

**DEVELOPMENT OF A GENETIC NEURO FUZZY INFERENCE SYSTEM
FOR DIAGNOSIS OF DIABETES MELLITUS**

OLADIPO, IDOWU DAUDA

[07/68FH005]

APRIL, 2018

DEVELOPMENT OF A GENETIC NEURO FUZZY INFERENCE SYSTEM FOR DIAGNOSIS OF DIABETES MELLITUS

By

**OLADIPO, IDOWU DAUDA
(07/68FH005)**

B.Sc.(Edu) (UNAD) 2005, M.Sc. (Unilorin) 2010

**A Thesis Submitted to the Department of Computer
Science, Faculty of Communication and Information
Sciences in Partial Fulfillment of the Requirements for
the Award of the Degree of Doctor of Philosophy in
Computer Science**

APRIL, 2018

CERTIFICATION

This is to certify that this thesis has been read and approved as meeting the requirements of the Department of Computer Science, University of Ilorin, Ilorin, Nigeria for the Award of Doctor of Philosophy degree in Computer Science.

Dr. A. O. Babatunde
(Supervisor)

Date

Dr. D. R. Aremu
(Head of Department)

Date

Dr.(Mrs.) Tinuke. O. Oladele
(Postgraduate Coordinator)

Date

(External Examiner)

Date

DEDICATION

This research work is dedicated to my parents.

DECLARATION

I, OLADIPO, Idowu Dauda with Matriculation Number 07/68FH005 hereby declare that this thesis entitled Development of a Genetic Neuro Fuzzy Inferential System for Diagnosis of Diabetes Mellitus is a record of my research work. It has neither been presented nor accepted in any previous application for a higher degree. All sources of information have been specifically acknowledged.

In addition, the research work has been ethically approved by the University Ethical Review Committee.

Oladipo, I. D.

Date

ACKNOWLEDGEMENTS

I wish to express my profound gratitude to God Almighty for seeing me through to the successful completion of this Ph.D. Program, my appreciation also goes to my Supervisor Dr. A. O. Babatunde for his constructive criticisms encouragement during the course of this research work, my thanks also goes to all Postgraduate committee staff member and other staff of the department of Computer Science Dr. R. G. Jimoh, Dr.(Mrs.) C. O. Abikoye, Dr. R. O. Oladele, Dr. D. R. Aremu, Dr. (Mrs.) T. O. Oladele, Dr. A. O. Bajeh, Dr.(Mrs.) M. A. Mabajoje, Dr.(Mrs.) Akintola, Dr. A.R. Ajiboye, Dr. (Mrs.) G. B. Balogun, Dr.(Mrs.) F. E. Usman-Hamza, Mr. A. O. Balogun, Mr. P. Sadiku, Mr. H. A. Mojeed, Mrs. S. A. Saliu and all non-academic staff too numerous to mention are highly appreciated.

Special thanks to Dr. A. L. Azeez of the Department of mass communication, University of Ilorin, for his advice and guidance at all times. I am also indebted to the following erudite Professors who has been my guide, supporter, adviser and father before and during this program: Prof. J. S. Sadiku, Prof. R. A. Ipinyomi and Prof. O. S. Adegboye.

I am also indebtedness to my immediate family, most especially my wife Mrs. B. I. Oladipo and my children Musharafat, Abdul Rahman, Hanifat and Zainab for tolerating my absenteeism in house during the course of the program. The contribution, discussion of idea shared from my friends and Colleagues in person of J. B. Awotunde, Aro T. O. , Odetunde Opeyemi and Fatai Tomori cannot be underestimated.

The expertize knowledge and technical assistance obtained from University of Ilorin Teaching Hospital (UTH) especially from Dr. J. K. Olarinoje of the department of internal medicine during the course the research work is appreciated.

I gratefully acknowledge the entire family of Late Alhaji Rahman Oladipo both nuclear and extended, most especially my beloved mother Mrs. Mariyam Arinpe Oladipo, my sister and brother for their encouragement, moral and financial support.

Last but not the least, the contribution of Dr. Rasheed A. Tomori Dr. Adewole, Kayode, Engr. Yinka Okeola and Alhaja S. I. Adio-Yusuf and all other academic and non-teaching staff of the faculty of Communication and Information Sciences, University of Ilorin is highly appreciated.

TABLE OF CONTENTS

TITLE	i
CERTIFICATION	iii
DEDICATION	iv
DECLARATION	v
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	viii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF EQUATION	xiv
LIST OF ABBREVIATIONS AND NOTATIONS	xvi
ABSTRACT	xviii

CHAPTER ONE: INTRODUCTION

1.1	Background to the Study	1
1.2	Statement of the Problem	7
1.3	Aim and Objectives	8
1.4	Significance of the Study	9
1.5	Scope of the Study	10
1.6	Definition of Operational Terms	10
1.7	Organization of the Thesis	14

CHAPTER TWO: REVIEW OF LITERATURE

2.1	Computational Intelligent Techniques (CIT)	15
-----	--	----

2.2	Basic Concepts in Medical Informatics	17
2.3	Basic Concept of Fuzzy Logic Approach	20
2.3.1	Practical Issues of Fuzzy Logic in Medicine Applications	25
2.3.2	Applications of Fuzzy Logic in Diabetes	26
2.3.3	Drawbacks of Fuzzy Logic	27
2.4	Artificial Intelligence	27
2.4.1	Neural Network	28
2.4.2	Artificial Neural Networks	29
2.4.3	Mathematical Background	31
2.4.4	Overview of ANNs in Diabetes	34
2.4.5	Drawbacks of ANNs	35
2.4.6	ANN Approach in Diabetes Mellitus Prediction	36
2.5	Adaptive Neuro-Fuzzy Inference System	39
2.5.1	Neuro-Fuzzy System (NFS) in Medical System	40
2.6	Genetic Algorithm	43
2.6.1	Pseudo code for Genetic Algorithm	46
2.6.2	Overview of Genetic Algorithm (GA) in Diabetes	48
2.6.3	Drawbacks of Genetic Algorithm	49
2.7	Diabetes Mellitus	49
CHAPTER THREE: RESEARCH METHODOLOGY		
3.1	Data for the Proposed System	52
3.2	Dataset for the Proposed System	54
3.3	Indicators of Diabetes Mellitus	55
3.4	Neural Network (NN)	56

3.5	Fuzzy Logic	57
3.5.1	Algorithm for Fuzzy logic	57
3.6	Genetic Algorithm	63
3.7	Genetic-Neuro-Fuzzy Inference System (GNFIS)	63
3.8	System Development and Design	64
3.9	Performance Evaluation Metrics	68
3.9.1	Receiver Operating Characteristics (ROC)	69
3.9.2	10-Fold Cross Validations	70
CHAPTER FOUR: RESULTS AND DISCUSSION		
4.1	Genetic Neuro-fuzzy Inference Diagnostic System Experimental Results	72
4.1.1	The experimental results of the GNFIS	72
4.2	Performance Evaluation	75
CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS		
5.1	Summary	92
5.2	Conclusion	93
5.3	Recommendation	93
5.4	Suggestion for Future Research	93
5.5	Contribution to Knowledge	94
	Reference	95
	Appendix A	105
	Appendix B	123

LIST OF TABLES

Table 2.1: Comparison of Type – I and Type-II Diabetes	51
Table 3.1: Physicians Rating for the Indicators of the Diabetes Mellitus	53
Table 3.2 Features of PIMA Indians Diabetes Dataset for Diagnosing of Diabetes	54
Table 3.3: Ranges of the Output Fuzzy set for Diabetes Mellitus Application	59
Table 4.1: Evaluation Metrics obtained from GNFIS diagnosis	73
Table 4.2: Accuracy obtained from the three systems for diagnosis	74
Table 4.3: Classification Accuracy and time taken to diagnose when First feature is removed (Number of time Pregnant)	75
Table 4.4: Classification Accuracy and time taken to diagnose when second feature is removed (Diastolic Blood pressure)	75
Table 4.5: Classification Accuracy and time taken to diagnose when third feature is removed (Triceps skin fold thickness)	76
Table 4.6: Classification Accuracy and time taken to diagnose when fourth feature is removed (Number of Time Pregnant and Triceps Skin Fold Thickness)	76
Table 4.7: Classification Accuracy and time taken to diagnose when fifth feature is removed (Number time pregnant and Diastolic Blood Pressure)	77
Table 4.8: Classification Accuracy and time taken to diagnose when sixth feature is removed (Body Mass Index)	77
Table 4.9: Classification Accuracy and time taken diagnose when seventh feature is removed (Diabetes Pedigree Function)	78
Table 4.10: Classification Accuracy and time taken to diagnose when eighth feature is removed (Age)	78
Table 4.11: Classification Accuracy and time taken to diagnose when seventh feature is removed (Plasma Glucose Concentration)	79
Table 4.12: Classification Accuracy and time taken to diagnose when eighth feature is removed (2-Hour Serum Insulin)	79

Table 4.13: Results for sensitivity and Specificity of GNFIS, Fuzzy logic and Neural network	82
Table 4.14 Experimental Results for 100 runs	85
Table 4.15: The Performance of Three Systems for Diabetes Mellitus Incidence	90
Table 4.16: Comparison of the proposed GNFIS with the Existing Systems	91

LIST OF FIGURES

Figure 2.1: Examples of elements of the network of medical knowledge	20
Figure 2.2: Membership functions for consequence B using minimum Combination	24
Figure 2.3 Principle of FRBS responses	25
Figure 2.4: Schematization of Biological neuron	29
Figure 2.5: Activation functions	30
Figure 2.6: Back-propagation ANN	31
Figure 3.1: Sentence Analysis represents the physical data of the person	58
Figure 3.2: Sample of Pima Indians Diabetes Dataset (PIDD)	65
Figure 3.3: Block Diagram for the Proposed System	67
Figure 4.1: Graphical Representation of Evaluation Metrics for GNFIS	73
Figure 4.2: The accuracy (%) for the three System for diabetes diagnosis	74

LIST OF EQUATION

Equation (2.1)	21
Equation (2.2)	22
Equation (2.3)	22
Equation (2.4)	23
Equation (2.5)	23
Equation (2.6)	23
Equation (2.7)	23
Equation (2.8)	30
Equation (2.9)	30
Equation (2.10)	30
Equation (2.11)	32
Equation (2.12)	46
Equation (2.13)	46
Equation (2.14)	47
Equation (3.1)	56
Equation (3.2)	56
Equation (3.3)	58
Equation (3.4)	60
Equation (3.5)	62
Equation (3.6)	64
Equation (3.7)	64
Equation (3.8)	68

Equation (3.9)	68
Equation (3.10)	68
Equation (3.11)	68
Equation (3.12)	68

LIST OF ABBREVIATIONS AND NOTATIONS

(i)	FL	-	Fuzzy Logic
(ii)	FIS	-	Fuzzy Inferential System
(iii)	AI	-	Artificial Intelligence
(iv)	PIP	-	Present Illness Program
(v)	DM	-	Diabetes Mellitus
(vi)	PCA	-	Principal Component Analysis
(vii)	DIALOG	-	Diagnostic Logic
(viii)	IT	-	Information Technology
(ix)	DoF	-	Degree of Fulfillment
(x)	FRBS	-	Fuzzy Rule Based System
(xi)	FLC	-	Fuzzy Logic Component
(xii)	ES	-	Expert System
(xiii)	NFS	-	Neuro-Fuzzy System
(xiv)	LSE	-	Least Square Estimator
(xv)	ANFIS	-	Adaptive Neuro-Fuzzy Inferential System
(xvi)	LM	-	Levenberg Marquardt
(xvii)	RMSE	-	Root Mean Squared Error
(xviii)	GD	-	Gradient Descent
(xix)	RPROP	-	Resilient Propagation
(xx)	QP	-	Quick Prop
(xxi)	SCG	-	Scaled Conjugate Gradient
(xxii)	PSO	-	Particle Swamp Optimization
(xxiii)	DCT	-	Discrete Cosine Transformation
(xxiv)	EEG	-	Electro Encephalogram

(xxv)	LDA	-	Linear Discriminant Analysis
(xxvi)	BBO	-	Biogeography Based Optimization
(xxvii)	EA	-	Evolutionary Algorithm
(xxviii)	NFL	-	Neuro-Fuzzy Logic
(xxix)	FANFIS	-	Fast Adaptive Neuro-Fuzzy Inference System

ABSTRACT

Diabetes Mellitus (DM) is one of the most chronic and debilitating diseases in the modern society and represents not only a medical problem, but also a socio-economic problem. Computational Intelligence Techniques (CIT) have been successfully employed in diabetes disease diagnosis, risk evaluation, patient monitoring and prediction in medical field. Using single techniques in the diagnosis of diabetes has been comprehensively investigated showing some level of accuracy in Fuzzy Logic (FL) and Artificial Neural Network (ANN) for diagnosis of diabetes mellitus. Therefore, this study aimed at developing a Genetic Neuro Fuzzy Inferential System (GNFIS) for the diagnosis of diabetes. The objectives were to: (i) develop an enhanced hybrid system for diagnosis of diabetes mellitus using genetic algorithm; (ii) implement the enhanced system using Java programming language; (iii) evaluate the performance of the proposed system based on accuracy, sensitivity and specificity; and (iv) perform a comparison of the proposed system with two existing systems; Fuzzy Logic (FL) and Artificial Neural Network (ANN) for diagnosis of diabetes mellitus.

The Neuro fuzzy inferential system was used to classify diabetes mellitus with genetic algorithm applied to obtain most relevant attributes from Pima Indian Diabetes Dataset (PIDD). Direct rating method was used for acquiring data used for the system. These data were presented in a series of objects to a domain expert who was requested to rate the membership function for the eight most significant attributes which also serves as input to the system. The ratings were aggregated for membership function calculation. The lowest was used as the minimum and highest as the maximum values. Triangular membership function was used because of its flexibility and fewer complexities when splitting values (low, medium and high), compared to other membership functions.

The findings of the study were that:

- i. the enhanced hybrid system using genetic algorithm performed better during the diagnosis process for diabetes mellitus;
- ii. the enhanced system was implemented using Java Programming Language and it achieved a high level of accuracy;
- iii. the GNFIS gave a minimum diagnosis accuracy of 98.26%, maximum diagnosis accuracy of 99% and the average accuracy of 97.76 %; sensitivity of 96% and specificity of 99% for the reduced dataset; and
- iv. results of comparison showed that GNFIS had a better performance with 99.34% accuracy on the whole dataset used when compared with FL and ANN with 96.14% and 95.14% respectively.

The study concluded that the genetic algorithm is a good attributes reduction technique for Neuro Fuzzy Inferential System. The developed GNFIS performed efficiently and also outperformed the existing systems. Thus, the study recommended GNFIS as a good technique for the screening and diagnosis of DM.

CHAPTER ONE

INTRODUCTION

This chapter discusses the underlying concepts and ideologies which serve as a background to the research. It states the statement of problem, aim and objectives of the study, significance of the study and the scope of the study. The chapter also provides the definitions of some operational terms employed in the research and organization of the thesis.

1.1 Background for the Study

Medical field is a large domain that involves a considerable number of expertise, knowledge and information. Information collection, sharing expertise and knowledge have been difficult tasks among medical experts and frequently become problem when knowledge is not easy to access. In order to overcome this problem, researchers like: {Adlassnig (2001); Dazzi, Simpson and Goldman (2001); Lehmann (2004) and Liao (2005)} have used Information and Communication Technology (ICT) to manage knowledge, working hand in hand with medical professionals. Knowledge collected and shared can be used for prediction and diagnosis of diseases (Ambilwade & Manza, 2014).

Malaria, hypertension, heart diseases and diabetes are chronic diseases that require tedious tests to confirm its signs and symptoms, which may require hours and days (Chang-Shing

& Mei-Hui, 2011). Therefore, there is a need to have a medical diagnostic system that can help experts to make timely diagnosis and proceed with initial treatments. Though, the diagnostic decision support system also relies on the experience and perception of the medical experts, with increase in complexity of system, it may not be easy to follow a specific path of diagnosis without any mistake (Prasath, Lakshmi, Nathiya & Neetha, 2013). Thus, it is necessary to have a strong reasoning method which can control uncertainties and vagueness of which Fuzzy Systems (FS) can provide (Galina, Mariusz, Wioletta, Pirog-Mazur & Tomasz, 2009).

Chioma, Obinna (2014), reported in Vanguard Newspaper of November 19, 2014, diabetes is on the increase in Nigeria and the provision for adequate equipped medical and personnel to handle the various intricate lifelong management of diabetes is very low. In a survey conducted in the year 2005 in the United States of America, by Centre for Disease Control (CDC), it was revealed that at age 60 years and above, 1 in 5 people have Diabetes Mellitus (DM). American Indians and Alaskan natives have the highest incidence of (17.9%) (CDC, 2009). In the global trend Nigeria and other developing countries are not left out. Over-dependent on western diet in many societies in developing countries have attributed to this increase. The rate of diabetes incidence is due largely to central obesity, physical inactivity, unhealthy dietary habits and obesity, in particular. (Hu, 2011).

FS establishes imprecise knowledge and proposes a linguistic idea with quality approximation to medical texts (Adlassnig, 2001). FS can be described as a branch of applied artificial intelligence. It is a software application used to take an expert view in the absence of a human expertise (Ambilwade & Manza, 2014) and has been applied to a

number of diseases diagnoses which includes Diabetes Mellitus (DM). (American Diabetes Association, 2006). Diabetes is defined as a group of metabolic activity with absolute or comparatively insulin deficiency. There are certain set of criteria used to diagnose diabetes as developed by American Diabetes Association (American Diabetes Associate, 2006). Medical doctors diagnose diabetes mellitus if the fasting blood glucose is higher than 7.0mM and if blood glucose is higher than 11.0mM in 2h oral glucose tolerance test (Tadić, et. al., 2010). Diabetes patients have good glycemic control if HbA1c (as the retrograde parameter of glycol regulation within the past 2 to 3 months) is lower than 7% (Nonso et al, 2015).

