14 inputs and one output.

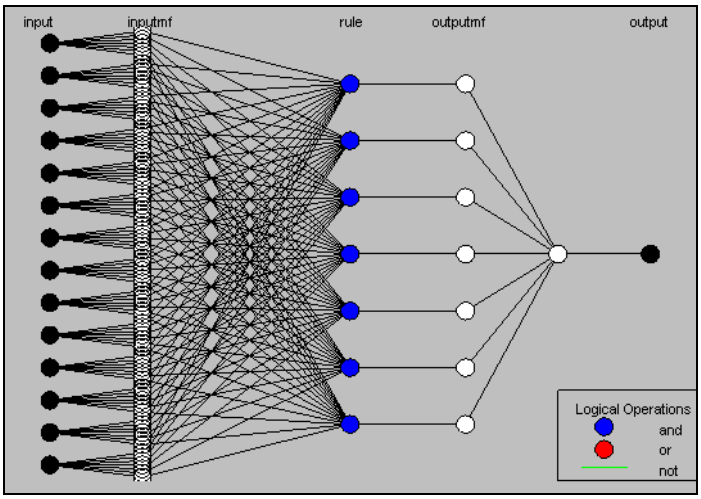

Figure 3.10: Anfis model structure of Diamicron dataset.

Another dataset is Glucophage which results I will introduce. 78 data used for training and 39 data used for checking from Glucophage dataset. Training and checking data distribution are shown in figure 3.11 and 3.12. The performance of ANFIS for

Glucophage dataset has seen in figure 3.12. Points show actual output values and predicted output values.

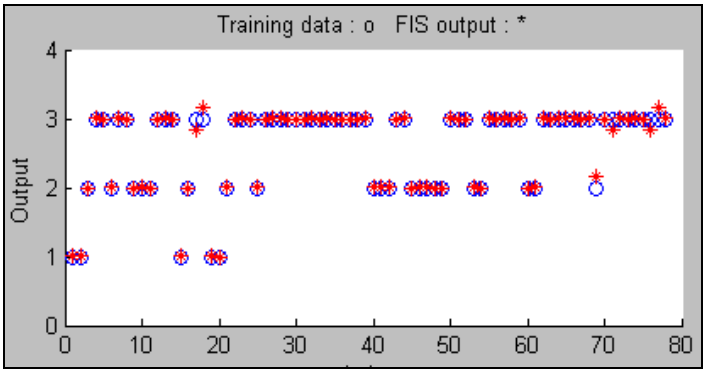

Figure 3.11: ANFIS Training data Plot

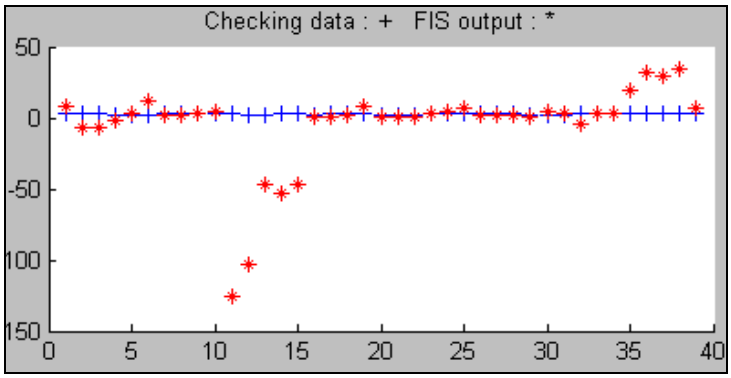

Figure 3.12: ANFIS checking data Plot

For Glucophage dataset ANFIS generates 7 rules. The checking data error of Glucophage is 0.2163 and the RMSE of Glucophage dataset is 21 percent. The sensitivity rate of Glucophage is 75 percent. The correctness of these rules is 72 percent. These rules are expressed below:

Rule 1 : [0 5 4 0 3 3 1 2 2 2 2 3 2 1][4]

If Gender=0 and Age=5 and Body Mass Index=4 and Genetic=0 and

Insulin=3 and Fast Blood Sugar=3 and Urine=1 and Creatinine=2 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=1 then Output is 4.

Rule 2 : [1 6 3 0 3 3 2 3 2 2 2 3 2 1][2]

If Gender=1 and Age=6 and Body Mass Index=3 and Genetic=0 and Insulin=3 and Fast Blood Sugar=3 and Urine=2 and Creatinine=3 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=1 then Output is 2.

Rule 3 : [1 6 3 1 3 4 2 3 2 2 2 3 2 2][3]

If Gender=1 and Age=6 and Body Mass Index=3 and Genetic=1 and Insulin=3 and Fast Blood Sugar=4 and Urine=2 and Creatinine=3 and Cholesterol=2 and Triglyceride=2 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=2 then Output is 3.

Rule 4 : [0 7 4 1 3 4 2 3 2 2 3 3 2 2][4]

If Gender=0 and Age=7 and Body Mass Index=4 and Genetic=1 and Insulin=3 and Fast Blood Sugar=4 and Urine=2 and Creatinine=3 and Cholesterol=2 and Triglyceride=2 and HDL=3 and LDL=3 and Uric Acid=2 and Glucose in Urine=2 then Output is 4.

Rule 5 : [0 6 4 1 3 4 2 3 3 2 3 3 2 2][3]

If Gender=0 and Age=6 and Body Mass Index=4 and Genetic=1 and Insulin=3 and Fast Blood Sugar=4 and Urine=2 and Creatinine=3 and Cholesterol=3 and Triglyceride=2 and HDL=3 and LDL=3 and Uric Acid=2 and Glucose in Urine=2 then Output is 3.

Rule 6 : [1 5 3 1 3 4 2 3 3 2 3 3 2 3][3]

If Gender=1 and Age=5 and Body Mass Index=3 and Genetic=1 and Insulin=3 and Fast Blood Sugar=4 and Urine=2 and Creatinine=3 and Cholesterol=3 and Triglyceride=2 and HDL=3 and LDL=3 and Uric

Acid=2 and Glucose in Urine=3 then Output is 3.

Rule 7 : [1 5 3 0 3 3 2 3 3 2 3 3 2 2][1]
If Gender=1 and Age=5 and Body Mass Index=3 and Genetic=0 and Insulin=3 and Fast Blood Sugar=3 and Urine=2 and Creatinine=3 and Cholesterol=3 and Triglyceride=2 and HDL=3 and LDL=3 and Uric Acid=2 and Glucose in Urine=2 then Output is 1.

The above rules are in different situations for trained ANFIS model. Each rule is represented by a vector. It consists of input values for the system. For Glucophage dataset age parameter's ranking ratio is 12 percent, uric acid parameter's ranking ratio is 9 percent, BMI parameter's ranking ratio is 8 percent, creatinine parameter's ranking ratio is 6 percent, HDL parameter's ranking ratio is 6 percent, LDL parameter's ranking ratio is 6 percent and insulin parameter's ranking ratio is 4 percent. These parameters are more effective than the others. For ANFIS gender, age, BMI, genetic, insulin, FBS, creatinine parameters are used more effective than others. Descriptions of the rules which are generated by ANFIS for Glucophage dataset are below.

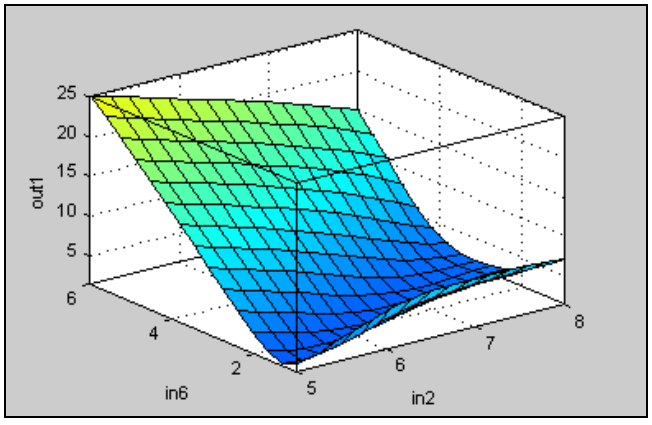

Figure 3.13: Surface Plot of FBS and Age versus Output.

Rule 1, patient is middle age and obese. Genetic is negative and LDL level is near high level. Insulin level is more than normal range, FBS level is high and she has a little glucose in her urine. Other parameters' levels are acceptable normal ranges. According

to this parameter levels patient is Type 2 diabetes and output is 4 which corresponds 3400 mg Glucophage dosage.

Rule 2 is similar with rule 1. As distinct from rule 1 patient is later middle age, over weight and creatinine level is high. The generated output is 2 which equal to 1700 mg Glucophage dosage.

Rule 3 is similar with rule 2. Differences from rule 2 are the increase in FBS level and glucose in urine and genetic is positive. ANFIS generates the output 3 which equal to 2550 mg Glucophage dosage.