Diabetes is increasing rapidly in developed countries, the increase is widespread, because over 246 million people worldwide were treated from this chronic disease according to data from International Diabetes Foundation in 2005. Further report states that by 2025 more than 300 million people worldwide will have diabetes (Tadić, et. al., 2010). The exact number of people with diabetes mellitus is absolutely higher since by definite epidemiological study on each diagnosed patient there is one non-diagnosed patient (International Diabetes Foundation, 2010). A large number of patients with diabetes belong to the group of the working population.

Meanwhile, emphasis on the need for preventive method has been the interest of the health care, researchers, and medical personnel around the world. The preventive method that would support good planned assessment that could moderate or lower the risk of transitioning from pre-diabetes to outright diabetes. Ryman (1992), discovered that modification of diet and exercise patterns of people lower the occurrence of diabetes by 58%. Likewise, the management of pre-diabetes with drugs (Glitazones and Metformine)

lowers the likelihood of transitioning to diabetes by 25% to 49%. Diabetes mellitus is classified into four main type viz: "type1", "type 2", "other specific types" and gestational diabetes (Nonso et al., 2015).

In the early second half of the 20th century, clinical expert system has improved the decision making process in the field of medicine, and it has increased into which different medical knowledge have been built. Nevertheless, it should be mentioned here that the implementation of such system are associated with several problems. (Lehmann, 2004; Wrong, Legnini, & Whitmore, 2000). Akter, UddinShorif and Haque (2008) developed a computerized, low-cost and knowledge-based system that supports self-diagnosis and management of this life threatening disease for patients as well as doctors.

The needs for laboratory test results used in diagnosed diabetes consumes a lot of time which delayed patients treatment in starting their medications. Hence, there is need for the introduction of intelligence health care monitoring system; this will probably put an end to frequent speculation by diagnosticians. The development of artificial intelligence system used to overcome the conventional methods for the diagnosis of diseases. Artificial Intelligence has also been applied widely in most of the fields of computation studies. The ability of self-learning and self-predicting desired outputs has been the main feature of this concept. Supervised or an unsupervised are two ways by which learning can be achieved.

Radha and Rajagopalan (2007) implemented a “real-life experiment” in which the results confirmed the effectiveness of the fuzzy logic technique for diagnosis of diabetes. Several physicians know that his or her medical knowledge and experience consists of nearly 70% of uncertain data for example; symptoms, test analysis, and prognostic information among

others (Radha & Rajagopalan, 2007). In addition, data from literature shows that clinical ambiguities can be adequately well-illustrated by the fuzzy sets theory (Zimmermann, 1996, Pedrycy & Gomide, 1998, John & Innocent, 2005).

Fuzzy set theory is a valuable tool which can provide solution to three problematic areas of optimal therapeutic procedures, determining: imprecision, randomness and ambiguity (Singh, Cariappa, Gupta & Gupta, 2014). These, according to Kaur and Chakraborty (2007) enable us to imitate the reasoning process of human and make decision based on vague or imprecise data. Nonso et al, (2015) concluded that the utility of fuzzy set theory (especially in management system of diabetes) allow patients to make short term alternations (particularly lifestyle) to their overall schedule as required.

However, a large number of studies have shown cases the blood glucose control done in an exact way, by modeling mathematical application (Pedrycy & Gomide, 1998). Dazzi, et al (2001), described a model in which the blood glucose control is done by application of the fuzzy logic principle and neural networks techniques. It was revealed that the neuro-fuzzy control system is effective in improving the blood glucose control in unstable diabetes patients without increasing either the number of blood control determinations or the threat of hypoglycemia (Prasath, et. al., 2013).

Kahramanli and Allahverdi (2008) developed a new method for classification of data from medical database. This is to increase the reliability of the results obtained from the data. A hybrid neural network that comprises artificial neural network and fuzzy neural network was developed to determine the applicability of the proposed model and tested on real

data. The system crisp and fuzzy the value of medical data (Rajeswari & Vaithiyanathan, 2011).

In another work, Roudsari (2004) developed online diabetes management system (DiabNet) with online appointment and consultation. The system help in the needs of the individual patient by integrating appropriate historic stretch in blood glucose data and with the ability of including an adaptive capability, but it however, fails to measure the accuracy and effectiveness of the developed system.

Stelfox, et al. (2006) reported that about 44,000 people with DM die yearly in hospitals as a result of preventable medical errors. This was in addition to the delayed detection of clinical deterioration which results in high rates of avoidable in-hospital deaths (Marco et al., 2013). The most common errors include faulty systems, processes and conditions which inhibit clinical attendants to notice emergency situations in prior (Sumathi & Mudigonda, 2015). These errors can be avoided by designing health care monitoring systems and incorporating intelligence at all stages to make it more safe and reliable. For instance in Nigeria, many diabetic patients are faced with challenges ranging from insufficient number of diabetologists and endocrinologists, lack of appropriate medical infrastructure, among others. Nigeria still lack appropriate diagnostic system for diabetes (Akinjinmi, Adeyooye, Akingbade & Okerentugba, 2014).

This research, therefore proposes the development of an intelligent diagnostic system for diabetes using Genetic-Neuro-Fuzzy Inferential system. The system is aimed at providing a Fuzzy Inferential System (FIS) platform that can aid medical practitioners in administering accurate, timely, and cost effective diagnosis of diabetes in Nigeria. Thus,

the focus of this study is to develop an interactive genetic- neuro-fuzzy computation for the diagnosis of Diabetes Mellitus.

1.2 Statement of the Problem

One of the most dangerous life threatening diseases is Diabetes Mellitus as it represents both medical and socio-economic problem in modern society. In developed and developing countries the disease is widespread. It was estimated that 175 million people having diabetes in 2004, worldwide and it was projected that an estimate of 354 million will have diabetes by 2030 (Akinjinmi, et. al., 2014). According to the International Diabetes Foundation, 2010 Nigeria is one of the most populous countries in Africa that has the highest number of people with diabetes, with 3.0 million, South Africa 1.9 million, Ethiopia 1.4 million, and Kenya 769,000. Providing a solution for inadequate medical services using the education of human resources requires long time and high expenses that may result an increased morbidity rate in patients.

Artificial Intelligence (AI) in medicine has provided numerous advantages in diagnosis, management and prediction of highly complicated and uncertain diseases. Despite the high rate of complexity and uncertainty in this field, computational intelligent systems like artificial neural network (ANN), fuzzy logic (FL) and genetic algorithm (GA) have been used in order to improve health care, minimize treatment expenses and improve the quality of life (Lin, 2009; Mendes et al., 2001). A new category of systems is the Hybrid intelligent system which relies on the artificial intelligence techniques, such as fuzzy logic, neural network, expert system and genetic algorithms (Soha, 2015).

Using single techniques in diagnosis of diabetes has been comprehensively investigated showing some level of accuracy. Researchers have investigated the effect of hybridizing more than one technique to show enhanced results in the diagnosis of diabetes. However, using the combination of two or more technique for the treatment and diagnosis of diabetes patients has received less attention (Abdulla, Mujallid & Bakhsh, 2011; Singh & Singh, 2014; Priti, 2017). Hence there is need to research in to diabetes diagnosis to investigate the use of combining two or more techniques to diabetes diagnosis data to provide a better performance using the following evaluation metrics such as classification accuracy, specificity, sensitivity and Receiver Operating Characteristics (ROC).

In this work, three intelligent approaches are used to develop the system: fuzzy logic, neural network and Genetic Neuro-fuzzy techniques for diagnosis of DM. These approaches are combined in order to overcome the weakness in one approach while applying it and bringing out the strength of other approaches to find an optimal solution. It will describe the current status of fuzzy systems and neural networks in the diagnose of diabetes. This system will help people to diagnose if they are suffering from diabetes based on identified symptoms.

1.3 Aim and Objectives

The study aimed at developing a Genetic-Neuro-Fuzzy Inferential System (GNFIS) for diagnosis of diabetes.

The objectives are to:

- (i) develop an enhanced hybrid system for the diagnosis of diabetes mellitus using genetic algorithm;
- (ii) implement the enhanced system using Java programming language;

- (iii) evaluate the performance of the proposed system based on accuracy, sensitivity and specificity; and
- (iv) perform a comparison of the proposed system with two (2) existing systems Fuzzy Logic (FL) and Artificial Neural Network (ANN) for the diagnosis of diabetes mellitus.

1.4 Significance of the Study

Despite the fact that machine learning algorithms (MLA) such as Neural Network (NN), Fuzzy Logic (FL), Probabilistic Reasoning (PR), Support Vector Machines (SVM) and Evolutionary Computing, are techniques that lay emphasis on computational intelligence and provide the best techniques for learning, knowledge engineering, classification and searching, etc, available evidence had shown that the disease of Diabetes Mellitus has been on the increase across the globe, and one of the major causes of this rise is due to inaccurate diagnosis, management and poor treatment. Therefore there is need to develop a better and an improved fuzzy system using a unified Genetic-Neuro-fuzzy approach for accurate diagnosis of Diabetes Mellitus.

The hybridization of different Machine Learning methods has been proved useful for different jobs in medicine and other fields. Hence, there is need for an inferential technique that will integrate Genetic Algorithm, Fuzzy Logic and Neural Network components for handling uncertain and imprecise data for diagnosis of diabetes for self-learning and adaptive system. Genetic-Neuro-Fuzzy computing which is the combination of the merits of genetic algorithm, neural network and fuzzy logic, will enable one to build more intelligent system for decision-making. The concept of integrating genetic algorithm, neural network and fuzzy systems is to have an architecture that uses a fuzzy

system to represent knowledge in an interpretable manner and the learning ability of a neural network to optimize its parameters.

Hence, shortcomings of both of the individual approaches could be avoided. Integration can establish an interpretable model capable of learning and can use problem-specific prior knowledge. Consequently, Genetic-Neuro-fuzzy methods are especially suited for applications, where user interaction and interpretation is desired. This technique provides ease implementation and low cost solution benefit.

1.5 Scope of the Study

The proposed study develops a computational intelligent system for the diagnosis of DM using Genetic-Neuro-Fuzzy Inferential technique to provide a Fuzzy Inferential System (FIS) platform that can assist medical practitioners in administering accurate, timely, and cost effective diagnosis of DM. This system uses the diseases symptoms as inputs. The performance analysis and evaluation of the improvement of the patients' health is achieved. The proposed system will only focus on outcomes in diagnosis of DM although; Genetic-Neuro-fuzzy system has been applied for other disease such as tuberculosis (Omisore, Samuel & Atajeromavwo, 2017).

1.6 Definition of Operational Terms

Adaptive intelligent systems: A systems that is able to coordinate perception, reasoning and action to pursue multiple goals while functioning in a dynamically changing environment.

Blood Glucose Level or Blood Sugar Level: This is the amount of glucose present in the blood of human being and other animals

Carbohydrates: This is a food source of energy for human being, animal and plant. It is the main source of energy in our diet, and it includes all sugars and starches, for example: pasta, bread, rice and potatoes.

Data Object: is a logical cluster of all tables in the dataset that have one or more columns containing data related to the same business entity

Data Subject: this is an individual whom particular personal data is about. It is an individual who is the subject of personal data.

Decision support: This is the activity that is reinforced on models undoubtedly explained but not necessarily completely formalized; it helps in getting answers to the questions asked by an interviewer in a decision-making process. It can be built on methods like statistics, operations research, multi-criteria methods, etc.

Decision-making systems: This is an AI system that chooses one among many variants as a solution to a particular problem. The solution is then suggested to the user.

Defuzzification: This is an inferred membership function for a variable that is used in calculating a single output numerical value for a fuzzy output variable.

Diabetes mellitus: This is a disease that occurs when the amount of glucose (sugar) in the blood is too high in the body and the body cannot use it properly.

Diagnosis: This is a general term used in finding faults in a system.

Distributed representation: This is a system used in neural network, for encoding information where a value for a variable is represented by a collective activation of a group of neurons.

Expert system: A computer system created by using AI algorithms for solving difficult problems is called expert system. The system can provide a similar expertise to the one provided by human experts in a constrained area, for example diagnosis of malaria fever, finding a cheap route for a round-the-world trip, breast cancer, etc.

Fasting blood glucose: This is a test conducted in blood glucose and the test is usually taken in the morning before eating, and is used when diagnosing diabetes.

Fuzzification: The process is used for finding the membership degree $\mu_A(x')$ to which input value x' for a fuzzy variable x defined on a universe U belongs to a fuzzy set A defined on the same universe.

Fuzzy control: This is an application used to control a problem. It is an expert system called fuzzy system applied to solve a control problem.

Fuzzy expert system: The use of fuzzy logic to create an expert system is called fuzzy expert system. The expert system uses fuzzy rules, fuzzy data and fuzzy inference for designing the fuzzy expert systems, in addition to the standard ones implemented in the ordinary expert systems.

Fuzzy relations: The fuzzy relations make it possible to represent a vague or ambiguous relationship. It is used to link two fuzzy sets together by assigning a number between 0 and 1 to each element of the cross-product $U \times V$ of the universes of the two fuzzy sets.

Gestational diabetes: This is diabetes discovered usually during the second or third trimester in pregnancy patient. This occurs when the body does not produce enough insulin to meet the extra needs of pregnancy.

Glucagon: The hormone raises the blood glucose level of a person and it is produced in the pancreas in the alpha cells in islets of Langerhans. To reverse a severe glucagon, hypo glucagon can be given as injection.

Glucose: This occurs from taking starchy foods, sugar and sweet foods, and from the liver that produce glucose. Examples of such starchy foods are bread, rice, plantain, Amala, Fufu etc.

Glycaemic index (GI): The term is used to grade the foods according to the level by which they affect blood glucose levels.

Inference: It is a term used to match current data from existing knowledge in a knowledge-based information system and inferring new facts until a solution is reached in the solution.

Medical decision support system: This is a computer software created for the purpose of providing physicians with useful, quick and useful information unfolding the clinical conditions of the patient and appropriate knowledge of this situation, properly filtered and presented to improve the quality of care and patient's health.

Membership function: It is a universal feature function that describes the degree by which an object from a universe belongs to a fuzzy set.

Parallel Computing: A kind of computation in which many calculations or the execution of processes are carried out simultaneously.

Type 1: The diabetes always appears in a patient before the age of 40 and it happen when the body is unable to produce any insulin. This type of diabetes is usually treated with a healthy diet, insulin injections and regular physical activity.

Type 2: This type of diabetes develops if patient does not respond properly to the insulin produced in the body, when it is not enough, or when the insulin that is produced does not work properly. Type 2 diabetes usually appears in patients that are over 40 years and above.

1.7 Organization of the Thesis

The remaining chapters of this thesis contain the following parts:

Chapter two discusses the literature review, where relevant subjects of the research topic were comprehensively discussed. Chapter three is the Research Methodology which mentions the method and evaluation metric used to achieve the objective of the study. Results and Discussion were carried out in chapter four. Summary, recommendation and conclusion were discussed in chapter five.

CHAPTER TWO

LITERATURE REVIEW

This chapter reviews relevant literatures to understand some basic concepts and related studies in Computational Intelligence Techniques (CIT) employed in diabetes disease diagnosis, which includes Fuzzy logic approach, Artificial Neural Networks, Neuro-Fuzzy System and Genetic Algorithm.

2.1 Computational Intelligent Techniques (CIT)

Different types of intelligent system techniques have been applied to predict, classify and diagnose diabetes and its complications. Many of these works employ Fuzzy logic (FL), Artificial Neural Networks (ANN), Genetic algorithm (GA) or combination of these techniques to build Expert System (ES). For instance, Kalpana and Senthil (2012) used fuzzy logic to design an expert system for the diagnosis of diabetes with verdict mechanism. Triangular membership functions with Mamdani's inference are used in fuzzy verdict mechanism. The conversion of fuzzy values into crisp values was done using Defuzzification method. Verdict mechanism was used to execute rules in order to make a decision on the possibility of individuals suffering from diabetes and to present the knowledge with descriptions. The results of the experiment show that the expert system can transfer the acquired information into the knowledge to simulate the thinking process of humans and can analyze data from the database (Omisore, et al, 2017).

Acar, et al, (2001) forecasting diabetic mellitus using ANN. The methods used 768 data from PIMA Indian database for learning and testing where 268 patients were diagnosed with diabetes using back-propagation algorithm, to adjust the weights until the error measured between the desired output and the actual output is reduced; the input data was trained. When the sufficiently low value is reached the training stop. MATLAB is used to analyze the data with Neural network toolbox. The evaluations were compared in term of error with eight inputs and four outputs results. The system helps to provide solutions in determining if a patient is suffering from diabetes or not, the system is much easier to determine if a patient suffers from diabetes or not rather than using the blood test.

Kavitha and Duraiswamy (2011) developed an automated classifier system for fundus images hold diabetic retinopathy. An Adaptive Neuro-Fuzzy Inference System (ANFIS) is an expert system used to determine a normal or non-proliferative eye. Back propagation with Least Squares method was used to train the system. Since the system provides best classification, it was used as a screening tool to the analysis and diagnosis of retinal images.

To improve the diagnostic accuracy of diabetes disease, Harun (2010) used adaptive neuro-fuzzy inference system (ANFIS) and principal component analysis (PCA) for diabetes diagnosis. He divided the system into two stages, the first stage of 8 features dataset is reduced to 4 features by PCA and in stage two the reduced dataset was used to diagnose patients with diabetes disease using the designed system classifier (adaptive neuro-fuzzy inference system). The proposed system recorded 89.47% classification accuracy.