Rule 4, patient is over aged and obese. Genetic is positive and insulin level is more than normal range. FBS level is high. Creatinine level is high and other parameters are acceptable normal ranges. These parameters levels denote that patient has Type 2 diabetes. Generated output is 4 which corresponds 3400 mg Glucophage dosage.

Rule 5 is similar with rule 4. As distinct from rule 4 is the increase in cholesterol level. Patient is later middle age. Other parameters' ranges are like rule 4. The generated output is 3 which equal to 2550 mg Glucophage dosage.

Rule 6, patient is middle age and overweight. Genetic is positive and insulin level is more than normal range. Creatinine and cholesterol levels are high. Patient has high level glucose in his urine and FBS level is high. According to these parameter levels, patient has Type 2 diabetes. Generated output is 3 which correspond to 2550 mg Glucophage dosage.

Rule 7, patient is middle age and overweight. Genetic is negative and insulin level is more than normal range. FBS level takes place between 126 and 180. Creatinine, LDL and cholesterol levels are high. Patient has glucose in his urine. Generated output is 1 which correspond 850 mg Glucophage dosage.

Roc curve of Glucophage dataset is shown in figure 3.14:

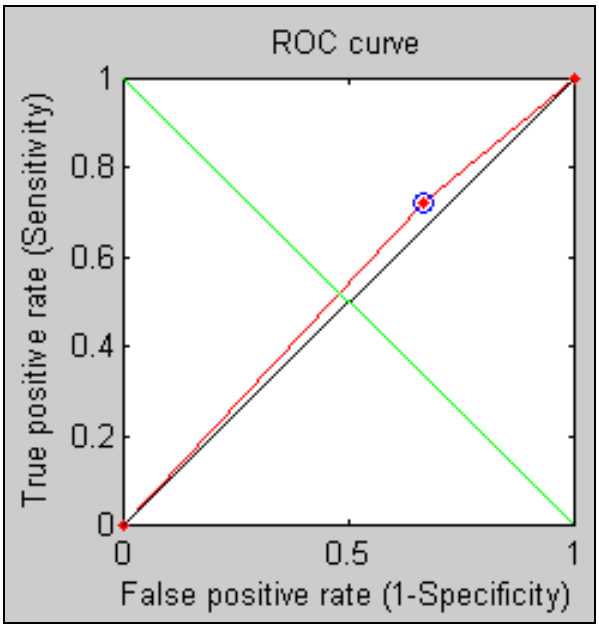

Figure 3.14: Roc curve of Glucophage dataset.

ANFIS model structure of Glucophage dataset is shown in figure 3.15. There are 14 inputs and one output.

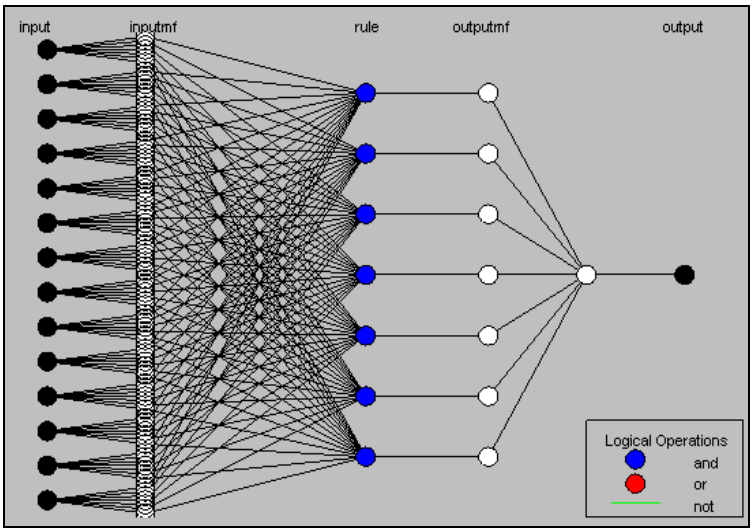

Figure 3.15: ANFIS model structure of Glucophage dataset.

The other dataset is Humilin_M which results I will introduce. 25 data used for training and 12 data used for checking from Humilin_M dataset. Training and checking data distribution are shown in figure 3.16 and 3.17. The performance of ANFIS for

Humilin_M dataset has seen in figure 3.17. Points show actual output values and predicted output values.

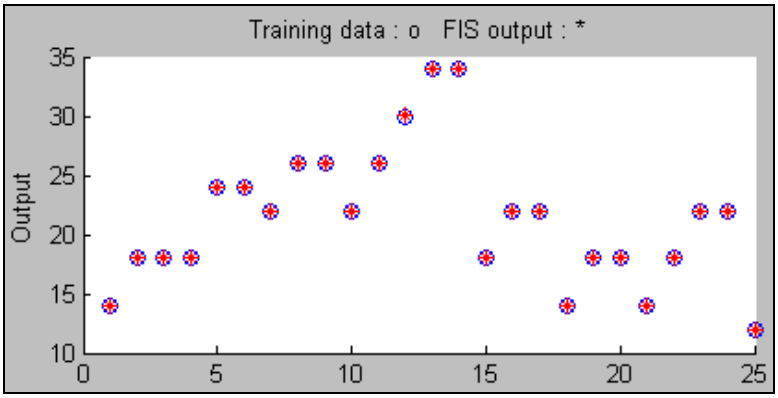

Figure 3.16: ANFIS Training data Plot

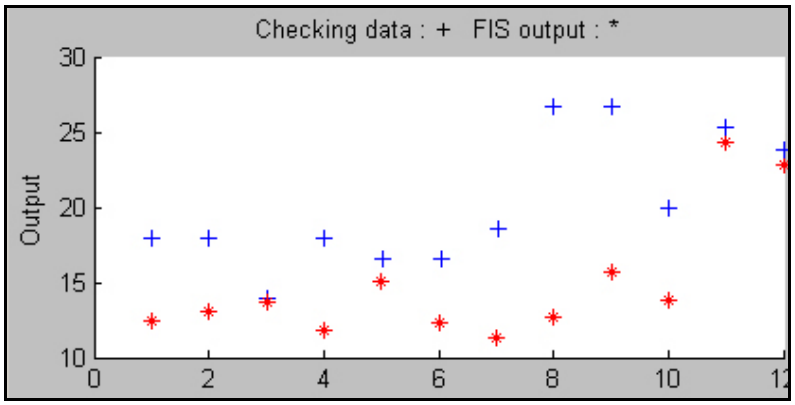

Figure 3.17: ANFIS checking data Plot

For Humilin_M dataset ANFIS generates 6 rules. The checking data error of Humilin_M is 0.1991 and the RMSE of Humilin_M dataset is 19 percent. The sensitivity rate of Humilin_M is 72 percent. The correctness of these rules is 60 percent. These rules are expressed below:

Rule 1 : [0 6 3 0 3 2 2 2 2 1 2 2 1][20]

If Gender=0 and Age=6 and Body Mass Index=3 and Genetic=0 and Fast Blood Glucose=3 and Urine=2 and Creatinine=2 and Cholesterol=2 and

Triglyceride=2 and HDL=1 and LDL=2 and Uric Acid=2 and Glucose in Urine=1 then Output is 20.

Rule 2 : [0 5 4 1 5 2 3 2 3 1 3 3 2][14]
If Gender=0 and Age=5 and Body Mass Index=4 and Genetic=1 and Fast Blood Glucose=5 and Urine=2 and Creatinine=3 and Cholesterol=2 and Triglyceride=3 and HDL=1 and LDL=3 and Uric Acid=3 and Glucose in Urine=2 then Output is 14.

Rule 3 : [1 7 3 1 6 3 4 3 3 2 3 2 1][18]
If Gender=1 and Age=7 and Body Mass Index=3 and Genetic=1 and Fast Blood Glucose=6 and Urine=3 and Creatinine=4 and Cholesterol=3 and Triglyceride=3 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=1 then Output is 18.

Rule 4 : [0 6 3 1 5 3 4 3 3 2 3 2 2][16]
If Gender=0 and Age=7 and Body Mass Index=3 and Genetic=1 and Fast Blood Glucose=5 and Urine=3 and Creatinine=4 and Cholesterol=3 and Triglyceride=3 and HDL=2 and LDL=3 and Uric Acid=2 and Glucose in Urine=2 then Output is 16.

Rule 5 : [0 5 3 1 4 2 3 2 3 2 2 1 2][12]
If Gender=0 and Age=5 and Body Mass Index=3 and Genetic=1 and Fast Blood Glucose=4 and Urine=2 and Creatinine=3 and Cholesterol=2 and Triglyceride=3 and HDL=2 and LDL=2 and Uric Acid=1 and Glucose in Urine=2 then Output is 12.