Tejashri and Satish (2015) proposed fuzzy logic expert system in order to improve on Pima diabetes dataset classification accuracy. Artificial neural network (ANN) with two neural networks were combined to form a hybrid system, back propagation algorithm was used to train the proposed system. Fuzzy like age, blood pressure and the rest were used as crisp data after dividing the inputs into two groups. The standardized crisp input values are fed to the first ANN in the first stage (ANN1) and the result of the defuzzified from fuzzing fuzzy data and their values are presented to FNN. To calculate the final output, the results from ANN1 and FNN were fed to the second ANN (ANN2). If the output value is different from the actual value, the weights of these networks will change, and the process is repeated until reasonable results are reached. The designed system was evaluated using K-fold cross validation and an accuracy of 84.24%.was recorded.

The proposed hybrid intelligent system used the fuzzy expert system in addition to the neural network base. The inputs have been separated into a couple of groups: fuzzy such as blood pressure and medical tests and rest are deemed to be crisp data. Fuzzy system is applied to integrate the fuzzy inputs then feeding them to ANN together with the crisp inputs. ANN has been used for the prediction of DM.

2.2 Basic Concepts in Medical Informatics

Clinical knowledge is the application and study of methods of improving patient data, community population data, health, medical informatics, and other information relevant to patient care. The application is a young science that emerged after the invention of digital computer in 1940s. The mechanical computing in medicine by the 19th century had a much earlier origin. The punched-card originally used for the US census is a data processing system that was later improved to support surveys in public health and

epidemiology developed by Herman Hollerith's (Shortliffe & Blois, 2001). This shows the multidisciplinary nature of the system in medical informatics, which include the clinical sciences, cognitive, computing, information, epidemiology, and health services research as examples of public health sciences. The terminologies used in medical informatics backgrounds can be confusing to workers and starters in the field. It is very obvious that the basic introduction of new concepts would be of use to those interested in learning more about medical informatics.

Public health informatics, consumer health informatics, and clinical informatics are new disciplines that emerged in recent years with different branches. It has advanced to the transmission of medical information to storing, structuring, and processing that are used for different purposes (Shortliffe & Blois, 2001). One of the purpose is the development of decision-support systems to enhance the human ability to diagnose, treat, and assess prognosis of pathologic conditions. Personalized diagnosis, treatment, and prognosis all essential parts of good health care will still be proved difficult as classification tasks with population unpredictability, even if disease processes were fully understood. This shows that the population variability was not fully taken into consideration in many decision-making and that diseases are not fully understood. The basic principles learned in the basic and clinical sciences to determine whether a patient has a given disease is not always enough for a clinician to know whether patients should be given a certain treatment, and how long he or she will survive.

The serious and intensive collaboration between different fields of study like computer scientists, physicians, electrical engineers and mathematicians gave birth to the history of computerized medical diagnosis. Researchers have works intensively by studied and wrote

on the use of card and needle systems for classification and storage of medical data and systematic medical decision-making in the late 1950s. However, the development of different techniques to diagnosis using Bayes' rule factor analysis and decision analysis in the 1960s and 1970s (Ledley, 1960; Woodbury, 1963 and Überla, 1965). Artificial intelligence approaches to simulate the physicians reasoning in information gathering and diagnosis using databases in form of networks of symptoms and signs for diagnosis are Present Illness Program (PIP) and Diagnostic Logic (DIALOG) and many others (Pople, Myers and Miller, 1975; Parker, Gorry, Kassiner and Schwartz, 1976).

The signs, information about patient's state of health, laboratory test results of different forms, medical history of patient and X-ray findings are termed as symptom. To find a list of diagnostic possibilities for a patient, information from the symptom will be used by a physician. The physicians have to master this process by studying excluding symptoms for diagnosis, relationships of obligatory or facultative proving in books and journals plus his practical experience. The information about relationships that exist between symptoms and symptoms, symptoms and diagnosis, diagnosis and diagnosis and more complex relationships of combinations of symptoms and diagnosis to a symptom or diagnosis formalizations are called medical knowledge.

Perez-Ojeda (1976) used medical knowledge called network linked by logical relations to create a prototype system used to search for an adequate strategy to simulate an approximate reasoning model in medical decision making. This is shown in figure 2.1.

- A common cold is always associated with a running nose.
- "A bladder with irritation and infection occur from acute pyelonephritis."

- “Occasionally fever or chills, and malaise occur from acute pyelonephritis.”

The common cold and acute pyelonephritis diseases are abbreviated by D1 and D2 and runny nose, fever, bladder irritation, infection, chills, and malaise are represented by S1 to S6 respectively. This is graphically represented and constructed by elementary knots and arcs. However, a mathematical probability modifier (usually, occasionally, and almost always) was used to model the relations by the author:

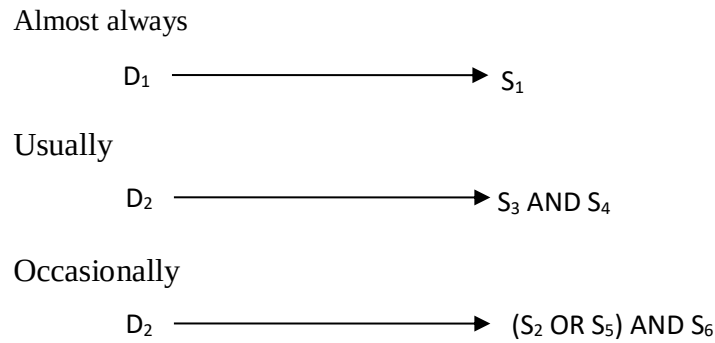


Figure 2.1: Examples of the network of medical knowledge (Perez-Ojeda, 1976)

2.3 Basic Concept of Fuzzy Logic Approach

Fuzzy logic (FL) is used to simulate human reasoning and to deal with uncertainty in knowledge. It is one of the AI techniques that are most applicable in an incomplete or fuzzy data. Fuzzy logic is an input dataset of nonlinear mapping used to a scalar output dataset; this is aimed at providing approximate reasoning. Pattern recognition, and decision making methods in this modern information technology has attracted growing attention among others methods (Tejashri & Satish, 2015). Fuzzy logic theory has the mathematical strength to capture uncertainties associated with human cognitive processes, such as thinking and reasoning. It is the most suitable and applicable method to develop knowledge based systems in different sectors of life especially in health sectors. The

method is very useful and suitable for real-time monitoring of patient data, interpret sets of medical findings; syndrome differentiation in eastern medicine and diagnosis of diseases in western medicine (Ephzibah, 2011).

Fuzzy logic is an example of structured numerical estimators (Lin, 2009). It used Boolean logic to grip the concept of partial truth by taking a value between a complete true value and a complete false value. The variables like “not very true, true, more true, false or less false” could be used for the partial truth. Fuzzy *sets* have to be introduced in order to accomplish the notion of partial truth. The process is the collection of objects that might belong to the set of degree that take any values between 0 and 1 represent (full non-belongingness) and (full belongingness) respectively, instead of taking a crisp value (0 or 1).

To deal with problems associated with uncertainties and imprecise information, fuzzy logic is the most suitable method. However, there are difference between uncertainty due to randomness and uncertainty due to the imprecision. Exactness in information and absence of a sharp boundary is called imprecision, and randomness is the occurrence of the event. The membership function is the intensity of belongingness assigning each element a number from the unit interval [0, 1]. The universal set is presented by X and the subset of X is represented by A , if A is a set of ordered pairs.

$$A = \{x, \mu_A(x): x \in X, \mu_A(x) \in [0,1]\} \quad (2.1)$$

The membership function of A is represented by μ_A and the grade of the membership of x in A is represented by $\mu_A(x)$. For example, animal with 5 years of age can be adult with a truth value of $\mu_A(x) = 0.9$, if the set of adult animal is fuzzy when the crisp numbers are

fuzzified. Some of the examples of membership functions shape are triangular, trapezoidal, bell-shaped etc. Membership Function choosing depends on the area of application or the controlling system. The most commonly used among them are triangular and trapezoidal because of their simplicity and more flexibility, but it is important to check the input data and controlling system before choosing membership function generators. Using triangular and trapezoidal membership for calculations are very easy, and shapes are simple and more flexible with fewer complexities when splitting values (low, med and high MF) or comparing other membership functions. Triangular and trapezoidal member function is show in equation 2.2 and 2.3 as follows;

$$\mu(x) = \begin{cases} 0, x < a \\ \frac{x-a}{b-a}, a < x \leq b \\ \frac{c-x}{c-b}, b < x < c \\ 0, x > c \end{cases} \quad \text{Triangular membership functions} \quad (2.2)$$

$$\text{trapezoid}(x: a, b, c, d) = \begin{cases} 0, x \leq a \\ \frac{x-a}{b-a}, a \leq x \leq b \\ 1, b \leq x \leq c \\ \frac{d-x}{d-c}, c \leq x \leq d \\ 0, d \leq x \end{cases} \quad \text{trapezoidal membership functions} \quad (2.3)$$

Operations like complement, intersection and union are some of the basic mathematical operations performed on fuzzy set. The rules such as if-then statement consists of arguments joined with logical operators performed on fuzzy set are verbally formulated, THEN the result has to be true IF the condition is true. The Boolean expression like AND,

OR, NOT and XOR are usually used formulated by logical expression. The truth value corresponding to the fulfillment rule conditions is the Degree of Fulfillment (DoF) for a given premise. The product and min-max inferences are used mostly to determine DoF, and the rules will be represented with combinations. Minimum, maximum and additive combinations are the combination used. The minimum technique at least remains a certain level in DoF with all the applicable rules tries to find a rule response. Considering the rule responses $B_1=(0,2,4)$ with DoF $v_1=0.4$ and $B_2=(3,4,5)$ with DoF $v_2=0.5$, the minimum combination of responses (B_i, v_i) is the fuzzy set B with the membership function:

$$\mu_B(x) = \min_{v_i > 0} v_i \times \mu_{B_i}(x) \quad (2.4)$$

Membership functions of x in fuzzy set B for rule i is represented by $\mu_B(x)$. B_1 and B_2 has intersection $[3, 4]$. The equation taking the minimum of two membership functions on $[3, 4]$ will be as follows:

$$\min\{0.4 \frac{4-x}{2}, 0.5 (x-3)\} \quad (2.5)$$

which results in:

$$0.4 \frac{4-x}{2} = 0.5 (x-3) \quad (2.6)$$

From the above equation, the value of $x = 23/7$. Thus, the minimum of membership function is:

$$\mu_B(x) = \begin{cases} x-3, & \text{if } 3 \leq x \leq \frac{10}{3} \\ \frac{4-x}{2}, & \text{if } \frac{10}{3} \leq x \leq 4 \end{cases} \quad (2.7)$$

From the graphical representation, the thick line represents the minimum combination while the thin lines represents the individual rule consequences as shown in Figure 2.3

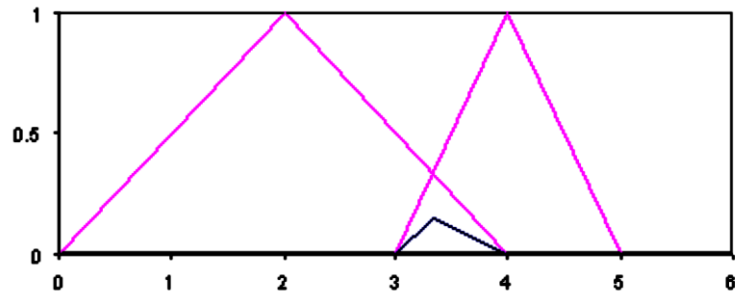


Figure 2.2: Minimum combination for consequence B membership functions (Duckstein Blinowska & Verroust, 1995)

To form a control algorithm a collection of fuzzy rules can be used. The basic four principal components that formed the fuzzy rule based system (FRBS) are as follows:

- (i) *Fuzzification Interface:* It is used to measure the inputs, fuzzified and mapped into the suitable universe of discourse.
- (ii) *Knowledge-base:* It is the fuzzy linguistic rule based section and the numeric database section.
- (iii) *Fuzzy Inference:* It is the fuzzing reason like product, max-min composition, the mechanism or engine that constitutes the core of the Fuzzy Logic Components.
- (iv) *Defuzzification Interface:* It is used to map the range of output variables into the corresponding universe of discourse then defuzzifies the results of fuzzy inference engine. This can be *maximum* of the rule consequence, *mean* and *median* methods (Duckstein, et al, 1995).

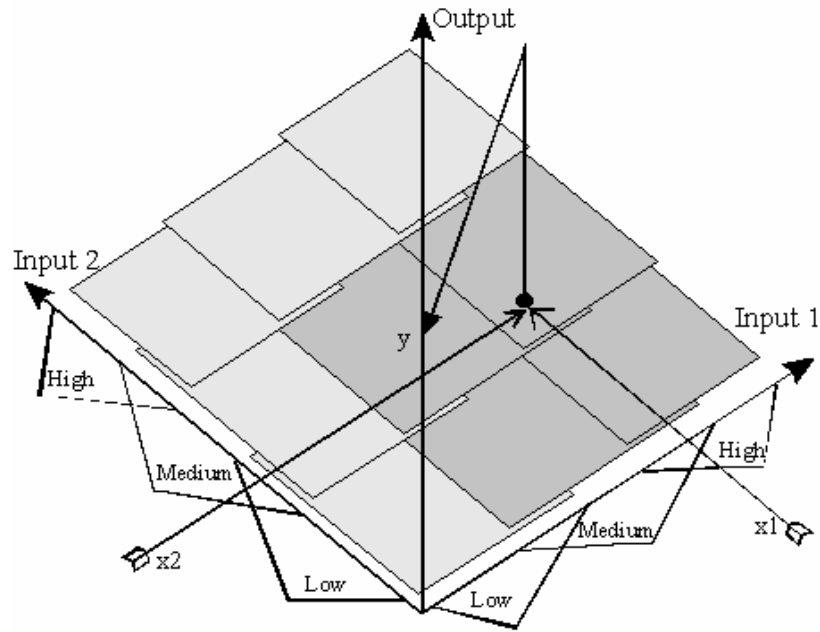


Figure 2.3 Principle of FRBS responses (Abebe, Solomatine & Venneker, 2000)

Figure 2.3 shows an example of a simple FRBS designed by Abebe et al with two inputs and one output using the principle of fuzzy rule base response (Abebe, et al., 2000). Input x_1 fits into *high* and *medium* and input x_2 fits to the *low* and *medium* membership functions with a different degree at the same time due to overlapping. Out of nine rules constructed, two inputs were activated from three input membership functions. The defuzzified numerical output y is the fuzzy inference results that was DoF determined.

2.3.1 Practical Problems in Medicine Applications for Fuzzy Logic

The belief that difficult clinical problems might yield mathematical formalism was first proposed in the 1960s in internal medicine. This is a classic field of research in computer-aided diagnosis. In internal medicine, pulmonology, gastroenterology, rheumatology and hematology are examples of some its main areas. MYCIN system is the first developed knowledge based expert system after the introduction of fuzzy logic by Zadeh. The designing and turning of fuzzy rule based system for medical diagnosis was discussed by

Rothstein. To aid the diagnosis of pulmonary embolism using ventilation perfusion scan and correlated chest X-rays, an inference system using fuzzy logic was developed. Mamdani fuzzy model was employed to implement the inference system (Mahfouf, Abbod & Linkens, 2001).

2.3.2 Applications of Fuzzy Logic in Diabetes

To prevent hypoglycemia and minimize high glucose peaks, MD-Logic Artificial Pancreas System (MDLAP) was developed. The system is a promising tool for individual glucose control in patients with type 1 diabetes (Radha & Rajagopalan, 2007). Fuzzy logic controller was proposed to maintain the normoglycaemic for diabetic patient of type I (Mahfouf, et al, 2001). The expert system is an artificial intelligence designed to give decision support to patients and physicians, it is also a telemedical monitoring platform. The system provides risk analysis among other important functions and control of the algorithms that represent the core of complex global agent for diabetes care (Kalpana & Senthil, 2012).

In pain relief control, similar nonlinear models with fuzzy measures can be used to determine the parameter that describes the dependence of the pain relief on the applied stimulation. Fuzzy measures was used to generate determination for a given pain distribution of the optimal pain relief stimulation (Tejashri & Satish, 2015). Since the diagnosis and treatment of clinical stroke is unique to individual patient, it can be captured by a scientific approach that is not only represented but also measures the changing causal role of known and unknown patient context in determining his/her condition (Kalpana & Senthil, 2012).

2.3.3 Drawbacks of Fuzzy Logic

To handle uncertainty and imprecision in real situation fuzzy system have been proved to be important and very useful, but the systems do not provide learning ability, which is one of its major disadvantage. The algorithms also have the problem of inability to adaption characteristics. Fuzzy systems have the following disadvantages in addition to the aforementioned limitations:

- (i) Inability to provide self- learning capability;
- (ii) Inability to provide adaption capability;
- (iii) Inability to provide parallel computation;
- (iv) It has the problem of optimization; and
- (v) Fuzzy logic lack the power to strengthen solution once achieved.

2.4 Artificial Intelligence

The enactment of Artificial Intelligence (AI) in term of research in the last two decades on both manufacturing and service systems has greatly improved their performance (Lin, 2009). AI, as it was first named by John McCarthy in the fifties, has been proved in the areas of the ‘hows’ and the ‘whys’ of human intelligence; however, this system has become useful research tools in virtually all fields including science, education, medicine and engineering, and also its applications in law, stock market, marketing and accounting, among others (Mendes et al., 2001). It is a system positioned by machines to handle complex imprecise tasks that require intelligence if done by humans. But AI has the following problems which include reasoning, natural language understanding, machine learning, programming, artificial life, knowledge representation, belief revision, and theory of computation (Tadic, Cvjetkovic & Milovanovic, 2009; Engelmores &

Feigenbaum, 2007). AI has achieved greatly in the practical application, despite some hindrances; its success is being reviewed with the commercial success in expert system (Kuncheva & Steimann, 1999).

2.4.1 Neural Network

The interconnected artificial neurons which are the properties of biological neurons were mimicked by Neural Network (NN). The algorithm follows analog and parallel computing system made up of simple processing elements. It communicates through a rich set of interconnections with erratic contributory weights. Artificial Neural Network (ANN) mimicked the functions of the human brain, it is a synthetic nervous systems (Zhang et al., 2008). It attempts to abstract the complexity of biological nervous system to focus on what may hypothetically matter most from an information processing point of view.

The cell body in the human neuron receives incoming impulses by means of chemical processes through dendrites (receiver). ANN work, mimic and performs basically like a human brain. The neuron will release it off to other neurons through its synapses if the number of incoming impulses exceed certain threshold value. This determines the impulse frequency to be fired off (Laleh, Ali, Shahriar and Fatollah, 2008).

Therefore, the neurons of an ANN or its processing units consist of three main parts; the summation function within the node, synaptic weights connecting the nodes, and the transfer function. Synaptic weights distinguish themselves and agree to the importance of the information coming from each neuron with their strength (value). This is the information encoded in strength-weights. To calculate a total input signal, the summation function is used by multiplying their synaptic weights and summing up all the products.

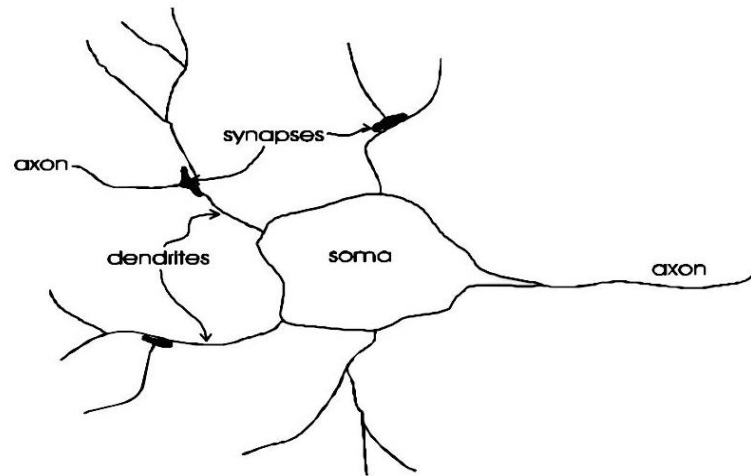


Figure 2.4: Schematization of Natural neuron (Elson, 2002)

The summed up input signal received from the summation function into an output is called activation function or the threshold function. This can be either linear or non-linear, which characterizes the neural network.

2.4.2 Artificial Neural Networks

An information-processing system inspired by the structure of the human brain is called artificial neural network. Soma, dendrites and axon forms the biological neuron as shown in figure 2.5. The cell body is called Soma, dendrites is the number of spine extensions and through synapse, axon twigs connects to the dendrites of other neurons and out from the soma (Nonso, et al., 2015).

ANNs consist of a number of interconnected neurons inspired by the natural neuron in the brain. A neuron is an information-processing unit that receives several signals from its input links and it has a weight assigned to it. In biological neurons synaptic efficiency corresponds to these weights. The basic means of the long term memory in ANNs are the weights (Sreedevi and Padmavathamma, 2015). The initial weights of the network are set

to random numbers and thereafter corrections of the weights are performed from a set of training examples.

The transfer function is used to transform the activation level of a neuron into an output signal. The activation function and weights for the neuron specified with each layer has its own transfer function enhanced by the behaviour of the ANN. Some examples of activation function are sign, sigmoid, linear and step function (Ryman, 1992). Activation functions of the neural networks are defined as follows in Figure 2.6:

Linear

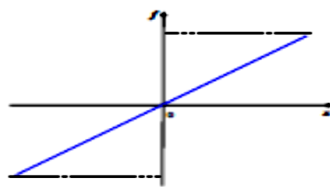
$$f(x) = x \quad (2.8)$$

Sigmoid

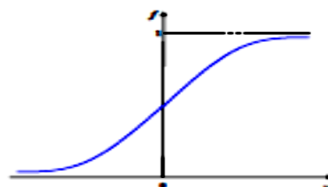
$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.9)$$

Hyperbolic tangent

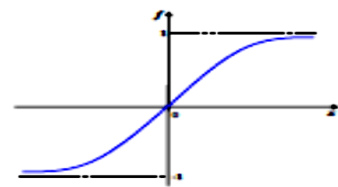
$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.10)$$



(a) Linear function



(b) sigmoid function



(c) hyperbolic tangent function

Figure 2.5: Activation functions

2.4.3 Mathematical Background

Neural network was formed with series nodes called neurons that are organized in layers. Each neuron in a layer is connected through a weighted connection with each neuron in the next layer. The w_{ij} is the weight value which specifies the strength of the connection in i -th neuron in the i th layer and the j -th neuron in the next layer. The input layer of one or more hidden layers and the output layer formed the structure of a neural network. The number of neurons in a layer and the number of layers shaped the complexity of the system studied. Therefore, the complexity of this system determined the optimal network architecture. In figure 2.6, three-layered ANN architecture general scheme is given.

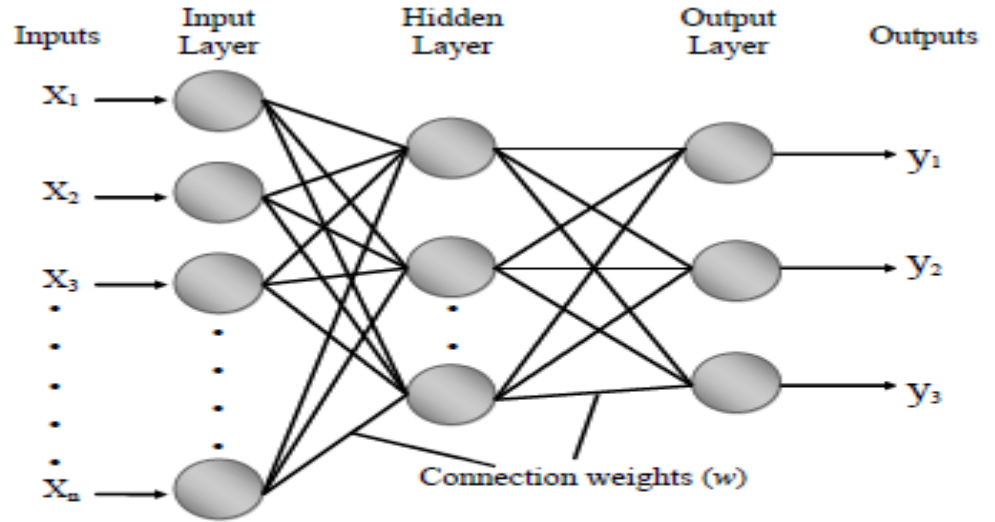


Figure 2.6: Back-propagation ANN (Hopfield, 1982 cited in Pradhan et al, 2012)

The network's output was formed from the neurons in the last layer. According to Equation 2.10, calculate the weighted sum and add a bias term (θ_j), and the j -th neuron in a hidden layer processes the incoming data (x_i). The neurons in the first hidden layer through the weighted links receive data transfer from the neurons in the input. Here, the result is transferred to the neurons in the next layer after the data are mathematically processed.

$$net_j = \sum_{i=1}^m x_i \times w_{ij} + \theta_j \quad (j=1,2,\dots,n) \quad 2.11$$

Learning algorithm used strongly determined the structure of neurons. Learning is when the weight is adjusted and until threshold values of certain criteria are encountered. For example if the training error falls below some user-defined level, then it means that epochs or cycles exceeds a certain value or all the validation tests have been performed (Singh, Rahul, & Yaduvir, 2011). There are three learning types that can be performed on neural network architecture namely: Supervised, unsupervised and reinforcement learning.

Supervised learning is a process of training artificial neural network and it provides steps to follow and concrete learning results from such supervision. It gives dataset with input vectors with desired output associated with each input vector. From the derived algorithms, to minimize the error of the desired output, the weights can be suitably adjusted from the real output of the neuron. Each set of the training vector is the desired output that serve as the representative of each set of the training vector (Pradhan et al, 2012). Back-propagation algorithm, used in artificial neural network and most popular training techniques is an example of supervised learning. The technique used at least three layers of units. The first is an input layer, with at least one hidden layer and an output layer. The output from the input layer are connected as an input into the hidden layer.

Similarly, the output from the hidden layer is connected as an input into the output layer to produce the final output of the ANN. In back propagation the network propagates the input pattern from layer to layer until the output pattern is generated by the output layer. An example of back propagation is shown in figure 2.7. If there is any difference between actual and desired output pattern, the weights are adjusted to minimize the error, and then

propagated backwards through the network from the output layer to the input (Mohammad, et al, 2013).

Unsupervised learning is a training model where the network is only given a set of input data; the learning network must search out shared properties of the dataset, connect these resemblances and learn how to classify input data into appropriate categories. Each input vector involves weight modifications without representation of the desired output. Some of the examples of unsupervised learning are Discrete Hopfield and Kohonen self-organizing map (Pradhan et al, 2012).

The reinforcement learning is used to reward the neuron for good performance and penalize neuron for bad performance. The technique is used to adjust the parameters until an equilibrium state is achieved after which there will be no more changes in its parameters. The learning algorithm is mainly used for control applications (Mohammad, et al, 2013).

ANNs have been applied in different fields like medicine, industrial business, manufacturing and robotics etc. the technique is proved useful in all these areas as solutions to different problems such as image processing, pattern recognition, signal processing, neurological and cognitive modeling, classification of text, forecasting, decision making, speech recognition, and controllers. Since late 1980s ANNs have been used in the field of medicine, the algorithm was first used for the probability of disease-free survival, to identify accuracy of survival prediction, or the time of recurrence of disease after surgery (Delen et al. 2005; Rama & Nagaveni, 2010; Vithyatheri, Vimala & Huck-Soo, 2012).

2.4.4 Overview of ANNs in Diabetes

In health care delivery, neural networks are widely used for a range of different purposes. In diabetic disease, it has been proved useful and applied in the area of diagnosis, treatment, classification and management of the disease. (Rao et al. 2002) developed a prototype and used neural network to measure the quality of life in diabetic patients with variables such as age, gender, weight, or fasting plasma glucose as input data. These variables are biological or biographical attributes used in measuring the level of diabetes in patient. The evaluation of the results were performed and compared with those from traditional statistical methods. Tudor and colleagues hybridized ANNs and nonlinear model predictive control as a control strategy. The proposed neural predictive control simulation was used on glucose concentration in subcutaneous tissue (Tudor et. al., 1998). In the said approach large noise levels and time delays were used and allowed. The result showed that administration of insulin or analogues was inappropriate for rapid control of diabetes in patient, but was suitable for controlling slow perturbations.

Nidhi & Jha, (2014) developed a neuro-fuzzy control system combined both fuzzy logic and neural networks algorithms using minimal number of input data for correct output. The combined algorithms proved suitable for the risk of hypoglycemia, reduced the need for increased blood glucose testing and direct control of diabetic patients on intravenous nutrition. As a preliminary study in patients with type I diabetes, inverse optimal neural control was used to monitor the blood glucose level (Rahman et al., 2015). Recurrent neural network was used to avoid hypoglycemia and hyperglycemia and to control the level of glucose and insulin following a desired trajectory. ANNs has also been used in the diagnosis of diabetes and not only in the tracking of glucose. Neural network was used to

develop an expert system to predict fatal hypoglycemia episodes in type I diabetes patients (Chan et al. 2011).

To evaluate the proposed system the records of 430 patients were collected based on the following features (i) rate of change of QT interval of electrocardiogram signal (ii) physiological parameters (iii) rate of change of heart rate and (iv) rate of change of corrected QT interval. The network was trained using 320 patients and the remaining 110 patients for verification. Comparing the system with fuzzy logic, statistical or fuzzy regression methods, and the system performed better with a sensitivity of 79.30% and specificity of 60.53%. The application, in conclusion has improved the accuracy, sensitivity, and specificity of diabetes patient compared to other methods, thus contributing to the betterment and improvement to clinical diagnosis, treatment, and management of diabetes mellitus.

2.4.5 Drawbacks of ANNs

- (i) Neural networks can be used if training data is available because it is not necessary to have a mathematical model to the problem of interest, since there is no need to provide any form of prior knowledge.
- (ii) To interpret the learning process of the solution produced from neural network is difficult.
- (iii) The final state of ANN cannot be interpreted in terms of rules and their solution cannot be checked whether it is plausible and their architectures are black boxes.
- (iv) Since neural network cannot be initialized with prior knowledge even if it is available, then the network must learn from the scratch.

- (v) The learning process usually takes a long time and process, since there is usually no guarantee of success.

2.4.6 ANN Approach in Diabetes Mellitus Prediction

Different approaches have been used in research works that uses ANN in diabetes diagnosis, treatment, management and prediction. Novruz (2014) used ANN to improve classification accuracy of diabetes mellitus diagnosis. Adeli & Mehdi, (2010) combined both fuzzy logic in form of computational intelligence and artificial neural network with case-based reasoning to enhance the ability of diagnosis and an early prediction of the diabetes mellitus. Adeli & Mehdi, (2010) combined ANN and rule-based algorithm to classify and predict specific data. The expert system improves the accuracy of the prediction results on each approach.

Sikchi, Sikchi and Ali (2013) performed a comparison between the performance of ANN prediction model, decision tree and logistic regression using 2 risk factors. The result from the study shows that decision tree algorithm produces the highest level of classification accuracy. Sumathy et al., (2010) diagnosed diabetes using designed expert system based on the risk factors. The proposed system eased patients who went through a number of medical tests as they could perform the tests by themselves thereby saving time. Artificial Neural Network (ANN) architecture was used for classification with back propagation learning algorithm that had a supervised multilayer feed forward network. The inputs to the system were the symptoms which appeared in diabetic patients. About 75 datasets were used for the work. When compared with the existing methods, the ANN technique gave better results (Sumathy et al, 2010).

Kumar and Kumaravel (2016) used different supervised learning algorithms of ANN for diabetes diagnosis. To train the network, patients between ages 35 to 78 year of 250 diabetes patients were used. Regression analysis was used on algorithm to further compare their performances while the performance accuracy of the each algorithm is computed to validate accurate prediction of each algorithms.

Baha and Wajiga (2012) combined ANN and risk factors with 7 risk factors and their strength of association to the development of Type 2 diabetes as relative weight of input variables. Back propagation algorithm with a multilayer feed forward architecture was designed using MATLAB. Batch mode back propagation with gradient descent and momentum was used to train the system. During the training the best performed networks identified were 2 hidden layers of 6 and 3 neurons, an output layer of 1 neuron, log sigmoid transfer function at the hidden layers and a linear transfer function at the output layer. At least 528 out of 1122 of the dataset used were found close to 1, which indicated high risk of Type 2 diabetes. The performance evaluation of the network recorded best validation of 0.10054 with correlation coefficient of 0.99705at 552th epoch. The graph of regression plot specified exact linear relationships with all the axes close to 1.

Ali (2008) developed an expert system using ANN algorithm to determine whether someone is a diabetes patient. Eight (8) inputs, three (3) hidden layers and one (1) output layer was used to train the system and result shown that 268 out of 768 patients diagnosed with diabetes with high performance prediction. Rao et al. (2002) with the use of a secondary data of 241 diabetic patients developed and trained expert system using ANN in predicting the well-being of patients living with diabetes. Java Swing Package using JDK 1.5 was used to implement ANN multilayer perception. Lavanya, Durai, & Sriraman (2011)

designed expert system using neural network with the aim to estimate the posterior probability for medical diagnosis and the probabilities of estimating diabetes mellitus.

The system proved that neural networks can be used to estimate posterior probabilities from the obtained result, comparing with discriminant analysis and logistic regression that are also traditional statistical procedures, but neural network estimates are distribution-free and direct. Al-Shayea (2011) proposed an expert system using ANN to classify type-2 diabetes patients of Pima Indian Diabetes dataset and study the impact of preprocessing for diagnosis of diabetes. The result when compared with other existing works shows tremendous improvements in accuracy when combined with principal component analysis preprocessing, and the use of replace with k-nearest neighbor and replace with mean.

Baha & Wajiga (2012) used ANN with analytical hierarchy process as an input to determine the relative severity of risk factors associated with Type 2 diabetes using analytical hierarchy process as input to AN, degree of relative weights of 7 risk factors was used by the authors. Structured Likert format with 6 choices was used by professionals to test the degree of severity of the risk factors. Performance evaluation of the system showed that the highest contributor of the risk is Heredity with 0.5388, followed by Obesity, Age, IGT and Gestational diabetes each contributes 0.1038, Diet and Physical inactivity each contributes 0.0230. Baha & Wajiga (2012) work was adopted by the result of the research as inputs to the neural network.

Different types of AI algorithms such as neuro-fuzzy systems, genetic fuzzy systems and genetic programming neural networks etc. can be hybrid or combined to design an expert system. The combination of ANN and fuzzy logic is referred to as Neuro-fuzzy systems

(Jang, Sun & Mizutani, 1996). The system combined human-like reasoning using fuzzy logic algorithm with learning and connectionist of neural network, which form the basic idea behind this NFS. The ability to explore interpretable IF-THEN rules, NFS provides flexible and powerful collective approximations. In many sector in our social and technological life NFS is thriving in the creation of AI.