Rule 6 : [1 8 4 1 5 3 3 2 3 2 2 2 2][26]
If Gender=1 and Age=8 and Body Mass Index=4 and Genetic=1 and Fast Blood Glucose=5 and Urine=3 and Creatinine=3 and Cholesterol=2 and Triglyceride=3 and HDL=2 and LDL=2 and Uric Acid=2 and Glucose in

Urine=2 then Output is 26.

The above rules are in different situations for trained ANFIS model. Each rule is represented by a vector. It consists of input values for the system. For Humilin_M dataset creatinine parameter's ranking ratio is 78 percent, FBS parameter's ranking ratio is 73 percent, LDL parameter's ranking ratio is 69 percent, glucose in urine parameter's ranking ratio is 65 percent, Age parameter's ranking ratio is 58 percent, cholesterol parameter's ranking ratio is 53 percent and urine parameter's ranking ratio is 53 percent. These parameters are more effective than the others. For ANFIS gender, age, BMI, genetic, FBS, creatinine, HDL and glucose in urine parameters are used more effective than others. Descriptions of the rules which are generated by ANFIS for Humilin_M dataset are below.

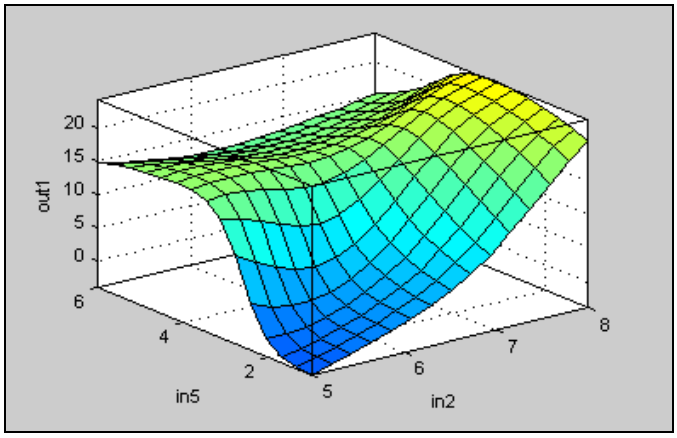

Figure 3.18: Surface Plot of FBS and Age versus Output.

Rule 1, patient is overweight and later middle age. Genetic is negative, FBS level is high, HDL level is low and has glucose in her urine. Other parameters are acceptable normal ranges. For these parameters, patient has Type 2 diabetes and generated output is 20 IU Humilin_M dosage.

Rule 2, patient is middle age and obese. Genetic is positive and patient has glucose in her urine. Creatinine, LDL, triglyceride, uric acid and FBS levels are high. Patient is Type 2 diabetic and generated output is 14 IU Humilin_M dosage.

Rule 3, patient is over aged, overweight and has glucose in his urine. Genetic is positive and FBS level is very high. Urine, creatinine, cholesterol, triglyceride, and LDL levels are high. Patient is Type 2 diabetic and generated output is 18 IU Humilin_M dosage.

Rule 4 is similar with rule 3. Differences from rule 3 are FBS level is high but not very high like rule 3, patient has further glucose in her urine and slightly younger than rule 3. Patient is Type 2 diabetic and generated output is 16 IU Humilin_M dosage.

Rule 5, patient is middle age and overweight. Genetic is positive. Creatinine, triglyceride and FBS levels are high. Other parameters' levels are acceptable normal ranges. Patient has glucose in her urine and type 2 diabetics. The generated output is 12 IU Humilin_M dosage.

Rule 6, patient is very old and obese. Genetic is positive and FBS level is very high. Patient has glucose in his urine. Creatinine, urine and triglyceride levels are high. Other parameters' levels are acceptable in normal ranges. Patient is Type 2 diabetic and generated output is 26 IU Humilin_M dosage.

Roc curve of Humilin_M dataset is shown in figure 3.19.

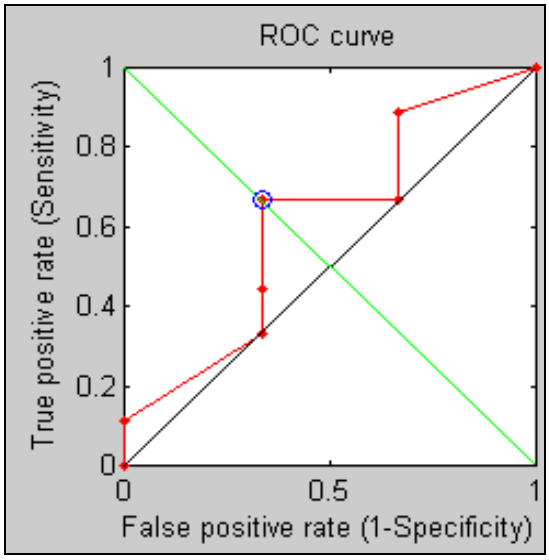

Figure 3.19: Roc curve of Humilin_M dataset.

ANFIS model structure of Humilin_M dataset is shown in figure 3.20. There are 13 inputs and 1 output for Humilin_M data.

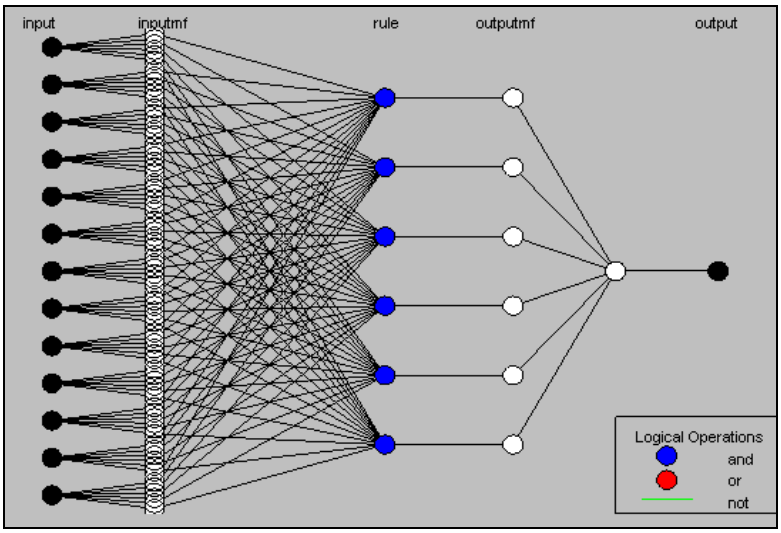

Figure 3.20: Anfis model structure of Humilin_M dataset.

Another dataset is Novorapid which results I will introduce. 120 data used for training and 59 data used for checking from Novorapid dataset. Training and checking data distribution are shown in figure 3.21 and 3.22. The performance of ANFIS for Novorapid dataset has seen in figure 3.22. Points show actual output values and predicted output values.

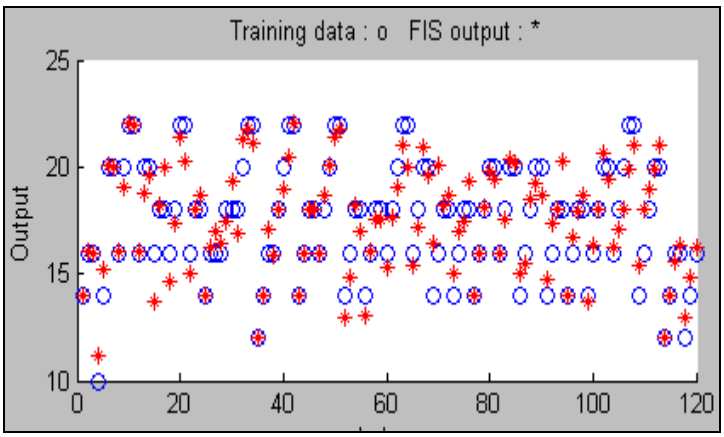

Figure 3.21: ANFIS Training data Plot.

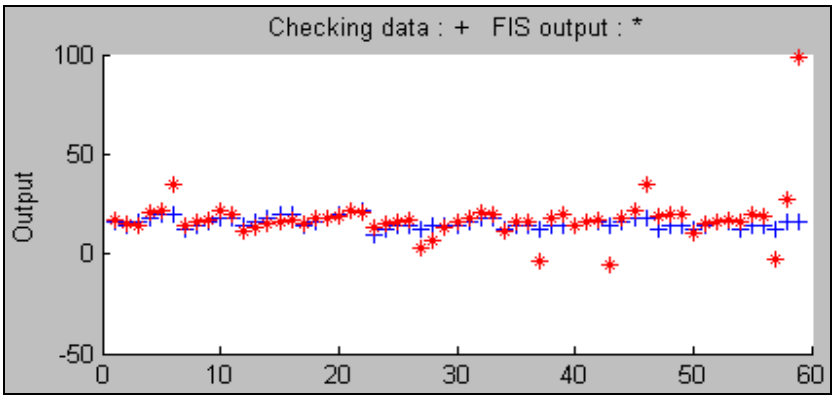

Figure 3.22: ANFIS checking data Plot.