2.5 Adaptive Neuro-Fuzzy Inference System

Fuzzy logic inference engine deal with linguistic expressions which make it distinct from all AI and neural network that use self-learn and self-improve on dataset as an advantage over other AI. Jang et al, (1996) combined the advantages of fuzzy logic and neural network to develop an Adaptive Neuro-Fuzzy Inference System (ANFIS). The system was used to enhance the learning ability derived from a neural network with fuzzy logic. This was done to adjust the membership functions and linguistic rules directly from data to enhance the system performance and represent knowledge in an interpretable manner (Wang, Amy, & Noémie, 2008). Five layer feed forward neural network was used for the system architecture. ANFIS implements a Sugeno fuzzy inference system to generating fuzzy rules from a given input output dataset for a systematic approach (Sreedevi & Padmavathamma, 2015).

Jang et al combined gradient descent and least squares estimator (LSE) to update the parameters of neural networks. A set of premise parameters and consequent parameters were used by the ANFIS system (Jang, et al, 1996). The system used alternative algorithm to reduce the error of the training network. The advantage of using the two algorithms is for an effective search for optimal parameters and to further speed up the learning process. The same authors proposed another method using Levenberg-Marquardt (LM) for ANFIS

learning in 1996, the result from system was compared with the first proposed hybridized ANFIS system.

The Levenberg-Marquardt (LM) gave a better Root Mean Squared Error (RMSE) after the predetermined iterations, but to assign appropriate linguistic interpretations became difficult and the final membership functions varied from the initial setting. To update the antecedent of the ANFIS network, many other algorithms have been employed (Chen, Li, Lu & Zhou, 2011). The consequent part remains as the recursive least square estimation after updating the membership functions, four algorithms were compared namely: LM, Quickprop (QP), Resilient Propagation (RPROP), and gradient descent (GD). For tuning parameters of ANFIS membership functions, Fletcher-Reeves, particle swarm optimization (PSO), and scaled conjugate gradient (SCG) have also been used (Tan, Shafiee, & Wu, et al, 2005; Seyedeh, Tawfik, & Yousef, 2013). The study shows that performance of the algorithms highly depends on the particular problems it is applied to (Chen, et al, 2011).

2.5.1 Neuro-Fuzzy System (NFS) in Medical System

Health care system or medical system is the organization of people, resources and institutions to deliver health care services to meet the health needs of target populations. Diabetes mellitus has been identified to be a life threatening diseases both in urban and rural areas of the Indian nation. Correct diagnosis and earlier diagnose of diabetes will reduce the death rate caused by the disease. Different types of clinical tests have been used to test patients for the presence of diabetes in their body. In medical system, neuro fuzzy applications have received increased attentions from experts, which are the reason for recent research work conducted in this area. The hybridized system has been applied in

different types of disease for example diabetes, breast cancer, alzheimer, thyroid disorder, leukemia, hypotension, brain disorder, cardiac disease, heart disease for diagnosis, treatment, prediction and management etc.

Hybridize expert system have contributed to the growth of medical disease diagnosis and the following briefly outline the contribution of neuro fuzzy system in medical disease diagnosis since 2003. Negoe et al (2003) developed a fuzzy gaussian neural network to recognize the ECG signals for ischemic heart disease diagnosis, with the help of Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA) (Sumathi and Mudigonda, 2015). Rashedur and Rahman (2013) used fuzzy neural filter for the separation of discontinuous lung sounds using dynamic orthogonal least square method.

Phoung and Kreinovich, (2001) investigated brain disorder for classification of electroencephalogram (EEG) signals using an adaptive neuro fuzzy inference system (ANFIS) model. In image processing, Neuro-fuzzy expert system could be one of the important modalities in the biomedical most diagnostic techniques using image (Ambilwade & Manza, 2014). The classification pixel contrivance was independent, powerful and highly useful for prior knowledge and for any expert. Lavanya et al. (2011) separated the lung sounds involve the patients with pulmonary pathology using Takagi–Sugeno–Kang (TSK) recurrent network with fuzzy neural filter.

Polat and Gunes (2007) proposed ANFIS to designed expert system with (PCA) for diabetes patients. The expert system improved diagnostic accuracy of diabetes disease with 89.47% classification accuracy, promising with regard to other classification applications. Aruna, Puviarasan and Palaniappan (2005) combined ANFIS and LDA to

investigated the normal and abnormal heart valves by diagnosing heart sounds. For improvement of diagnostic accuracy Ibrahim et al. used ANFIS with Lyapunov exponents for ECG signal classification. Vijaya and Padmavathamma (2014), combined Evolutionary Algorithm (EA), neuro fuzzy rules and biogeography based optimization (BBO) for cardiac diagnosis. Their focus is on cardiac disease to investigate the expert system classification accuracy.

For better performance neuro-fuzzy model was optimized by particle swarm optimization (PSO) algorithm for ultrasound image segmentation, which was used for detecting the micro calcifications in breasts on orgasm (Ambica, Satyanarayana and Amarendra, 2013). Neuro fuzzy inference was used for diagnosis of Alzheimer disease and in the same year the authors paid attention to leukemia diagnosis (Agboizebeta and Chukwuyeni, 2012). Kumar et al., (2011) developed fast adaptive neuro fuzzy inference system (FANFIS) with analysis of variance (ANOVA) called gene ranking technique to detect the occurrence of cancer. The expert system was later modified by Levenberg–Marquardt algorithm. Agboizebeta and Chukwuyeni (2012) used a set of symptoms with neuro fuzzy expert system to diagnose thyroid disorder. The system was able to interact with the patient and update his current condition. In the same year cognitive neuro fuzzy expert system was used for hypotension control. Ephzibah (2011) combined neuro-fuzzy and genetic algorithm for diagnosis of heart disease. Prasath et al (2013) proposed ANFIS to classify blood samples into normal and abnormal category and also detect abnormality in red blood cell. The system was used to identify, classify, diagnose different type of disease.

2.6 Genetic Algorithm

This algorithm is used to search problems and generate solutions to optimization and it is very useful in solving related problems. It is a search heuristic that mimics the process of natural selection. Genetic algorithms belong to the family of evolutionary algorithms used to generate solutions related to optimization problems like crossover problems, selection based on relative fitness, mutation, and inheritance inspired by using natural evolution technique (Bojarczuk, Lopes & Freitas, 2000). It has been applied in many fields to solve different problems in medical science, bioinformatics, engineering, industry, science, computational mathematics etc. Genetic algorithm could be applied in areas that are too large to be thoroughly searched (Chien et al, 2002).

The technique is used to search for solutions to problems and an iterative procedure involves a population demonstrating as individuals, each one represented by a finite string of symbols, called the genome. An initial population of individuals is generated heuristically or at random. In generation the individuals in the current population are decoded and evaluated according to a fitness function that describes the optimization problem in the search space. The next generation is formed by selecting individuals according to their fitness; a fitness function is used to measure how close the given individual is to reach the set aims of the problem (De Falco, Della, & Tarantino, 2002).

Fitness-proportionate selection is a simple method where individuals are selected with a probability proportional to their relative fitness compared with other selection ways that are available. The performance in the population depends on the expected number of times an individual will be chosen. High-fitness of individuals will give chance to reproduce and bring new individuals to the population, but low-fitness will not bring new

members to the population. Since GA is stochastic iterative processes, it is not guaranteed to converge and the stopping condition may be specified as a maximal number of generations or a chosen level of the fitness.

For feature selection, GA is an important optimization method based on Darwin's theory of natural selection and 'survival of the fittest. It uses the process of competition and controlled variation in representing feasible solutions to particular problem over successive iterations (generations). Competition process is used to select the chromosomes in the population that satisfy the associated fitness to form new ones. Crossover and mutation are commonly used genetic operators to create new chromosomes. To create next generation from current population, three methods are involved namely: Selection, crossover, and mutation.

1. Selection rules: It create individual parent that contribute to the population at the next generation.
2. Crossover rules: It combines two parents to create children for the next generation
3. Mutation rules: It applies randomly to form children from changes to individual parents.

In creation of new population $P(t)$ with chromosomes in $P(t-1)$ are used as selection mechanism. Higher fitness chromosomes usually have a greater chance of contributing copies to $P(t)$ and the number of copies received for each chromosome depends on its fitness. Both crossover and mutation are then applied to $P(t)$. Crossover takes two individual parents to produce two new individuals called the offspring, by swapping parts of the parents. The operator works by exchanging substrings after a randomly selected crossover point. In the new population, the crossover operator is not usually applied to all

pairs of chromosomes. Where the likelihood applied depends on probability defined by a crossover rate. To prevent premature loss of population diversity, sampling of new points in the search space mutation is applied.

Flipping one or more random bits with a probability equal to the mutation rate is used by mutation. Finding an acceptable solution by some criterion or termination may be triggered by reaching a maximum number of generations. Each chromosome consists of 24 genes and each gene represents the connection weight of a diagnosis variable in a length of 1 bit. Feasible solution is used to generate an initial population holding a set of possible solutions from random chromosomes. Each chromosome is denoted as a vector $C = (CA...CX)$ of binary decision variables $C_i=0, 2, 3$; encoded in binary representation as string consisting $\{0, 1\}$ genes. A gene $C_i = 1$ if the i th variable is included in a solution set of a diagnostic process otherwise 0. Fitness function is used to optimize each chromosome by evaluating the genes that constitute the chromosome using their fitness value.

Fitness value of each chromosome keeps improving till it reaches an optimum value when it can no longer improve, as the algorithm continues to complete its cycle, a number of constraints have been considered in carrying out appropriate management of disease in medical diagnosis. Therefore, fitness evaluation must be done with proper constraint for proper estimation. Constraints must be achieved by rendering most of the solutions from the search space, hence; its application in GA is problem specific. The following equation is used to evaluate an individual $F_{(i)}$:

$$F_i = \left(1 + \sum_i^n W_i * C_{i(p)}\right)^{-1} \quad (2.12)$$

n represent the number of diagnosis variables, W_i present weight associated with i th mutable and $C_i(p)$ represent the number of violations for i th constraint at solution p . This fitness function has a range of $[0, 1]$ and an optimal solution occurs when we have 0 violations thus $\sum_{i=0}^n W_i * C_{i(p)}$ which results in $F_i = 1$. For mating in order to produce outstanding candidates and maximize the fitness function, chromosomes with higher fitness value are selected as parents. If the fitness value of an individual is F_i , then the probability, P_i , of choosing the individual is:

$$P_i = \frac{F_i}{\sum_{i=0}^{n-1} F_i} \quad (2.13)$$

This process is repeated until an optimal connection weight is achieved. It is worthing note that the probability of choosing an individual for genetic operation is proportional to its fitness.

2.6.1 Pseudo code for Genetic Algorithm

The following steps are the pseudo code for Genetic algorithm, which is an iterative.

Step 1: The choice of value m is important selection is made randomly from the set G_n by selected an initial population, $P(k)$, of a given size m , where $k=1$. If it is too large, the algorithm does not change from an exhaustive search, but if it is too small, the algorithm may not reach the optimal solution.

Step 2: To determine each chromosome x in the population the value of the fitness function $f(x)$, evaluate each chromosome in population $P(k)$ in terms of its fitness.

Step 3: For the new population $P_n(k)$ from the given population $P(k)$ used procedure of natural selection. The only possible procedure of natural selection is referred to as

deterministic sampling. Calculate the value $e(x) = mg(x)$ for each x in $P(k)$, where $g(x)$ is a relative fitness by:

$$g(k) = \frac{f(x)}{\sum_{n \in p^{(k)}} f(x)} \quad (2.14)$$

From the equation the number of copies of each chromosome x in $P(k)$ is chosen for $P_n(k)$, given by the integer part of $e(x)$. If the total number of chromosomes chosen is smaller than m (the usual case), then it selects the remaining chromosomes for $P_n(k)$ with fractional parts of $e(x)$, from the highest values down. The purpose of this technique is to duplicate those with high fitness and eliminate chromosomes with low fitness.

Step 4: If the criteria use for stop are not met, then go to step 5, otherwise stop.

Step 5: Repeat the operation on chromosomes in population $P_n(k)$ to produce a population of new chromosomes $P(k+1)$. The step attempts to mimic genetic operation observed in biological systems. Repeating the procedure over and over until the proper solution is revealed.

2.6.2 Overview of Genetic Algorithm (GA) in Diabetes

GA has already been used by a lot of authors. Lim, Loh and Shih (2000) used hybrid GA Algorithms and BPN with ANN for classification of diabetes using PIMA Indians database. Polat, et al (2008) combined principal component analysis, neuro-fuzzy inference and Generalized Discriminant Analysis (GDA) with least square support vector machine (LS-SVM) for diabetes data classification. It recorded performance accuracy of 89.47% and 79.16% respectively (Harun, 2010).

Hasan, Nijat, and Feyzullah (2009) classified diabetes patients using a neural approach and achieved 82.37% accuracy. Pradhan et al. (2011) designed a 2-class classifier using Comparative Partner Selection (CPS) and GP for detecting diabetes disease. Cheung, Wee, Li *et al.* (2005) used C4.5, Naive Bayes, BNND and BNNF algorithms for prediction and classification of diabetes patients. The result recorded 81.11%, 81.48%, 81.11% and 80.96%, classification accuracies respectively. Arcanjo et al. (2011) based on the three basic assumptions of semi-supervised learning proposed a KNN-GP (KGP) algorithm with semi-supervised transductive. Inferior results were obtained for diabetes dataset because the system was implemented on 8 datasets of UCI repository (Arcanjo, et al., 2011).

From previous research works, few researchers have used multi-class classifiers, Kishore et al., (2000); Lim, Loh and Shih, (2000) had tried to attempt with it. Kishore et al. (2000) used multi-class classifiers to consider a class problem as a set of two-class problems. Smart used multi class method using modified genetic operators and point out that GP can be used to improve multi class problems. Lim et al, (2000) investigated and presented a tremendous comparison of different 33 classification algorithms. The system used large

number of benchmark datasets for the system evaluation and comparison, but none of these 33 classification algorithms use GP (Lim et al, 2000).

2.6.3 Drawbacks of Genetic Algorithm

Genetic Algorithms according to Maxim have shown several weaknesses and some drawback which make researchers think of using another algorithms or hybrid methods (Ephzibah, 2011):

- (i) Domain knowledge cannot be directly documented using genetic algorithm structure in Genetic Algorithms.
- (ii) Since genetic algorithms require the entire population it take lots of time and memory and in real time applications, GA may not alone work well.
- (iii) It cannot be used in a problem involved in real life application, because it is difficult to use genetic algorithm in problems that deal with imprecision and uncertainty that is the basic requirement for real life applications. However, its inability in storage and handling imprecision is the significant limitation of Genetic Algorithms.

2.7 Diabetes Mellitus

Diabetes has been identified to be a major health problem in both industrial and developing countries, and on a daily basis its incidence is rising. It occurs when the hormone that “unlocks” the cells and the body requires more insulin or does not produce/properly, uses insulin and glucose used to fuel them (Mohammad et al., 2013). If the body is not responding to insulin or produces insulin which is not enough to regulate glucose then the patient is suffering from diabetes. Type I, Type II and Gestational diabetes are the three types of diabetes.

(i) Type I diabetes

Type 1 diabetes is responsible for 5% to 10% of all the cases of diabetes diagnosed. The type is formerly referred to as insulin-dependent diabetes mellitus (IDDM). Its risk factors are not well distinct when compared to type 2 diabetes. Genetic and environmental factors are the most common sources or causes of type 1 diabetes.

(ii) Type II diabetes

Type II diabetes is responsible for 90% cases of diabetes in patients, which make it to be the most common among all types of diabetes and it is an adult-onset diabetes earlier called non-insulin-dependent diabetes mellitus (NIDDM). Old age, family history, prior history of gestational diabetes, impaired glucose tolerance, obesity, physical inactivity and ethnicity/race are the risk factors of type 2 diabetes.

(iii) Gestational Diabetes

Gestational Diabetes responsible for 2% to 5% in all pregnant patients. It occurs during pregnancies and normally disappears after delivery of the patients. Pregnant women have enough insulin, its effect is partly by variation of other hormones produced in the placenta, it is called insulin resistance (Tawfik, Mohammad, Yousef and Tabibi, 2012).

The development of kidney disease can be aided by diabetes. Diabetes also increases the case of heart disease, blood vessel damage, blindness, and nerve damage. The proper diagnosis of diabetes and interpretation of its data will help diabetes classification problem (Temurtas, 2009). In Gestational Diabetes, the body fails to properly use insulin, it is insulin resistance combined with relative insulin deficiency. Type 2 diabetes happen

if the cell ignores the insulin or the body is not producing enough insulin (Mohammad, et. al., 2013).

Table 2.1: Comparison of Type – I and Type II Diabetes (Lalka and Jain, 2015)

Feature	Type I	Type II
Ketoacidosis	Common	Rare
Autoantibodies	Usually Present	Absent
Endogenous Insulin	Low or absent	Normally Increase or decrease
Onset	Sudden	Gradual
Age at Onset	Mostly in children	Mostly in adult
Body Habitus	Thin or Normal	Often Obese
Prevalence	10%	90%

The comparison of type I and Type II diabetes is shown in Table 2.1. Most of the cases of detection of diabetes in patients are type 2 and are on the increase recently. Type 2 takes time before it is discovered and the delay from disease onset to diagnosis may exceed 10 years. Therefore, treatment of hyper glycaemia and related metabolic abnormalities and earlier detection of Type 2 diabetes are of importance.

CHAPTER THREE

RESEARCH METHODOLOGY

This chapter discusses the methodology that guides the research work. A detailed description is given on the necessary step for the development of the designed Genetic Neuro Fuzzy Inferential System for diagnosis of diabetes. The data collection, knowledge acquisition and dataset used are presented. The various parameters (metrics) used for final system evaluation were also enumerated in this chapter.