For Novorapid dataset ANFIS generates 8 rules. The checking data error of Novorapid is 0.1812 and the RMSE of Novorapid dataset is 18 percent. The sensitivity rate of Novorapid is 73 percent. The correctness of these rules is 84 percent. These rules are expressed below:

Rule 1 : [0 2 1 0 1 0 7 1 2 2 2 1 2 1 0][8]
If Gender=0 and Age=2 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=7 and Urine=1 and Creatinine=2 and Triglyceride=2 and HDL=2 and Uric Acid=1 and Glucose in Urine=2 and Acetone=1 and DM=0 then Output=8.

Rule 2 : [0 1 1 0 1 0 7 1 1 1 2 1 4 1 0][18]
If Gender=0 and Age=1 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=7 and Urine=1 and Creatinine=1 and Triglyceride=1 and HDL=2 and Uric Acid=1 and Glucose in Urine=4 and Acetone=1 and DM=0 then Output=18.

Rule 3 : [1 1 1 0 1 0 7 1 2 1 2 1 2 1 0][16]
If Gender=1 and Age=1 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=7 and Urine=1 and Creatinine=2 and Triglyceride=1 and HDL=2 and Uric Acid=1 and

Glucose in Urine=2 and Acetone=1 and DM=0 then Output=16.

Rule 4 : [1 2 1 0 1 0 6 1 1 1 1 1 2 1 0][12]

If Gender=1 and Age=2 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=6 and Urine=1 and Creatinine=1 and Triglyceride=1 and HDL=1 and Uric Acid=1 and Glucose in Urine=2 and Acetone=1 and DM=0 then Output=12.

Rule 5 : [1 2 1 0 1 0 7 1 2 1 2 1 4 1 0][14]

If Gender=1 and Age=2 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=7 and Urine=1 and Creatinine=2 and Triglyceride=1 and HDL=2 and Uric Acid=1 and Glucose in Urine=4 and Acetone=1 and DM=0 then Output=14.

Rule 6 : [1 4 3 0 3 1 6 1 2 2 2 1 4 0 1][10]

If Gender=1 and Age=4 and Body Mass Index=3 and Genetic=0 and Insulin=3 and C_peptide=1 and Fast Blood Sugar=6 and Urine=1 and Creatinine=2 and Triglyceride=2 and HDL=2 and Uric Acid=1 and Glucose in Urine=4 and Acetone=0 and DM=1 then Output=10.

Rule 7 : [0 2 2 0 1 0 8 1 1 1 2 2 4 1 0][16]

If Gender=0 and Age=2 and Body Mass Index=2 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=8 and Urine=1 and Creatinine=1 and Triglyceride=1 and HDL=2 and Uric Acid=2 and Glucose in Urine=4 and Acetone=1 and DM=0 then Output=16.

Rule 8 : [0 1 1 0 1 0 8 1 1 2 2 2 4 1 0][18]

If Gender=0 and Age=1 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=8 and Urine=1 and Creatinine=1 and Triglyceride=2 and HDL=2 and Uric Acid=2 and Glucose in Urine=4 and Acetone=1 and DM=0 then Output=18.

The above rules are in different situations for trained ANFIS model. Each rule is represented by a vector. It consists of input values for the system. For Novorapid dataset gender parameter's ranking ratio is 40 percent, glucose in urine parameter's ranking ratio is 31 percent, Acetone parameter's ranking ratio is 30 percent, creatinine parameter's ranking ratio is 11 percent, triglyceride parameter's ranking ratio is 7 percent, age parameter's ranking ratio is 5 percent and HDL, genetic and BMI parameters' ranking ratio is 4 percent. These parameters are more effective than the others. For ANFIS gender, age, BMI, genetic, c_peptide, FBS, urine, uric acid, acetone and diabetes type are used more effective than others. Descriptions of the rules which are generated by ANFIS for Novorapid dataset are below.

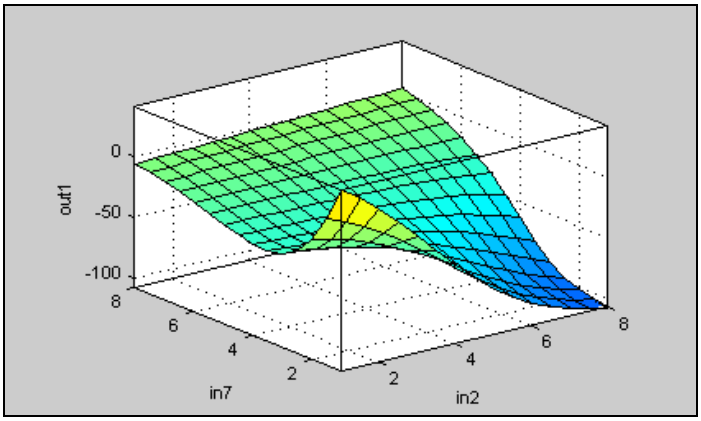

Figure 3.23: Surface Plot of FBS and Age versus Output.

Rule 1, patient is underweight whose age is between 8-14 years old. Genetic and C_peptide are negative and insulin level is under normal range. Patient has acetone and glucose in her urine. FBS level is very high and other parameters' levels are acceptable normal ranges. Patient is a Type 1 diabetic. According to these parameter levels ANFIS generates the output 8 IU Novorapid dosages.

Rule 2 is similar with rule 1. Differences from rule 1 are glucose in urine level is higher than rule 1, patient is younger than rule 1 and triglyceride level is low. Other parameters are in the same ranges with rule 1. The generated output is 18 IU Novorapid dosage.

Rule 3, patient is underweight whose age is between 0-7 years old. Genetic and C_peptide are negative. Patient has glucose and acetone in his urine. Triglyceride level is low. FBS level is between 401 and 500 mg/dl and other parameters' levels are acceptable in normal ranges. Patient is Type 1 diabetic and generated output is 16 IU Novorapid dosage.

Rule 4, patient is underweight whose age is between 8-14 years old. Genetic and C_peptide are negative and insulin level is under normal range. Patient has acetone and glucose in her urine. Triglyceride level is low. FBS level takes place between 301-400 mg/dl. Patient is Type 1 diabetic and generated output is 12 IU Novorapid dosage.

Rule 5 is similar with rule 4. Differences with rule 4 are FBS, creatinine, HDL glucose in urine parameters' levels higher than rule 4. Triglyceride level is low. The other parameters' levels are in the same ranges with rule 4. The generated output is 14 IU Novorapid dosage.

Rule 6, patient is overweight whose age is between 26-35 years old. Genetic is negative and c_peptide is positive. Insulin level is more than normal range. The patient's urine has glucose but do not have acetone. FBS level is between 301 and 400. Other parameters' levels are acceptable normal ranges. Patient is Type 2 diabetic and the generated output is 10 IU Novorapid dosage.

Rule 7, patient is normal weight whose age is between 8-14 years old. Genetic and c_peptide are negative. Insulin level is under normal range. Patient's urine have glucose and acetone. Triglyceride level is low and FBS level is between 501 and 600. The patient is Type 1 diabetic and generated output is 16 IU Novorapid dosage.

Rule 8 is similar with rule 7. Differences with rule 7 are patient's age is younger and underweight. Other parameters' levels are the same ranges with rule 7. The generated output is 18 IU Novorapid dosage.

Roc curve of Novorapid dataset is shown in figure 3.24.

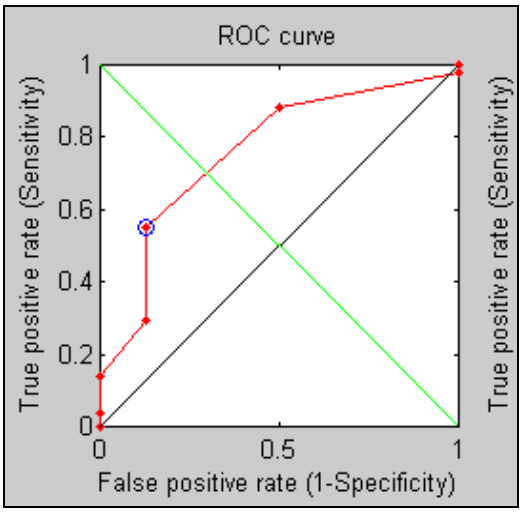

Figure 3.24: Roc curve of Novorapid dataset.

ANFIS model structure of Novorapid dataset is shown in figure 3.25. There are 15 inputs and 1 output.