3.1 Data for the Proposed System

In this work direct rating method was used for acquiring data, which is an effective way of constructing a membership function and direct means of collecting data. Also, to rate the membership function, all the rating indicators (subject) in sequences of data were presented (objects) to the domain experts. To construct a membership function, the responses of several physicians were collected. Both lowest and highest values were used finally to form the ranges for membership function calculation. Minimum values are the lowest and highest as the maximum values. The data used fell within the range, removing any chance of losing data. Table 3.1 shows the ratings of the data for the proposed system for diagnoses of diabetes mellitus.

Table 3.1: Results of rating for the Indicators of the Diabetes Mellitus by Physicians

Input Field	Range				Fuzzy Set
	1st Physician	2 nd Physician	3 rd Physician	4 th Physician	
1. Glucose	$\leq 71 - 110$	≤ 90	≤ 120	≤ 121	Low
	94-120	95-130	97-138	96-145	Medium
	≥ 121	≥ 122	≥ 124	≥ 125	High
2. INS	$\leq 15 - 80$	$\leq 15 - 78$	$\leq 20 - 86$	$\leq 17 - 87$	Low
	15-100	20-150	17-90	15-100	Medium
	$\geq 89 - 194$	$\geq 95 - 193$	≥ 192	≥ 194	High
3. BMI	≤ 24	≤ 27	≤ 33	≤ 30	Low
	24	35	39	40	Medium
	≥ 33	≥ 35	≥ 37	≥ 40	High
4. DPF	≤ 0	≤ 0.2	≤ 0.3	≤ 0.4	Low
	0.2	0.3	0.4	0.6	Medium
	≥ 0.4	≥ 0.6	≥ 0.7	≥ 0.4	High
5. Age	≤ 30	≤ 29	≤ 38	≤ 35	Young
	33	40	42	43	Mid
	$\geq 45 - 55$	$\geq 47 - 54$	≥ 50	≥ 58	Old

3.2 Dataset for the Proposed System

Pima Indians Diabetes Database (PIDDD) was used in this research work and it is freely available on the Internet for download. The last attribute was used as target variable and the first 8 attributes as input data. The database contains 768 samples, each sample with 8 features of clinical findings. Number of times pregnant, Plasma glucose concentration, Diastolic blood pressure (mm Hg), Triceps skin fold thickness (mm), 2-Hour serum insulin (μ U/ml), Body Mass Index (weight in kg/ (height in m) ²), Diabetes pedigree function and Age (years) represented the eight risks factor used in Pima Indians Diabetes Database (PIDDD). Table 3.2 shown the features and description and their attribute values

Table 3.2 Features of PIMA Indians Diabetes Dataset for Diagnosing of Diabetes

No of Attributes	Attributes	Description and Attribute Values
1	Number of Times Pregnant	Numerical values
2	Plasma Glucose Concentration (PGC)	Numerical values
3	Diastolic Blood Pressure (DBP)	Numerical values in (mm Hg)
4	Triceps Skin fold Thickness (TSFT)	Numerical values in mm
5	2-Hour Serum Insulin (2-HSI)	Numerical values in (μ U/ml)
6	Body Mass Index (BMI)	Numerical values in weight in kg/(height in m) ²
7	Diabetes Pedigree Function (DPF)	Numerical values
8	Age	Numerical
9	Diagnosis of types 2 diabetes disease	Yes = 1 No = 0

3.3 Indicators of Diabetes Mellitus

- i. **Number of pregnancy:** It is categorized as absent, low and risky. If a person is male then number of pregnancy will be absent.
- ii. **Diastolic blood pressure:** This field has four fuzzy sets low, medium, high and very high.
- iii. **Triceps skin fold thickness:** Is a value used to measure body fat.
- iv. **Glucose:** Glucose or blood sugar is the main source of energy found in the blood. It is measured by 2-hour glucose concentration after 2 hours of having breakfast (American Diabetes Association, 2014). Glucose has three fuzzy sets low, medium and high.
- v. **Insulin (INS):** Insulin is the hormone excreted by the pancreas to help move glucose from the blood into the cells to be used for energy. It is measured by 2- hour serum insulin (INS) after 2 hours of having breakfast (American Diabetes Association, 2014). If the cells do not respond well to insulin then glucose is unable to enter the cells. Consequently, the cells fail to get the fuel they need, and glucose builds up in the blood stream. INS has three fuzzy sets low, medium and high.
- vi. **Body Mass Index (BMI):** BMI is considered as an assessment of evaluating the weight of the body in relation to the height of a person. This field consists of three fuzzy sets in the developed system are low, medium and high.
- vii. **Diabetes Pedigree Function (DPF):** DPF is the statistical classification of a certain group of data (American Diabetes Association 2014). For example, of the age group 40-45, data are analyzed and calculated when determining statistical values for the age group. DPF consists of three fuzzy sets low, medium and high.
- viii. **Age:** Age is considered to be another indicator of diabetes. Age has three fuzzy sets young, medium and old.

3.4 Neural Network (NN)

Neural Network can easily handle both continuous and discrete data and it has the capability of capturing domain knowledge from available indicators. Neural network was used to train and test the designed fuzzy system, also to improve the performance of the overall system. It is made up of 8 attributes namely: Diastolic blood pressure, 2 hours in an oral glucose tolerance test (Plasma glucose concentration), Number of times pregnant, Triceps skin fold thickness, 2-Hour serum insulin (INS), Body mass index (BMI), Diabetes pedigree function (DPF) and Age. W_i was used to show that each attribute has a weight that contributes to the diagnosis process. The patient diagnosis attributes was fed into the neural network as input layer and contribution of each category of variables was determined at the hidden layer. The following is the equation used:

$$CAT_i = \sum_i^n A_i * W_{Ai} \quad (3.1)$$

The sets of variable i th is CAT_i , the number of attributes in CAT_i is represented by n , and A_i is the i th diagnosis variable with weight W_{Ai} . Result of the NN (output layer) represents the total output of diagnosis by Neural Network component of the architecture. The output result is illustrated in Equation (3.2)

$$Output_{Neural\ Network} = \sum_i^n CAT_i * W_{CAT_i} \quad (3.2)$$

Where W_{CAT_i} is the connection weight of CAT_i

3.5 Fuzzy Logic

Fuzzy logic has the strength of handling imprecise and incomplete data that is one of the characteristics of medical records, and it is a superset of the conventional Boolean. It finds a precise solution to a given problem and its ability to work from approximate reasoning also resembles human decision making. The following process is the stages involves in diagnosis and management of diabetes by fuzzy logic:

- (i) Fuzzification of the attributes input by the patient
- (ii) Formation of fuzzy rule base system.
- (iii) Inference engine: the building of decision making for the fuzzy logic component.
- (iv) Defuzzification of the output results from inference engine into crisp values.

3.5.1 Algorithm for Fuzzy logic

Step 1: Glucose, INS, BMI, DPF and age was used as input crisp values.

Step 2: Set the triangular membership function for the fuzzy number.

Step 3: Build the fuzzy numbers for the five (5) attributes for input set & output set

Step 4: For fuzzy inference method Mamdani's was used to execute.

Step 5: Input the rules and calculate the matching degree of rule with "OR" fuzzy disjunction for fuzzy input set (Glucoselow, Glucosemedium, Glucosehigh, INSlow, INSmedium, INShigh, BMIlow, BMImedium, BMIhigh, DPFlow, DPFmedium, DPFhigh, Ageyoung, Agemedium, Ageold).

Step 6: Compute the aggregation of the fired rules for fuzzy output set DM (DMverylow, DMlow, DMmedium, DMhigh, DMveryhigh).

Step 7: Defuzzify into the crisp values by:

$$z^* = \frac{\int \mu A(z).zdz}{\int \mu A(z)dz} \quad (3.3)$$

Where, $\int$ is the algebraic integration, $\mu A(z)$ is the number of fuzzy numbers of the output fuzzy variable DM and z represents the weight for $\mu A(z)$.

Step 8: Represent the knowledge in human language form.

End.

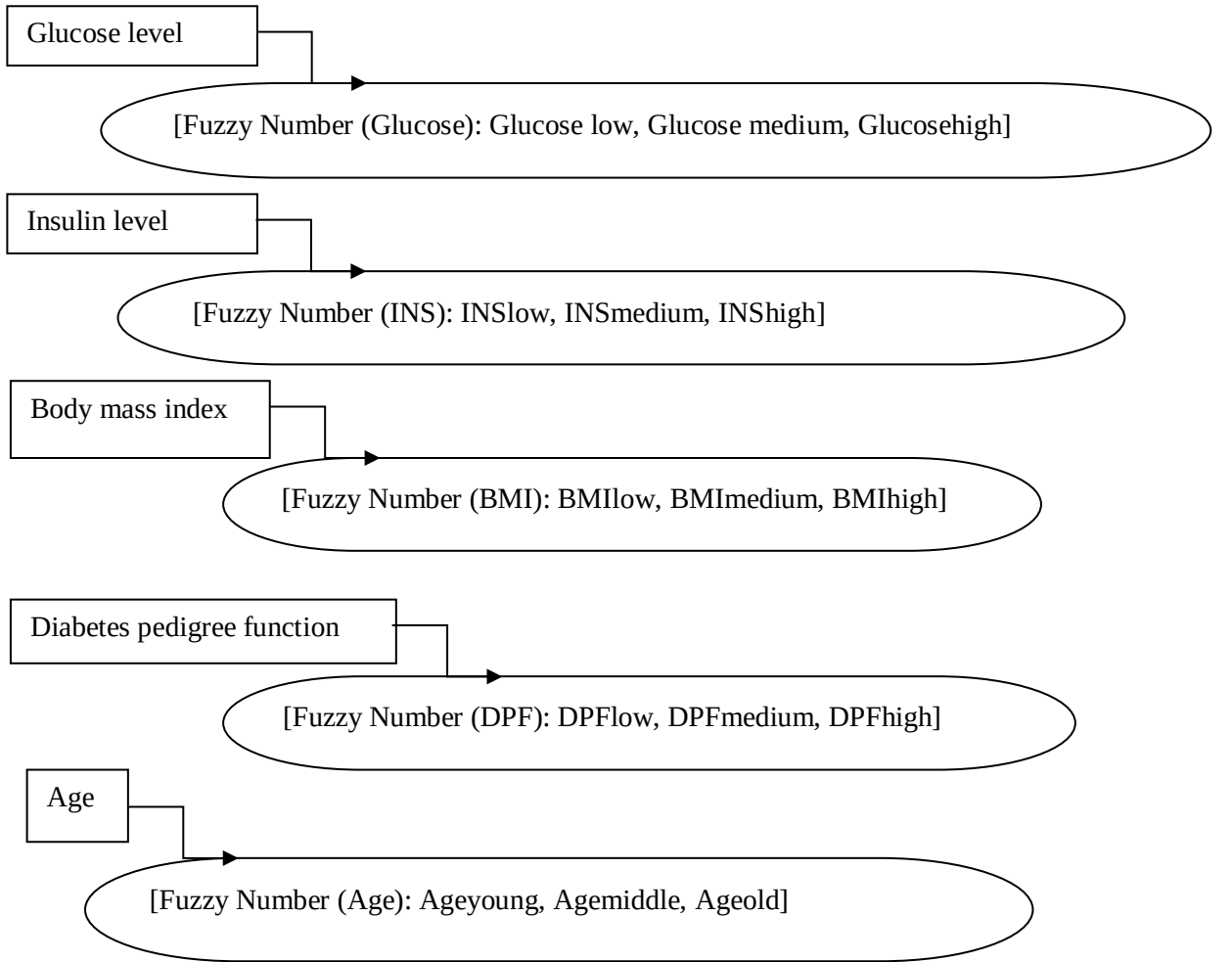


Figure 3.1: Sentence Analysis represents the physical data of the person.

(a) Fuzzification

In fuzzy inference system, fuzzification is the first step. It's a domain transformation by which crisp inputs are transformed into fuzzy inputs. The fuzzy sets for the indicators, and the output of diagnosis and management of diabetes along with the membership function were defined in fuzzification.

The Fuzzy Sets for the Indicators and for the Output of Diabetes Mellitus is as follow;

Number of Pregnancy: {Absent, Normal, Risk}.

Diastolic Blood Pressure: {Low, Medium, High, Very High}.

Triceps Skin Thickness: {Good, Average, Below Average}.

Glucose: {Low, Medium, High}.

Insulin :{Low, Medium, High}.

Body Mass Index (BMI): {Low, Medium, High}.

Diabetes Pedigree Function (DPF): {Low, Medium, High}.

Age: {Young, Medium, Old}.

Output: {Low, Medium, High}.

For the output fuzzy set, the system used 0 = Low, 1 = Medium, and 2= High. The following Table 3.3 shows the ranges of the output fuzzy sets.

Table 3.3: Ranges of the Output Fuzzy set for Diabetes Mellitus Application

Output Fields	Range	Fuzzy Set
Result	< 0.4	Low
	0.4 – 0.6	Medium
	0.61 – 1	High

For the final result, this study considered low as No Diabetes, medium and high as Diabetes. After the indicators and their fuzzy sets have been defined, the range values were prepared for the fuzzy sets of each indicator and the rating for the data used for the model was carried out by the physicians. When the range values for the fuzzy sets were ready, the equations were constructed by using the range values to generate the membership function. Triangular, trapezoidal and bell-shaped membership functions are some of the types of membership functions that have been proposed. In this work, triangular membership function was used because calculations with triangular membership are easy, and shapes are simple and more flexible and have fewer complexes when splitting values (low, med and high MF) comparing other membership functions. Triangular member function is show in equation 3.4.

$$\mu(x) = \begin{cases} 0, x < a \\ \frac{x-a}{b-a}, a < x \leq b \\ \frac{c-x}{c-b}, b < x < c \\ 0, x > a \end{cases} \quad \text{Triangular membership functions} \quad (3.4)$$

After generating the membership functions of the fuzzy sets, in order to get the most appropriate membership from the fuzzy set of each indicator, the maximum was taken from the generated membership function of the fuzzy set in each indicator. The maximum was considered because this study followed the Mamdani method to develop the FIS. In Mamdani, maximum is taken from the generated membership function of the fuzzy sets to choose the appropriate membership function.

(c) The Formation of the Fuzzy Rule Base

After obtaining the linguistic terms of the indicators, the fuzzy rules were generated heuristically by domain experts in this study. For this work, 25 fuzzy rules were developed for diagnosis and management of diabetes mellitus. These rules are as follows:

- i. If Glucose is Low then Diabetes Mellitus is Low
- ii. If Glucose is Medium then Diabetes Mellitus is Medium
- iii. If Glucose is High then Diabetes Mellitus is High
- iv. If Insulin is Low then Diabetes Mellitus is Low
- v. If Insulin is Medium then Diabetes Mellitus is Medium
- vi. If Insulin is High then Diabetes Mellitus is High
- vii. If Body Mass Index is Low then Diabetes Mellitus is Low
- viii. If Body Mass Index is Medium then Diabetes is Medium Mellitus
- ix. If Body Mass Index is High then Diabetes Mellitus is High
- x. If Diabetes Pedigree Function is Low then Diabetes is Low
- xi. If Diabetes Pedigree Function is Medium then Diabetes Mellitus is Medium
- xii. If Diabetes Pedigree Function is High then Diabetes Mellitus is High
- xiii. If Age is Young then Diabetes Mellitus is Low
- xiv. If Age is Medium then Diabetes Mellitus is Medium
- xv. If Age is Old then Diabetes Mellitus is High
- xvi. If Number of pregnancy is absent then Diabetes Mellitus is Low
- xvii. If Number of pregnancy is Normal then Diabetes Mellitus is Medium
- xviii. If Number of pregnancy is High then Diabetes Mellitus is High

- xix. If Triceps skin fold thickness is Good then Diabetes Mellitus is Low
- xx. If Triceps skin fold thickness is Average then Diabetes Mellitus is Medium
- xxi. If Triceps skin fold thickness is Below average then Diabetes Mellitus is High
- xxii. If Diastolic blood pressure is low then Diabetes Mellitus is Low
- xxiii. If Diastolic blood pressure is Medium then Diabetes Mellitus is Medium
- xxiv. If Diastolic blood pressure is High then Diabetes Mellitus is High
- xxv. If Diastolic blood pressure is Very high then Diabetes Mellitus is High

(d) Building the Inference Engine which is the Decision making of Fuzzy Logic Component

The process of drawing conclusion from existing data is called inference. Inference engine is the knowledge processor that is modeled after the expert's reasoning. Fuzzy inference is formed by membership functions, logical operations, and if-then rules processes.

(e) Defuzzification

Defuzzification is the process of producing a quantifiable result in crisp logic given fuzzy sets and corresponding membership degrees.

The last part of inference engine is to determine the firing strength of each rule. The firing rules shows 23 and 25 fired at 60% or 0.6 and 40 % or 0.4 respectively. Before defuzzification process for crisp output generation, the logical products of the rules must be inferred or combined (max-min'd, max-dot'd, averaged, root-sum-squared, etc.). The Root Sum Square (RSS) is used and has the formula below:

$$RSS = \sqrt{\sum R^2} = \sqrt{(R_1^2 + R_2^2 + R_3^2 + \dots + R_n^2)} \quad \text{Root Sum Square (RSS) formula} \quad (3.5)$$

Where $R_1^2, R_2^2, R_3^2, \dots, R_n^2$ are strength values of different rules which share the same conclusion.

RSS combines the effects of all applicable rules, scales the functions at their respective magnitudes, and computes the "fuzzy" centroid of the composite area. RSS is mathematically compared to other methods. but it gives best weighted influence to all firing rules, which form the reasons for its selection for this work (Saritas &Sert, 2003).