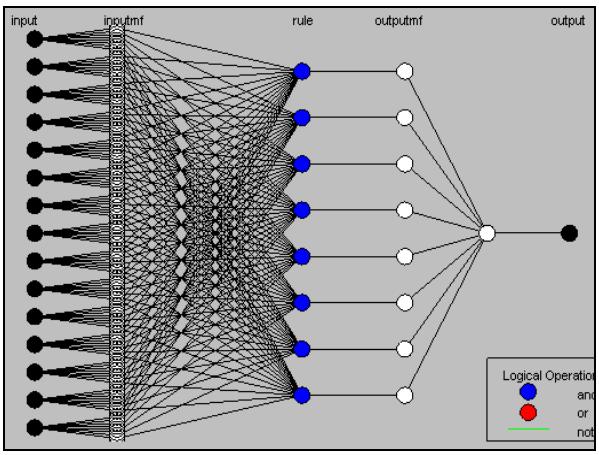

Figure 3.25: Anfis model structure of Novorapid dataset.

The other dataset is Insulin Lantus which results I will introduce. 125 data used for training and 62 data used for checking from Glifix dataset. Training and checking data distribution are shown in figure 3.26 and 3.27. The performance of ANFIS for Insulin Lantus dataset has seen in figure 3.27 Points show actual output values and predicted output values.

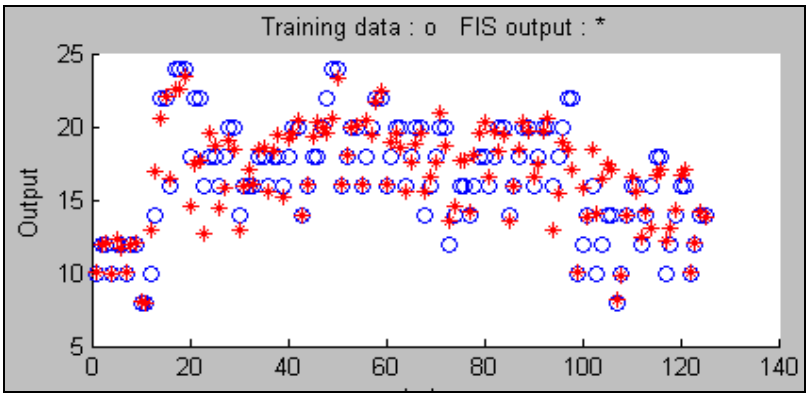

Figure 3.26: ANFIS Training data Plot.

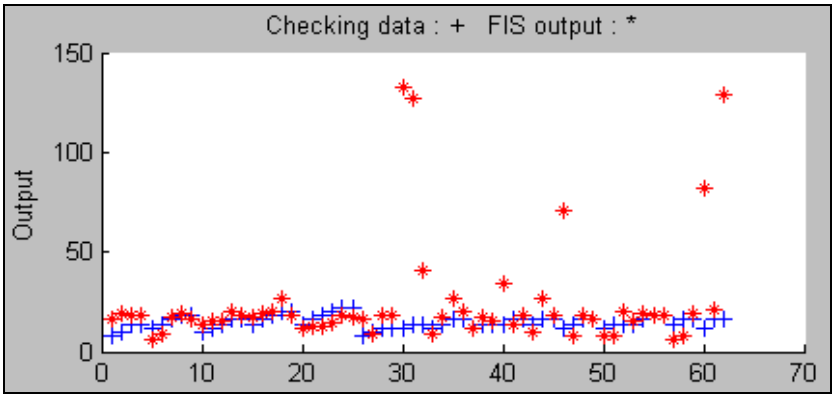

Figure 3.27: ANFIS checking data Plot.

For Insulin Lantus dataset ANFIS generates 7 rules. The checking data error of Insulin Lantus is 0.1859 and the RMSE of Novorapid dataset is 18 percent. The sensitivity rate of Insulin Lantus is 74 percent. The correctness of these rules is 80 percent. These rules are expressed below:

Rule 1 : [0 4 3 1 3 1 6 2 2 3 2 1 1 2 0 1][24]

If Gender=0 and Age=4 and Body Mass Index=3 and Genetic=1 and Insulin=3 and C_peptide=1 and Fast Blood Sugar=6 and Creatinine=2 and Cholesterol=2 and Triglyceride=3 and HDL=2 and LDL=1 and Uric Acid=1 and Glucose in Urine=2 and Acetone=0 and DM=1 then Output=24.

Rule 2 : [1 1 1 0 1 0 7 1 2 1 2 1 1 2 1 0][18]

If Gender=1 and Age=1 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=7 and Creatinine=1 and Cholesterol=2 and Triglyceride=1 and HDL=2 and LDL=1 and Uric Acid=1 and Glucose in Urine=2 and Acetone=1 and DM=0 then Output=18.

Rule 3 : [0 1 1 0 1 0 6 1 2 1 1 1 1 4 1 0][22]

If Gender=0 and Age=1 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=6 and Creatinine=1 and Cholesterol=2 and Triglyceride=1 and HDL=1 and LDL=1 and Uric Acid=1 and Glucose in Urine=4 and Acetone=1 and DM=0 then Output=22.

Rule 4 : [0 1 1 0 1 0 7 1 1 1 2 1 1 4 1 0][16]

If Gender=0 and Age=1 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=7 and Creatinine=1 and Cholesterol=1 and Triglyceride=1 and HDL=2 and LDL=1 and Uric Acid=1 and Glucose in Urine=4 and Acetone=1 and DM=0 then Output=16.

Rule 5 : [1 2 1 0 1 0 8 1 1 1 2 1 1 4 1 0][10]

If Gender=1 and Age=2 and Body Mass Index=1 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=8 and Creatinine=1 and Cholesterol=1 and Triglyceride=1 and HDL=2 and LDL=1 and Uric Acid=1 and Glucose in Urine=4 and Acetone=1 and DM=0 then Output=10.

Rule 6 : [1 2 2 1 1 0 7 1 1 1 2 1 1 4 1 0][20]

If Gender=1 and Age=2 and Body Mass Index=2 and Genetic=1 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=7 and Creatinine=1

and Cholesterol=1 and Triglyceride=1 and HDL=2 and LDL=1 and Uric Acid=1 and Glucose in Urine=4 and Acetone=1 and DM=0 then Output=20.

Rule 7 : [0 3 2 0 1 0 7 1 1 1 2 1 1 2 1 0][12]
If Gender=0 and Age=3 and Body Mass Index=2 and Genetic=0 and Insulin=1 and C_peptide=0 and Fast Blood Sugar=7 and Creatinine=1 and Cholesterol=1 and Triglyceride=1 and HDL=2 and LDL=1 and Uric Acid=1 and Glucose in Urine=2 and Acetone=1 and DM=0 then Output=12.

The above rules are in different situations for trained ANFIS model. Each rule is represented by a vector. It consists of input values for the system. For Insulin Lantus dataset FBS parameter's ranking ratio is 32 percent, glucose in urine and BMI parameters' ranking ratio are 24 percent, age parameter's ranking ratio is 23 percent, creatinine parameter's ranking ratio is 20 percent, acetone, c_peptide, diabetes type, insulin and triglyceride parameters' ranking ratio are 14 percent. These parameters are more effective than the others. For ANFIS gender, age, BMI, genetic, c_peptide, FBS, LDL, uric acid, acetone and diabetes type are used more effective than others. Descriptions of the rules which are generated by ANFIS for Insulin Lantus dataset are below.

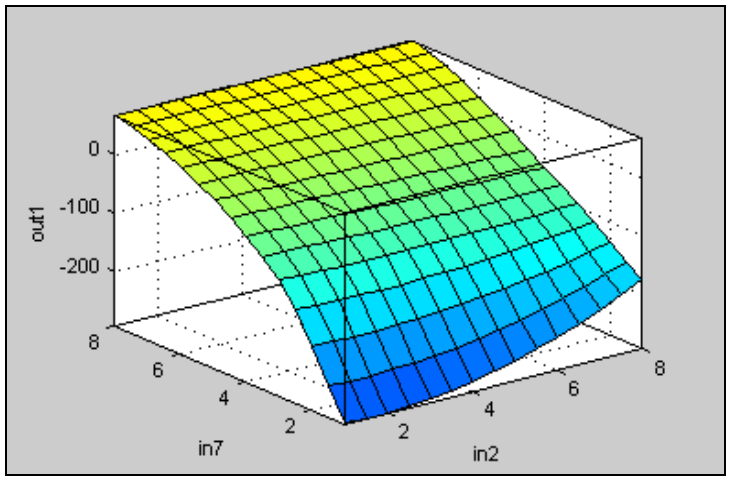

Figure 3.28: Surface Plot of FBS and Age versus Output.

Rule 1, patient is middle age and overweight. Genetic and c_peptide are positive. Insulin, triglyceride and FBS levels are more than normal range. Her urine has glucose but do not have acetone. Patient is a Type 2 diabetic. Other parameters' levels are acceptable in normal ranges. The generated output is 24 IU Insulin Lantus dosage.

Rule 2, patient is underweight whose age is between 0-7 years old. Genetic and C_peptide are negative. Patient has glucose and acetone in his urine. Insulin level is under normal range. FBS level is very high. . Patient is a Type 1 diabetic. Other parameters' levels are acceptable in normal ranges. The generated output is 18 IU Insulin Lantus dosage.

Rule 3 is similar with rule 2. Differences with rule 2 are FBS and HDL levels are less than rule 2 and her urine has more glucose than rule 2. Other parameters' levels are in same ranges with rule 2. The generated output is 22 IU Insulin Lantus dosage.

Rule 4 is similar with rule 3. Differences with rule 3 are FBS and HDL levels are more than rule 3 and cholesterol level is less than rule 3. Other parameters' levels are in same ranges with rule 3. The generated output is 16 IU Insulin Lantus dosage.

Rule 5, patient is underweight whose age is between 8-14 years old. Genetic and C_peptide are negative. Patient has glucose and acetone in his urine. Insulin level is under normal range. FBS parameter's level is between 501 and 600 mg/dl. Other parameters' levels are acceptable in normal ranges. Patient is Type 1 diabetic. The generated output is 10 IU Insulin Lantus dosage.

Rule 6 is similar with rule 5. Differences with rule 5 are patient is in normal weight and FBS level is between 401 and 500 mg/dl. Other parameters' levels are in same ranges with rule 5. The generated output is 20 IU Insulin Lantus dosage.

Rule 7, patient is in normal weight whose age is between 15-25 years old. Genetic and C_peptide are negative. Patient has glucose and acetone in her urine. Insulin level is under normal range. FBS parameter's level is between 401 and 500 mg/dl. Other parameters' levels are acceptable in normal ranges. Patient is Type 1 diabetic. The generated output is 12 IU Insulin Lantus dosage.

Roc curve of Insulin Lantus dataset is shown in figure 3.29:

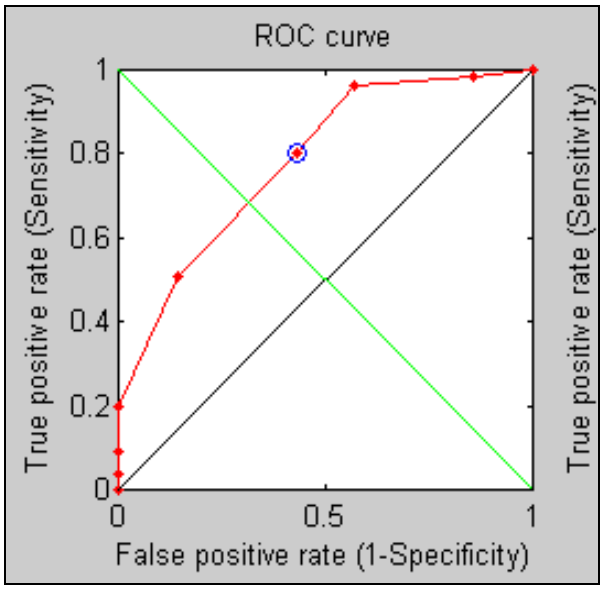

Figure 3.29: Roc curve of Insulin Lantus dataset.

ANFIS model structure of Insulin Lantus dataset is shown in figure 3.30. There are 16 inputs and 1 output.

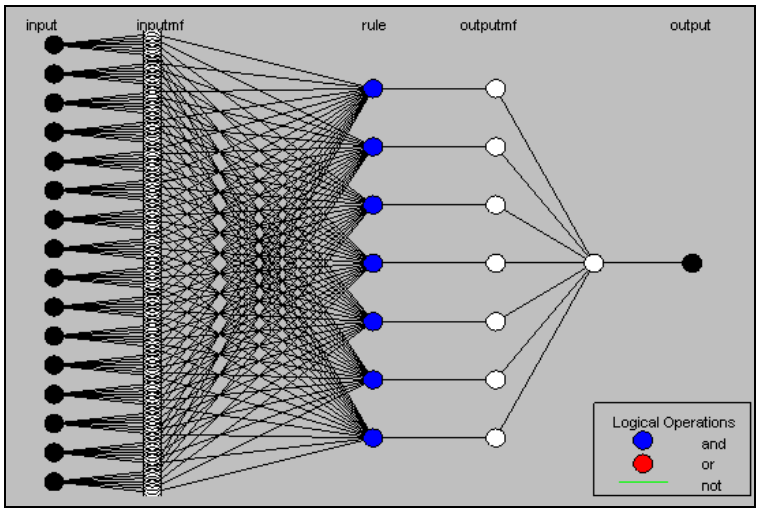

Figure 3.30: Anfis model structure of Insulin Lantus dataset.

3.2 ROUGH SET RESULTS

Rosetta software is used for RSES (a Rough Set algorithm implementation) algorithm.

I will start to introduce with Glifix dataset's RSES results. RSES algorithm generates 13 rules. Among those, I choose the specific ones and examine them. The selected rules are below:

Rule 1: BMI(4) => GlifixDosage(1)

Rule 2: FBS(3, 5) AND Creatinine(2) AND Triglyceride(2, 3) => GlifixDosage(3) OR GlifixDosage(1)

Rule 3: Age(6, 7) AND BMI(3, 4) => GlifixDosage(1)

Rule 4: Age(7) AND LDL([3) => GlifixDosage(2)

Rule 5: FBS(3, 5) AND Triglyceride(2) AND HDL(2) => GlifixDosage(3)

Rule 6: Gender(1) AND FBS(5) => GlifixDosage(1)

Rule 7: Gender(1) AND FBS(2, 3) AND Triglyceride(2, 3) => GlifixDosage(1)

When rules of above applied, possible outputs seen by the given inputs. To obtain result for a given input set, we use the resultant classification and each point of rules. The root mean square error of Glifix dataset is 0.5 and the RMSE rate is 50 percent.

Rule 1, the patient is obese and generated output is 1.

Rule 2, patient's FBS cluster range is between 3 and 5; creatinine cluster is 2, and triglyceride level within range of 2 and 3. The output can be 3 or 1.

Rule 3, patient's age is in cluster 6 or 7 and BMI is in cluster 3 or 4 than the output is 1.

Rule 4, patient's age is in cluster 7 and LDL level is in cluster 3 than the output is 2.

Rule 5, FBS cluster range is in 3 and 5, and triglyceride cluster range is 2, and HDL cluster range is 2, than the output is 3.

Rule 6, patient's gender is male and FBS cluster range is in 5, than the output is 1.

Rule 7, patient's gender is male and FBS cluster range is between 2 and 3, and triglyceride cluster range is between 2 and 3, than the output is 1.

Diamicron dataset RSES results are introduced below. RSES algorithm generates 23 rules. Among those, I choose the specific ones and examine them. The selected rules are below:

Rule 1: Creatinine(2, 3) AND HDL(3) => DiamicronDosage(2)

Rule 2: Urine(2) => DiamicronDosage(3) OR DiamicronDosage(4)

Rule 3: Cholesterol(2, 3) AND LDL(2, 3) => DiamicronDosage(4)

Rule 4: Gender(0) AND GU(1, 2) => DiamicronDosage(3)

Rule 5: Age(7, 8) AND BMI(3, 4) => DiamicronDosage(4)

Rule 6: Triglyceride(3) => DiamicronDosage(3) OR DiamicronDosage(4)

Rule 7: BMI(4) AND FBS(3, 4) => DiamicronDosage(3)

When rules of above applied, possible outputs seen by the given inputs. To obtain result for a given input set, we use the resultant classification and each point of rules. The root mean square error of Diamicron dataset is 0.107143 and the RMSE rate is 10 percent.

Rule 1, told that creatinine level is in cluster 2 and 3, and HDL range is in cluster 3 then the output is 2.

Rule 2 express that patient whose urine is in cluster 2 then the output is 3 or 4.

Rule 3, shows that cholesterol level is in range of cluster 2 or 3, and LDL level is in cluster 2 or 3 then the output is 4.

Rule 4, if patient's gender is female and has glucose level in her urine is 1 or 2 then the output is 3.

Rule 5, shows that patient's age is in range of cluster 7 and 8 and BMI is in range of cluster 3 or 4 then the output is 4.

Rule 6 express that the patient whose triglyceride level is in cluster 3 then the output is 3 or 4.

Rule 7, if the patient is obese and FSB level is in range of cluster 3 and 4 then the output is 3.

Glucophage dataset RSES results are introduced below. RSES algorithm generates 47 rules. Among those, I choose the specific ones and examine them. The selected rules are below:

Rule 1: Age(7, 8) AND Cholesterol(2) => GlucophageDosage(1) OR GlucophageDosage(2)

Rule 2: BMI(3, 4) AND Triglyceride(3) => GlucophageDosage(3)

Rule 3: BMI(3, 4) AND FBS(3, 4) AND Uric Acid(2, 3) => GlucophageDosage(2) OR GlucophageDosage(3)

Rule 4: Genetic(0) AND GU([1, 2)) => GlucophageDosage(3)

Rule 5: Age(7) AND FBS(3) AND HDL(3) => GlucophageDosage(3)

Rule 6: Gender(0) AND BMI(3) AND LDL(2, 3) => GlucophageDosage(2)

Rule 7: Creatinine(2, 3) AND Uric Acid(2) AND GU(1, 2) => GlucophageDosage(2)

When rules of above applied, possible outputs seen by the given inputs. To obtain result for a given input set, we use the resultant classification and each point of rules. The root mean square error of Glucophage dataset is 0.615385 and the RMSE rate is 61 percent.