3.6 Genetic Algorithm

Genetic algorithm was used to select optimal attributes (values) from the diagnostic parameters which serves as input, and also for the optimization of the system. PIMA Indian diabetes dataset has 8 attributes. In Neural Network, genetic algorithm was used to choose which attributes to be used as input in order to minimize computational complexity.

3.7 Genetic-Neuro-Fuzzy Inference System (GNFIS)

The proposed system integrate ANN, Fuzzy logic and Genetic Algorithm to build an inferential system called Genetic-Neuro-Fuzzy Inference System (GNFIS) to handle uncertain and imprecise data for diagnosis of diabetes, a self-learning and adaptive system was designed. To build the inference system, the feed forward propagation learning technique made up of nine layers of neurons was used. Both hidden and output layers consist of active nodes, nodes where computations take place at the input layer are passive. Mamdani's Inference Mechanism which is a production rules was used by the inference engine with reasoning algorithm. The active nodes denote inputs for the system and are one of the seven layers. To represent how severe a patient feels, numeric values

are used as the diagnosis variables. The output layer is the linguistic labels corresponding to each input value.

In receiving the output of preceding layer as input, second layer which is made up of adaptive nodes was used and produced their corresponding membership grade using the formula below:

$$L_2(x_i) = \mu_{Ai}(X_i) \quad (3.6)$$

The fuzzy value of each variable is computed using triangular MF, given as:

$$\mu_{Ai}(X_i) = \frac{x_i - b}{a - b} \quad (3.7)$$

3.8 System Development and Design

The system was developed for diagnosis of Diabetes Mellitus. Pima Indian Diabetes Database (PIDDD) was used to build the system. The database has 8 attributes as input and the last attributes as target variable. The following are the attributes of PIDDD database namely: Triceps skin fold thickness (mm), number of times pregnant, 2-Hour serum insulin (mu U/ml), Diastolic blood pressure (mm Hg), Diabetes pedigree function, Body mass index (weight in kg/ (height in m)²), Plasma glucose concentration a 2 hours in an oral glucose tolerance test, and Age (years). The attributes was used to test if a patient is tested-positive or tested-negative. The totals of 768 patients are available in PIDDD with the sample shown below in Figure 3.2.

```
6,148,72,35,0,33.6,0.627,50,1
1,85,66,29,0,26.6,0.351,31,0
8,183,64,0,0,23.3,0.672,32,1
1,89,66,23,94,28.1,0.167,21,0
0,137,40,35,168,43.1,2.288,33,1
5,116,74,0,0,25.6,0.201,30,0
3,78,50,32,88,31.0,0.248,26,1
10,115,0,0,0,35.3,0.134,29,0
2,197,70,45,543,30.5,0.158,53,1
8,125,96,0,0,0.0,0.232,54,1
```

4,110,92,0,0,37.6,0.191,30,0
 10,168,74,0,0,38.0,0.537,34,1
 10,139,80,0,0,27.1,1.441,57,0
 1,189,60,23,846,30.1,0.398,59,1
 5,166,72,19,175,25.8,0.587,51,1
 7,100,0,0,0,30.0,0.484,32,1
 0,118,84,47,230,45.8,0.551,31,1
 7,107,74,0,0,29.6,0.254,31,1
 1,103,30,38,83,43.3,0.183,33,0
 1,115,70,30,96,34.6,0.529,32,1
 3,126,88,41,235,39.3,0.704,27,0
 8,99,84,0,0,35.4,0.388,50,0
 7,196,90,0,0,39.8,0.451,41,1
 9,119,80,35,0,29.0,0.263,29,1
 11,143,94,33,146,36.6,0.254,51,1
 10,125,70,26,115,31.1,0.205,41,1
 7,147,76,0,0,39.4,0.257,43,1
 1,97,66,15,140,23.2,0.487,22,0
 13,145,82,19,110,22.2,0.245,57,0
 5,117,92,0,0,34.1,0.337,38,0
 5,109,75,26,0,36.0,0.546,60,0
 3,158,76,36,245,31.6,0.851,28,1
 3,88,58,11,54,24.8,0.267,22,0
 6,92,92,0,0,19.9,0.188,28,0
 10,122,78,31,0,27.6,0.512,45,0
 4,103,60,33,192,24.0,0.966,33,0
 11,138,76,0,0,33.2,0.420,35,0
 9,102,76,37,0,32.9,0.665,46,1
 2,90,68,42,0,38.2,0.503,27,1
 4,111,72,47,207,37.1,1.390,56,1
 3,180,64,25,70,34.0,0.271,26,0
 7,133,84,0,0,40.2,0.696,37,0
 7,106,92,18,0,22.7,0.235,48,0
 9,171,110,24,240,45.4,0.721,54,1
 7,159,64,0,0,27.4,0.294,40,0
 0,180,66,39,0,42.0,1.893,25,1
 1,146,56,0,0,29.7,0.564,29,0
 2,71,70,27,0,28.0,0.586,22,0
 7,103,66,32,0,39.1,0.344,31,1
 7,105,0,0,0,0.0,0.305,24,0
 1,103,80,11,82,19.4,0.491,22,0
 1,101,50,15,36,24.2,0.526,26,0
 5,88,66,21,23,24.4,0.342,30,0
 8,176,90,34,300,33.7,0.467,58,1

Figure 3.2: Sample of Pima Indians Diabetes Dataset (PIDD)

From Pima database diabetes dataset 100 patients' were formulated and stored as rules in the database (medical records of the patients) to evaluate the performance of the designed system. In order to have an easy estimation 100 patients were used from the dataset to test the proposed system. Experts in human respiratory diseases determine the output value of

each rule from the records retrieved for the system. Expertise of human expert are used for the assessment of the intensity of the input variables. Intensity of each variable represents its contribution to diabetes disease.

Java programming language was used to develop the expert system and the experimental setup was carried out with the system built. The programming language was chosen because of being a high level programming language and object oriented based program with functions and class which could be used to develop and link a graphic user interface for user interactive and responsiveness. The diagnostic system employed several components in Java Script to achieve the implementation of this system.

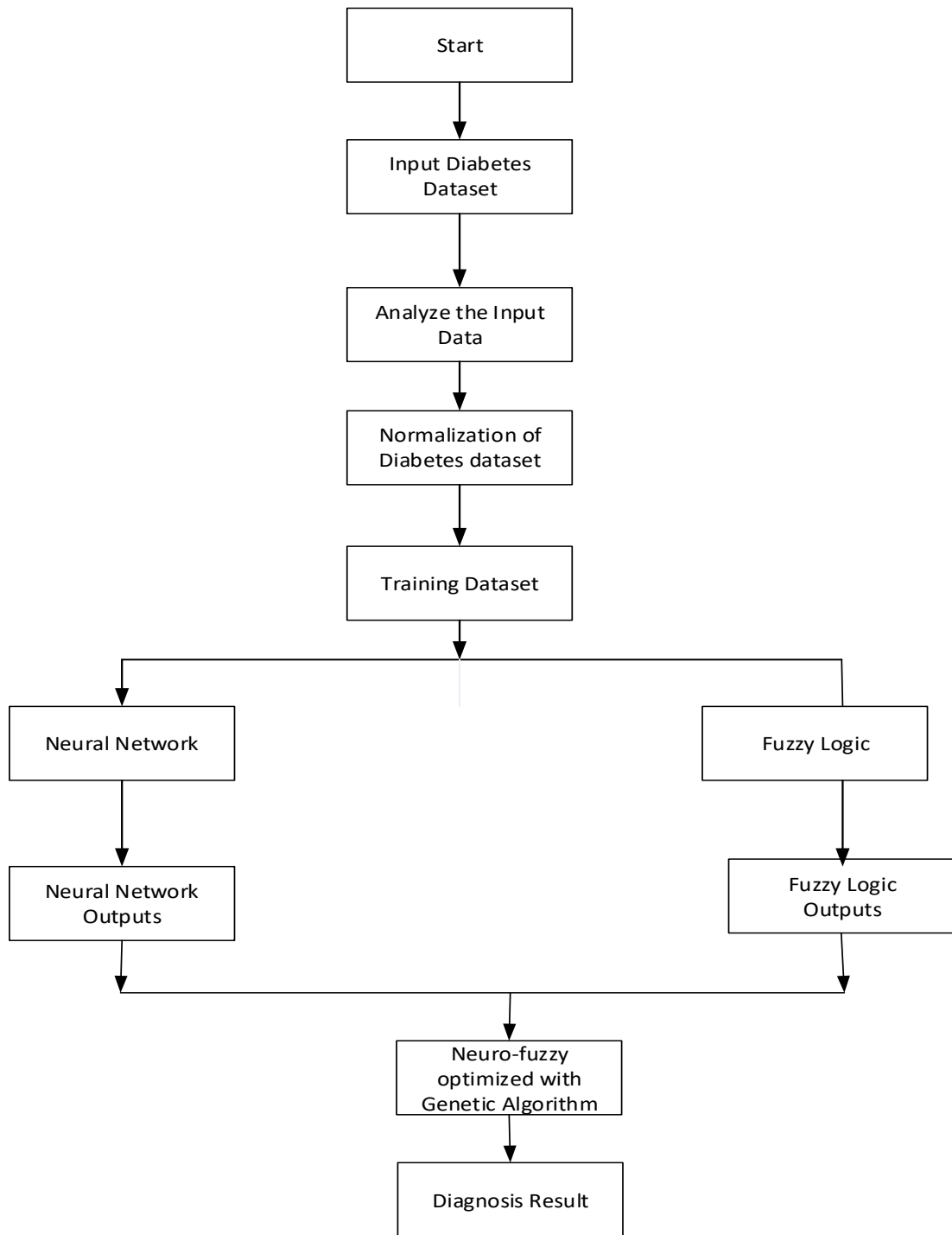


Figure 3.3: Block Diagram for the Proposed System

3.9 Performance Evaluation Metrics

The various stages identified in this research are implemented using appropriate tools and evaluation of the proposed approach are carried out using measure of prediction such as sensitivity, specificity and predictive values. Sensitivity and specificity have been proved useful in research measures to assess an instrument. Predictive values, positive predictive value (PPV) and negative predictive value (NPV) are more useful in clinical medicine. Basically, medical practitioners have evaluated the designed system. The basic evaluation criteria and assessment based on the classification accuracy level using True Negative (TN), True Positive (TP), False Negative (FN) and False Positive (FP). Sensitivity test is the ability to correctly detect diabetes disease and Specificity is the ability to avoid calling normal things as diabetes disease in medical tests. The classification accuracy is represented in equation (3.11)

$$\text{Sensitivity} = TP / (TP + FN) \quad (3.8)$$

$$\text{Specificity} = TN / (TN + FP) \quad (3.9)$$

$$PPV = TP / (TP + FP) \quad (3.10)$$

$$NPV = TN / (TN + FN) \quad (3.11)$$

Hence

$$\text{Classification Accuracy} = \frac{TN + TP}{TN + FP + FN + TP} \times 100 \quad (3.12)$$

Where: TN = state the number of people who are healthy among those who were diagnosed diabetic

TP = Implies the number of people who actually suffer from diabetes among those who were diagnosed as being diabetic

FN = the number of people found to be healthy among those who were diagnosed as being diabetic

FP = Depicts the number of persons who are unhealthy i.e. is diabetic but were diagnosed as healthy.

3.9.1 Receiver Operating Characteristics (ROC)

ROC curve is used as a graphical plot to illustrate the performance of a binary classifier system as its discrimination threshold is varied. The use of ROC curve to evaluate the performance of diagnostic tests is an effective method. It is related to natural way and direct to cost/benefit analysis of diagnostic decision making.

ROC curve was first developed by electrical engineers and radar engineers during war II for detecting enemy objects in battle fields, it increasingly used in machine learning and data mining research and it is used for analysis in different fields such as biometrics, medicine, radiology etc.

It was used in this research through the lens of diagnostic testing, even though the curves have been used for different applications across a number of fields of study. Individual or patient is measured given a score with the intent that the measurement will be useful in classifying the unit into one of two condition (e.g., Positive / Negative, Yes / No, Diseased / Non-diseased). If the score and condition is known based on a large number of individuals, researchers may use graph analysis to determine the ability of the score to

classify or predict the condition. ROC Curve analysis may also be used to determine the optimal decision threshold (optimal cutoff value).

A positive or negative diagnosis is made for each unit by comparing the measurement of the cutoff value for a given cutoff value. The predicted condition is negative if the measurement is less or greater than the cutoff, and the predicted condition is positive otherwise. However, the predicted condition does not necessarily match the true condition of the patient. From the inference engine results, there are four possible outcomes: true positive (tp), true negative (tn), false positive (fp), and false negative (fn).

Sensitivity and specificity and many thresholds can be measured by ROC curve simultaneously. ROC curve is very useful to decide threshold that is optimal for a given decision-making problem and by looking at the full spectrums of the thresholds possible it can directly compare different classification systems. It is very useful in an area where uncertain decisions are made especially in health care decision and ROC curve researchers can always compare two systems easily.

3.9.2 10-Fold Cross Validations

Cross Validation is a statistical way of dividing data into two segments for comparing and evaluating learning machine. Part of the partitions for training and learning a model and the other parts are used to validate the model. The training and validation sets must cross-over in successive rounds such that each data point has a chance of being validated against one another. K-fold validation is the basic form of cross-validation, in other special cases involve repeated rounds of k-fold cross validation.

It is used to predict data that is not trained on by determining how accurately a learning algorithm will be able to predict data. The 10-Fold Cross Validations are used in machine learning. After the dataset is randomly partitioned into 10 groups for training, the first 9 groups are used for the classifier and the remaining one is used as the testing dataset. This process is repeated until all the groups have been used as testing dataset, and then the classifier's performance is measured as an aggregate of all the 10 folds. The method was used to test the accuracy of the system when some of the symptoms are removed from the dataset to see if the system will still diagnose correctly. Also to validate that some of the symptoms were not as important as others in determined diabetes in patient.

The cross validations was used to validate the importance of the symptoms and also to know the accuracy of the designed system. The validations help to see the most important symptoms that directly affect patient suffering from diabetes.

CHAPTER FOUR

RESULTS AND DISCUSSION

This chapter presents the Genetic Neuro Fuzzy Inference Diagnostic System Experimental results. The results gathered from the developed GNFIS diagnostics system for Diabetes Mellitus were generated and presented in this chapter. It also states the result in order to show clearly how the objectives enumerated in the research are achieved.

4.1 Genetic Neuro-fuzzy Inference Diagnostic System Experimental Results

The results gathered from developed GNFIS diagnostics system for Diabetes Mellitus are shown in the following sections

4.1.1 The Experimental Results of the GNFIS

Fifty (50) samples were selected from the database for test runs. The one trail/run that is closer to the average classification accuracy value was selected to determine the various performance metrics. The performance classification average accuracy was 97.79%. The performance evaluation metrics from GNFIS system for classifying the reduced dataset is shown in Table 4.1 and Table 4.2; this shows that the classification accuracy of the designed system is better than the existing work.

Genetic algorithm (GA) was used to remove outliers, noisy and inconsistent samples and cases, missing values were replacing by mean. The algorithm was used as a feature selection tool and its output was fed to NFIS using 10-fold cross validation technique for classification. GA was used to select different features from the original set of attributes during each run and their classification accuracy was

recorded for each run. The experiment was repeated 50 times to achieve consistent result.