Rule 1, shows that patient whose age is in cluster 7 or 8 and cholesterol level is in cluster 2 then the output is 2.

Rule 2, the patient is overweight or obese and triglyceride level is in cluster 3 then the output is 3

Rule 3, told that the patient is overweight or obese and FBS level is in cluster 3 or 4 and uric acid level is in cluster 2 or 3 then the output is 2 or 3.

Rule 4, the patient's genetic is negative and urine's glucose level is 1 or 2 then the output is 3.

Rule 5; if the patient's age is in cluster 7 and FBS level is in cluster 3 and HDL level is in cluster 3 then the output is 3.

Rule 6, shows that patient whose is overweight and female and LDL level is in cluster 2 or 3 then the output is 2.

Rule 7, defines that patient whose creatinine level is in cluster 2 or 3 and uric acid level is in cluster 2 and urine has glucose level 1 or 2 then the output is 2.

Humilin_M dataset RSES results are introduced below. RSES algorithm generates 23 rules. Among those, I choose the specific ones and examine them. The selected rules are below:

Rule 1: FBS(5) AND LDL([2, 3)) => Humilin_Mdosage(18)

Rule 2: Gender(0) AND Creatinine([2, 3)) AND Triglyceride([2, 3)) => Humilin_Mdosage(18)

Rule 3: FBS(5) AND Urine(2) => Humilin_Mdosage(22)

Rule 4: Genetic(0) AND FBS(5) => Humilin_Mdosage(18)

Rule 5: Gender(0) AND GU(3) => Humilin_Mdosage(14)

Rule 6: FBS(3) AND Cholesterol([2, 3)) => Humilin_Mdosage(18)

Rule 7: BMI(3) AND FBS([3, 5)) => Humilin_Mdosage(22)

When rules of above applied, possible outputs seen by the given inputs. To obtain result for a given input set, we use the resultant classification and each point of rules. The root mean square error of Humilin_M dataset is 0.25 and the RMSE rate is 25 percent.

Rule 1; if FBS level is in cluster 5 and LDL level is in cluster 2 or 3 then the output is 18.

Rule 2, defines that patient who is female and creatinine level is in cluster 2 or 3 and triglyceride level is in cluster 2 or 3 then the output is 18.

Rule 3, shows that patient whose FBS level is in cluster 5 and Urine level is in cluster 2 then the output is 22.

Rule 4, told that patient whose genetic is negative and glucose level in urine is 3 then the output is 18.

Rule 5, if the patient is female and glucose level in urine is 3 then the output is 14.

Rule 6, shows that the patient whose FBS level is in cluster 3 and cholesterol level is in cluster 2 or 3 then the output is 18.

Rule 7, if the patient is overweight and FBS level is in range of cluster 3 and 5 then the output is 22.

Novorapid dataset RSES results are introduced below. RSES algorithm generates 66 rules. Among those, I choose the specific ones and examine them. The selected rules are below:

Rule 1: Gender(1) AND Age(3) AND BMI(2) AND FBS(4) AND HDL(3) => NovorapidDosage(18) OR NovorapidDosage(20)

Rule 2: Age(3) AND BMI([2, 3)) AND FBS([4, 6)) AND HDL(3) => NovorapidDosage(18) OR NovorapidDosage(16)

Rule 3: Age([3, 6)) AND Genetic(1) AND FBS([4, 6)) => NovorapidDosage(16) OR NovorapidDosage(18)

Rule 4: FBS(6) AND HDL(3) AND GU([1, 3)) => NovorapidDosage(20)

Rule 5: FBS([4, 6)) AND Uric Acid(2) => NovorapidDosage(16) OR NovorapidDosage(14)

Rule 6: BMI(2) AND HDL(3) AND GU(3) AND As(0) => NovorapidDosage(18)

Rule 7: BMI(2) AND FBS(6) AND Triglyceride([2, 3)) => NovorapidDosage(16) OR NovorapidDosage(12)

When rules of above applied, possible outputs seen by the given inputs. To obtain result for a given input set, we use the resultant classification and each point of rules. The root mean square error of Novorapid dataset is 0.322034 and the RMSE rate is 32 percent.

Rule 1, if the patient whose male and age is in cluster 3 and has normal weight and FBS level is in cluster 4 and HDL level is in cluster 3 then the output is 18 or 20.

Rule 2, shows that patient whose age is in cluster 3 and weight is normal or overweight and FBS level is in range of cluster 4 and 6 and HDL level is in cluster 3 then the output is 18 or 16.

Rule 3, defines that patient whose age is in range of cluster 3 and 6 and genetic is positive and FBS level is in range of cluster 4 and 6 then the output is 16 or 18.

Rule 4, if the patient's FBS level is in cluster 6 and HDL level is in cluster 3 and Urine Glucose level range is 1 and 3 then the output is 20.

Rule 5, told that patient whose FBS level is in range of cluster 4 and 6 and uric acid is in cluster 2 then the output is 16 or 14.

Rule 6, if the patient has normal weight and HDL level is in cluster 3 and urine glucose level is 3 and acetone is negative then the output is 18.

Rule 7, shows that patient who has normal weight and FBS level is in cluster 6 and triglyceride level is in range of cluster 2 and 3 then the output is 16 or 12.

Insulin Lantus dataset RSES results are introduced below. RSES algorithm generates 70 rules. Among those, I choose the specific ones and examine them. The selected rules are below:

Rule 1: Age(4) AND FSB(6) => InsulinLantusDosage(10)

Rule 2: Gender(0) AND Age(3, 4) AND BMI(2, 3) AND As(1) => InsulinLantusDosage(10) OR InsulinLantusDosage(16)

Rule 3: BMI(2) AND FSB(4, 6) AND Triglyceride(2, 3) => InsulinLantusDosage(22)

Rule 4: Age([3, 4)) AND BMI(2) AND FSB(4, 6) AND HDL(3) => InsulinLantusDosage(20)

Rule 5: Age(3) AND FSB(6) AND GU(1, 3) => InsulinLantusDosage(16)

Rule 6: Gender(0) AND Age(3, 4) AND Genetic(1) AND FSB(4, 6) => InsulinLantusDosage(14)

Rule 7: Genetic(1) AND FSB(4, 6) AND HDL(3) => InsulinLantusDosage(10) OR InsulinLantusDosage(20) OR InsulinLantusDosage(12) OR InsulinLantusDosage(14)

When rules of above applied, possible outputs seen by the given inputs. To obtain result for a given input set, we use the resultant classification and each point of rules. The root mean square error of Insulin Lantus dataset is 0.50645 and the RMSE rate is 50 percent.

Rule 1, told that patient whose age is in cluster 4 and FBS level is in cluster 6 then the output is 10.

Rule 2, if patient is female and age is in cluster 3 or 4 and has normal weight or overweight and Acetone is positive then the output is 10 or 16.

Rule 3, patient who has normal weight and FBS level is in range of cluster 4 and 6 and triglyceride is in cluster 2 or 3 then the output is 22.

Rule 4, shows that patient whose age is in cluster 3 or 4 and has normal weight and FBS level is in range of cluster 4 and 6 and HDL level is in cluster 3 then the output is 20.

Rule 5, defines that patient's age level is in cluster 3 and FBS level is in cluster 6 and urine glucose level range is 1 and 3 then the output is 16.

Rule 6, if the patient is female and age is in range of cluster 3 and 4 and genetic is positive and FBS level is in range of cluster 4 and 6 then the output is 14.

Rule 7, shows that patient whose genetic is positive and FBS level is in range of cluster 4 and 6 and HDL level is in cluster 3 then the output is 10 or 20 or 12 or 14.

4. CONCLUSION AND FUTURE PLANS

There are many data mining studies in the world. Using data mining techniques, many studies have been made about the Diabetes disease. During these studies the most important step to find right data and that data is correctly interpreted. The aim of this thesis study is dosage planning with ANFIS.

In this thesis study, I try to determine the degree of drug amount dosage will be use. 318 diabetic assays were used. As a result of these assays, more than one drug was used. These data were classified into 6 classes, according to the types of used drugs. As a result of this classification, six data sets are obtained. These data sets are Glifix dataset, Diamicron dataset, Glucophage dataset, Humilin_M dataset, Novorapid dataset and Insulin Lantus dataset. Glifix, Diamicron and Glucophage are oral medicines. Humilin_M, Novorapid and Insulin Lantus are injections. Each data sets' parameters were changes according to the ranking algorithm. The parameters were eliminated according to their ranking values that were zero. The eliminated parameters were not having much more importance as the others. Glifix dataset has 13 parameters, Diamicron dataset has 14 parameters, Glucophage dataset has uses 14 parameters, Humilin_M dataset has 13 parameters, Novorapid dataset has 15 parameters and Insulin Lantus dataset has 16 parameters.

According to ANFIS results FBS, genetic, BMI, age and HDL parameters more deterministic for all datasets. Also, doctors used to use these parameters to decide the drug dosage. All parameters were grouped in each other to get effective results from ANFIS. Such as a male patient has 208 mg/dl fast blood glucose level, his mother is a diabetic, he has normal weight and 40 years old. In the first stage, doctor defined that this patient will begin treatment with using Diamicron according to these parameter values. According to medical doctors this patient could be a Type 2 diabetic. The older age, genetic and FBS not be too high are the most decisive parameters for them. Checking c_peptide and insulin in blood those decision is approved. Because patient has not got c_peptide in his blood, insulin level is 23.6 above the normal range, triglyceride level is 287 and he has glucose in his urine. These are the most important parameters for doctors to determine the patient is diabetic or not. The doctor decided to begin 1*3

Diamicron in a day, these corresponds 90 mg Diamicron dosage. I prepared an input vector for ANFIS with the results of this patient's assay for my data classification rules which I used in this thesis study. I obtained the output 3 which equals 90 mg Diamicron dosage. As you can see from this sample we can achieve successful results.

In order to compare the success of the results I applied another data mining technique with RSES algorithm. The comparison of ANFIS and RSES, ANFIS has very acceptable output results. Especially aspect of applied ANFIS, great method for decided dosage of drug for diabetic diseases.

This thesis study will be helpful information for diabetic patients and doctors. This study can improve for the human being health.

REFERENCES

1. Avci, E. & Turkoglu, I., 2009. *An intelligent diagnosis system based on principle component analysis and ANFIS for the heart value disease.* Expert Systems with Applications **36**, pp: 2873-2878, Available at: http://www.sciencedirect.com/

2. Berka, P., Rauch, J. & Zighed A. D., 2008. *Data Mining and Medical Knowledge Management: Cases and Applications.* New York: Hersey, Medical Information Science reference, pp: 76-107.

3. Breault, J. L. *Data mining diabetic databases: are rough sets a useful addition?* Department of Health Systems Management, Tulane University.

4. Breault, J. L., Goodall, C. R. & Fos, P. J., 2002. *Data mining a diabetic data warehouse*. Elsevier, Artificial Intelligence in Medicine 26, pp 37-54.

5. Firat, M. & Gungor, M., 2006. *Sugeno fuzzy inference system for river flow estimation.* Available at: http://www.dsi.gov.tr/english/congress2007/chapter_4/120.pdf

6. Ghazavi, S. N. & Liao, T. W., 2008. *Medical data mining by fuzzy modelling with selected features.* Elsevier, Artificial Intelligence in Medicine, **43**, pp: 195-206

7. Goren, S., Karahoca, A., Onat, F. Y. & Goren, Z., 2008. *Prediction of cyclosporine A blood levels: an application of the adaptive- network-based fuzzy inference system (ANFIS) in assisting drug therap.,* Eur J Clin Pharmacol **64**, 807-814.

8. Guopeng, Z. & Levin, 2006. *Data analysis with fuzzy inference system.,* Available at: http://www3.ntu.edu.sg/home/aswduch/Teaching/.

9. Huang, Y., McCullagh, P., Black, N. & Harper, R., 2007. *Feature selection and classification model construction on type 2 diabetic patients' data*. Elsevier, Artificial Intelligence in Medicine **41**, 251-262.

10. Jaafar, A.F.B, Jais, J., Hamid M.H., Rahman, Z.B.A, & Benaouda, D, 2006. *Using Rough Set as a tool for knowledge discovery in DSS.* Malaysia, Current Developments in Technology-Assisted Education, pp: 1011-1015.

11. Jang, JSR., 1993. *ANFIS: Adaptive–network-based fuzzy inference system.* IEEE Trans Syst Man Cybern **23**, pp 665-685.

12. Kara, A. 2008. *Developing an expert-system for diabetics by supporting with ANFIS.* Master Thesis, Bahcesehir University Institute of Science Computer Engineering.

13. Luukka, P., 2006. *Similarity classifier using similarity measure derived from Yu's norms in classification of medical datasets.* Elsevier, Computers in Biology and Medicine (2007) **37**, pp: 1133- 1140.

14. Magnani, M., 2003. *Technical report on rough set theory for knowledge discovery in data bases.* Available at: http://magnanim.web.cs.unibo.it/data/pdf/roughkdd.pdf

15. Ohrn, A. & Komorowski, J. *Rosetta - a rough set toolkit for analysis of data.*

16. Pandey, B. & Mishra, R.B., 2008. *Knowledge and intelligent computing system in medicine.* India, Elsevier, Computers in Biology and Medicine **39**, pp: 215-230.

17. Polat, K. & Gunes, S., 2006. A*n expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease.* Elsevier, Digital Signal Processing.

18. Polkowski, L. & Skowron, A., 1998. *Rough Sets in Knowledge Discovery 2 Applications, Case Studies and Software.* Germany: Physica Verlag Heidelberg, pp 15-20, Available at: http://books.google.com.tr/

19. Porte, D., Robert, Jr., Sherwin, S. & Baron, A., 2003. *Ellenberg & rifkin's diabetes mellitus.* 6th edition, United States: Quebecor World Kingsport press, pp 331-367.

20. Powers, M. A., 1996. *Handbook of diabetes medical nutrition therapy*. 2nd edition, United States of America: Aspen Publishers Inc., pp 115-125.

21. Shafiq, Z. M., Farooq, M. & Khayam, S. A., 2008. *A comparative study of fuzzy inference systems, neural networks and adaptive neuro fuzzy inference systems for portscan detection.* Evo Workshop LNCS **4974**, pp 52-61.

22. Sigurdardottir, A. K., Jonsdottir, H. & Benediktsson, R. 2007. *Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis*. Elsevier, Patient Education and Counselling 67, pp 21-31.

23. Sivanandam, S. N., Sumathi, S. & Deepa, S. N., 2007. *Introduction to Fuzzy Logic Using Matlab.* New York: Springer Berlin Heidelberg, pp 11-40, Available at: http://books.google.com.tr/

24. Song, Q., Kasabov, N., Ma, T. & Marshall, M. R., 2005. *Integrating regression formulas and kernel functions into locally adaptive*

knowledge based neural networks: A case study on renal function evaluation. Elsevier, Artificial Intelligence in Medicine (2006) **36**, pp: 235-244.

25. Su, C. T., Yang, C. H., Hsu, K. H. & Chiu, W.K., 2005. *Data mining for the diagnosis of type 2 diabetes from three-dimensional body surface anthropometrical scanning data*. Elsevier, Computers and Mathematics with applications, **51**, pp 1075-1092.

26. Temurtas, H., Yumusak, N. & Temurtas, F., 2008. *A comparative study on diabetes disease diagnosis using neural networks.* Elsevier, Expert Systems with Applications **36**, pp 8610-8615.

27. Williams, L. & Wilkins, 2007. *Diabetes mellitus a guide to patient care.* United States: a Wolters Kluwer Business press.

28. Yun, J., Zhanhuai, L., Yong, W. & Longbo, Z., 2006. *A better classifier based on rough set and neural network for medical images.* Sixth IEEE International Conference on Data Mining, pp: 853-857.

29. Zadeh, LA, 1965. *Fuzzy sets*. Inf Control 8, pp 338-353.

30. ANFIS Gui, Available at: http://www.mathworks.com/products/matlab/ [Cited 2 February, 2008]

31. http://www.ai-cit.sk/source/publications/thesis/master_thesis/1997/holecy/html/ [Cited 17 March, 2009].

32. http://www.diabetesmellitus-information.com/ [Cited 13 February, 2008].

33. http://www.diabetesnet.com/diabetes_treatments/. [Cited 14 February, 2009].

34. http://en.wikipedia.org/wiki/Rough_set/ [Cited 21 March, 2009].

VITAE

Name Surname: Emirhan Gülçin YILDIRIM

Address: Merkez mh. Namıkkemal Cd. No: 15 Daire: 7 Avcılar / ISTANBUL

Birth place and year: Konya 1982

Foreign Language: English

Primary School: Mehmet Akif Ersoy Primary School

High School: Sinop Anatolian High School

Under Graduate: Istanbul University 2006

Graduate: Bahcesehir University

Institute Name: Institute of Science

Program Name: Computer Engineering

Working Life: Yapi Kredi Sigorta 2007-