Table 4.1: Evaluation Metrics obtained from GNFIS diagnosis

Performance Measures	Reduced Dataset
No. of Attributes Used	5
Sensitivity (%)	96.00
Specificity (%)	99.00
Positively Predicted value (%)	99.00
Negatively Predicted value (%)	98.00
GNFIS Accuracy (%)	98.00

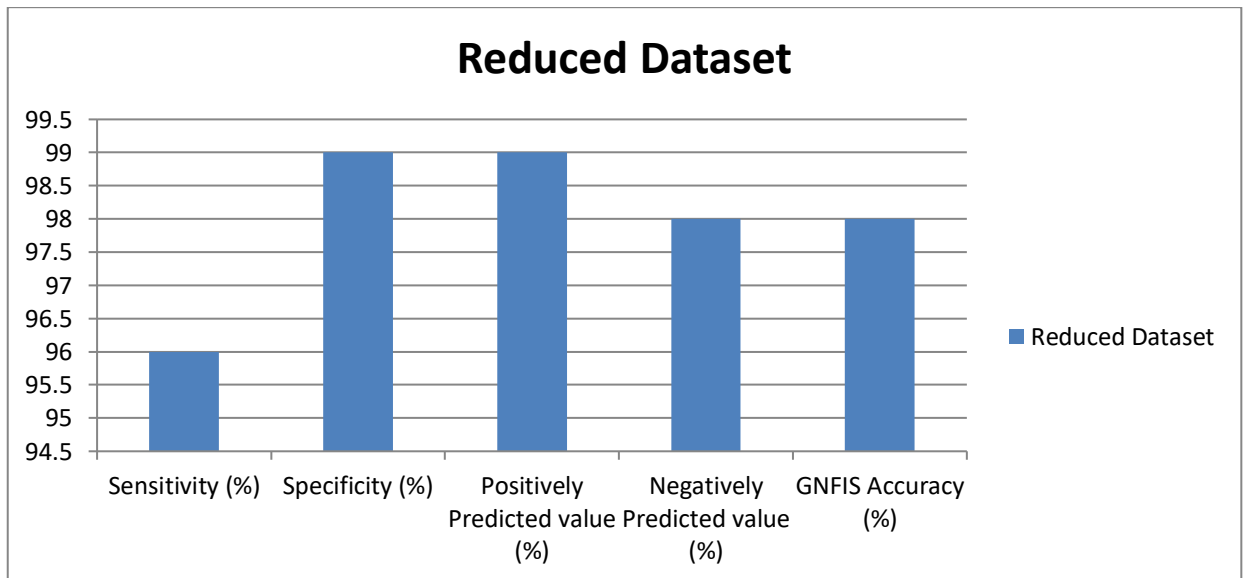


Figure 4.1: Graphical Representation of Evaluation Metrics for GNFIS

Table 4.2: Accuracy obtained from the three systems for diagnosis

Performance Measures	Reduced Dataset
No. of Attributes Used	5
Fuzzy Logic Accuracy (%)	97.00
ANN Accuracy (%)	96.00
GNFIS Accuracy (%)	99.00

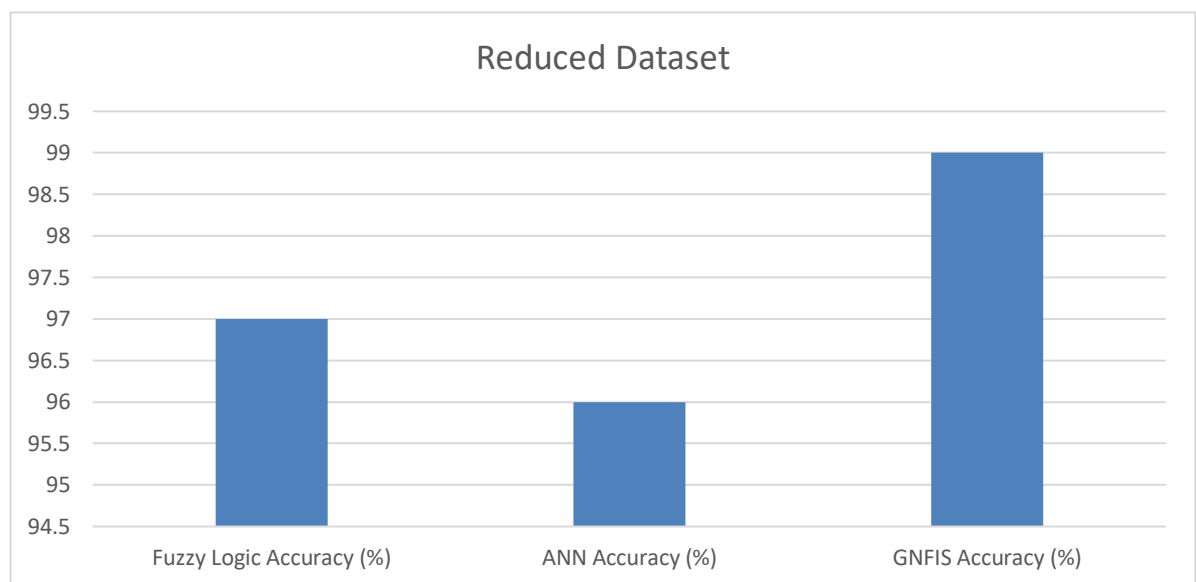


Figure 4.2: The accuracy (%) for the three System for diabetes diagnosis

4.2 Performance Evaluation

This evaluation was carried out using 10-fold cross validation in order to determine how accurately a learning algorithm will be able to predict data in the field of machine learning.

Table 4.3: Classification Accuracy and time taken to diagnose when First feature is removed (Number of time Pregnant)

Classification Techniques	Accuracy (%)	Time Taken to build model (secs)
FUZZY LOGIC	98.7342	0.97
ANN	97.7352	3.13
GNFIS	99.8352	1.25

After the first feature is removed, it was observed that none of the algorithms had less than 96% accuracy; this shows us that the “number of time pregnant” attribute is not one of most important feature in diagnosing whether a patient has diabetes in the body or not.

Table 4.4: Classification Accuracy and time taken to diagnose when second feature is removed (Diastolic Blood pressure)

Classification Techniques	Accuracy (%)	Time Taken to build system(secs)
FUZZY LOGIC	97.9832	0.89
ANN	97.8732	3.57
GNFIS	99.1234	1.34

After the second feature is removed, it was observed that none of the algorithms had less than 96% accuracy; this shows us that the “diastolic blood pressure” attribute is not one of the most important feature in diagnosing whether a patient has diabetes in the body or not.

Table 4.5: Classification Accuracy and time taken to diagnose when third feature is removed (Triceps skin fold thickness)

Classification Techniques	Accuracy (%)	Time Taken to build system(secs)
FUZZY LOGIC	98.9233	0.97
ANN	98.6654	3.69
GNFIS	99.7896	1.75

After the third feature is removed, it was observed that none of the algorithms had less than 96% accuracy; this shows us that the “Triceps skin fold thickness” attribute is one of the most important features in diagnosing whether a patient has diabetes in the body or not.

Table 4.6: Classification Accuracy and time taken to diagnose when fourth feature is removed (Number of Time Pregnant and Triceps Skin Fold Thickness)

Classification Techniques	Accuracy (%)	Time Taken to build system (secs)
FUZZY LOGIC	98.4532	0.89
ANN	97.8756	3.98
GNFIS	98.9876	1.45

After the forth feature is removed, it was observed that none of the algorithms had less than 96% accuracy; this shows us that the “Number of Time Pregnant and Triceps Skin Fold

Thickness” attributes are not one of the most important feature in diagnosing whether a patient has diabetes in the body or not.

Table 4.7: Classification Accuracy and time taken to diagnose when fifth feature is removed (Number time pregnant and Diastolic Blood Pressure)

Classification Techniques	Accuracy (%)	Time Taken to build model (secs)
FUZZY LOGIC	97.8723	0.79
ANN	97.2143	3.88
GNFIS	98.5391	1.44

After the fifth feature is removed, it was observed that none of the algorithms had less than 96% accuracy; this shows us that the “(Number Time Pregnant and Diastolic Blood Pressure)” attributes are not one of most important feature in diagnosing whether a patient has diabetes in the body or not.

Table 4.8: Classification Accuracy and time taken to diagnose when sixth feature is removed (Body Mass Index)

Classification Techniques	Accuracy (%)	Time Taken to build model (secs)
FUZZY LOGIC	94.4532	0.84
ANN	95.6614	3.47
GNFIS	95.7654	1.30

After the sixth feature is removed, it was observed that the algorithms had less than 96% accuracy; this shows us that the “body mass index” attribute is one of most important feature in diagnosing whether a patient has diabetes in the body or not.

Table 4.9: Classification Accuracy and time taken to diagnose when seventh feature is removed (Diabetes Pedigree Function)

Classification Techniques	Accuracy (%)	Time Taken to build model (secs)
FUZZY LOGIC	95.4563	0.98
ANN	95.9453	3.93
GNFIS	94.2756	1.34

After the seventh feature is removed, it was observed all the algorithms had less than 96% accuracy; this shows us that the “diabetes pedigree function” attribute is one of most important feature in diagnosing whether a patient has diabetes in the body or not.

Table 4.10: Classification Accuracy and time taken to diagnose when eighth feature is removed (Age)

Classification Techniques	Accuracy (%)	Time Taken to build model (secs)
FUZZY LOGIC	94.9493	0.87
ANN	93.9493	3.31
GNFIS	95.7356	1.68

After the eighth feature is removed, it was observed that all the algorithms had less than 96% accuracy; this shows us that the “age” attribute is one of most important feature in diagnosing whether a patient has diabetes in the body or not.

Table 4.11: Classification Accuracy and time taken to diagnose when seventh feature is removed (Plasma Glucose Concentration)

Classification Techniques	Accuracy (%)	Time Taken to build model (secs)
FUZZY LOGIC	94.4312	0.85
ANN	94.7985	3.74
GNFIS	94.2156	1.62

After the ninth feature is removed, it was observed that all the algorithms had less than 96% accuracy; this shows us that the “Plasma Glucose Concentration” attribute is one of most important feature in diagnosing whether a patient has diabetes in the body or not.

Table 4.12: Classification Accuracy and time taken to diagnose when eighth feature is removed (2-Hour Serum Insulin)

Classification Techniques	Accuracy (%)	Time Taken to build system (secs)
FUZZY LOGIC	94.2341	0.79
ANN	95.3431	3.35
GNFIS	95.1256	1.59

After the tenth feature is removed, it was observed that all the algorithms had less than 96% accuracy; this shows us that the “2-Hour Serum Insulin” attribute is one of most important feature in diagnosing whether a patient has diabetes in the body or not.

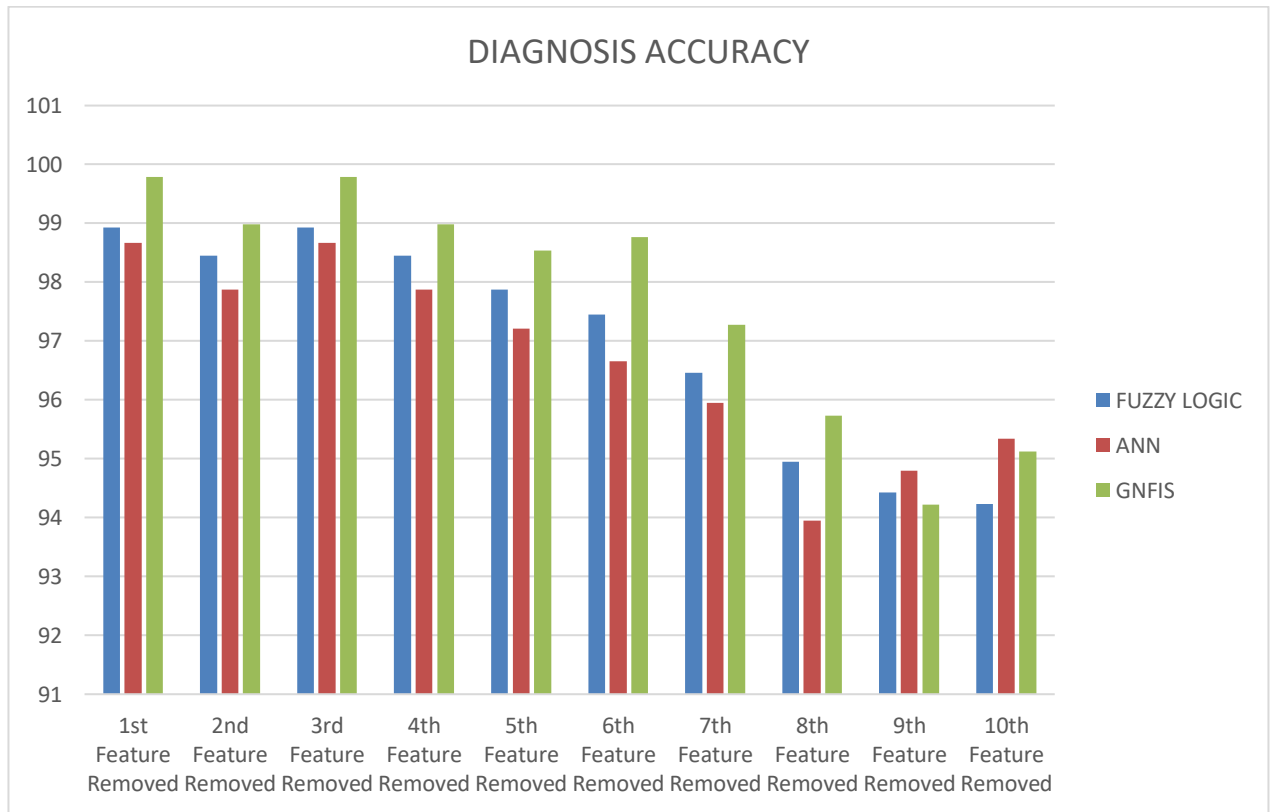


Figure 4 3: Classification Accuracy

After five (5) features were removed, it was observed that all the classifiers had less than 96% classification accuracy, this shows that the “Age, diabetes pedigree function, plasma glucose concentration, BMI and 2-Hour Serum Insulin” attributes are the most important feature in determining whether diabetes is present in a patient or not. But being a medical condition where increased accuracy in classification can mean the difference between life and death, other features need to be considered. GNFIS performed best when the first five features were removed but subsequently, removal of features reduced its accuracy.

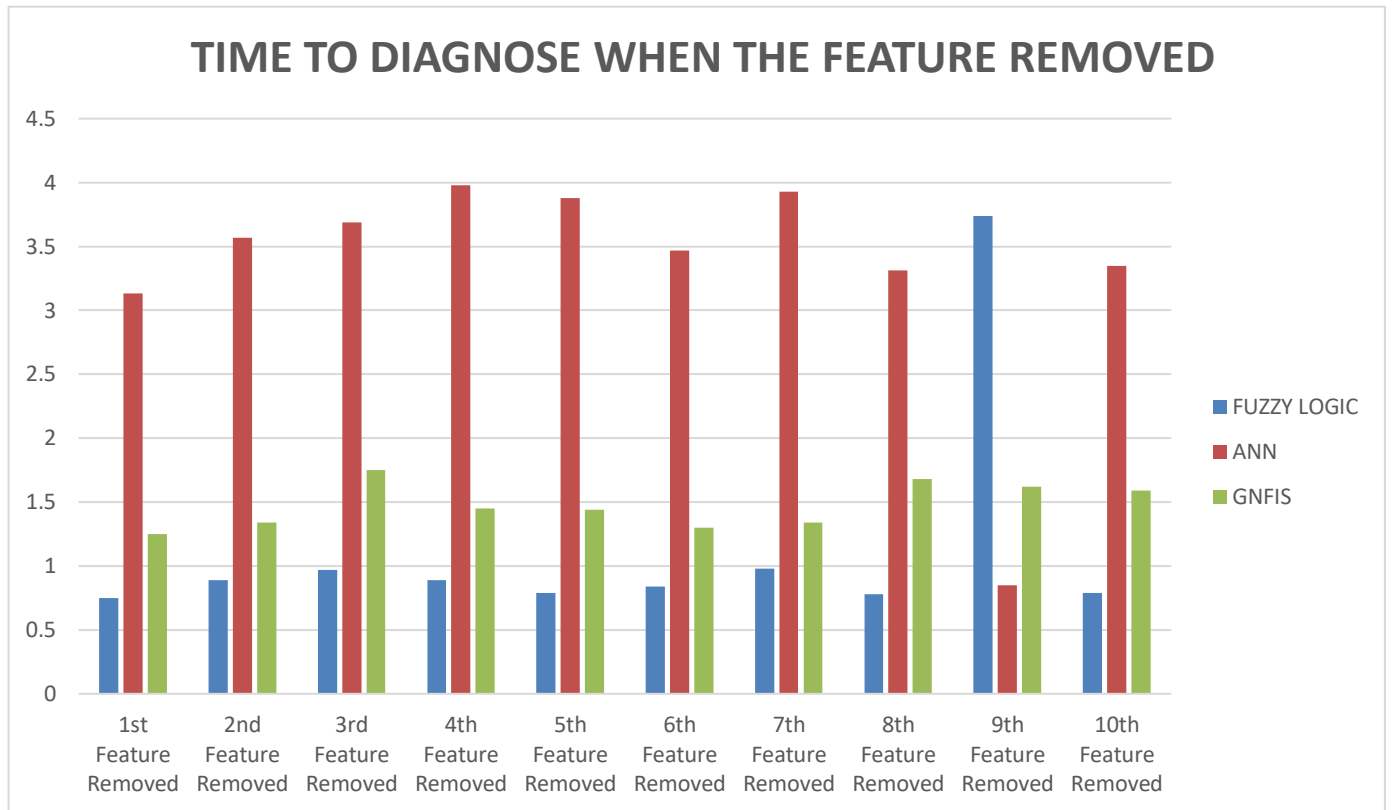


Figure 4.4: Time taken to build classification model

Fuzzy logic was extremely fast in building its classification model, while ANN took a longer time in building its classification model. In a condition that time is very crucial fuzzy logic will be the best to use, especially, when the data involved is very large, fuzzy logic can be used to determine the class of the diabetes in a patient. As expected, removal of more features generally reduces the time it takes for algorithms to build their models. GNFIS can be used for its highest accuracy which gives it an edge compare to other two algorithms.

Table 4.13: Results for sensitivity and Specificity of GNFIS, Fuzzy Logic and Neural Network

GNFIS		Fuzzy Logic		ANN	
Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
0.00	0.00	0.00	0.00	0.00	0.00
0.10	0.05	0.10	0.05	0.10	0.05
0.20	0.10	0.24	0.13	0.20	0.10
0.30	0.15	0.32	0.15	0.30	0.15
0.40	0.18	0.43	0.18	0.40	0.18
0.50	0.20	0.54	0.20	0.50	0.20
0.62	0.21	0.62	0.21	0.60	0.21
0.68	0.23	0.68	0.23	0.70	0.23
0.71	0.25	0.71	0.25	0.80	0.25
0.83	0.28	0.75	0.40	0.90	0.40
0.87	0.33	0.84	0.50	1.00	0.50
0.92	0.45	0.89	0.60	1.00	0.60
0.98	0.68	0.92	0.70	1.00	0.70
1.00	0.73	0.98	0.80	1.00	0.80
1.00	0.88	1.00	0.90	1.00	0.90
1.00	1.00	1.00	1.00	1.00	1.00

A reliable and perfect diabetes diagnosis test would have 100 percent sensitivity and specificity respectively. The system should positively identify all the true cases of the disease, and never mislabel anything normal as disease. Every designed system should try to strike a balance between sensitivity and specificity especially when a diabetes test is imperfect. Receiver Operator Characteristic (ROC) curve might be plotted, which is formed by sensitivity and 1-specificity on a graph to measure this.

ROC curve plot according to the Check Data of fuzzy logic, neural network and the designed GNFIS system was shown in figure 4.5.

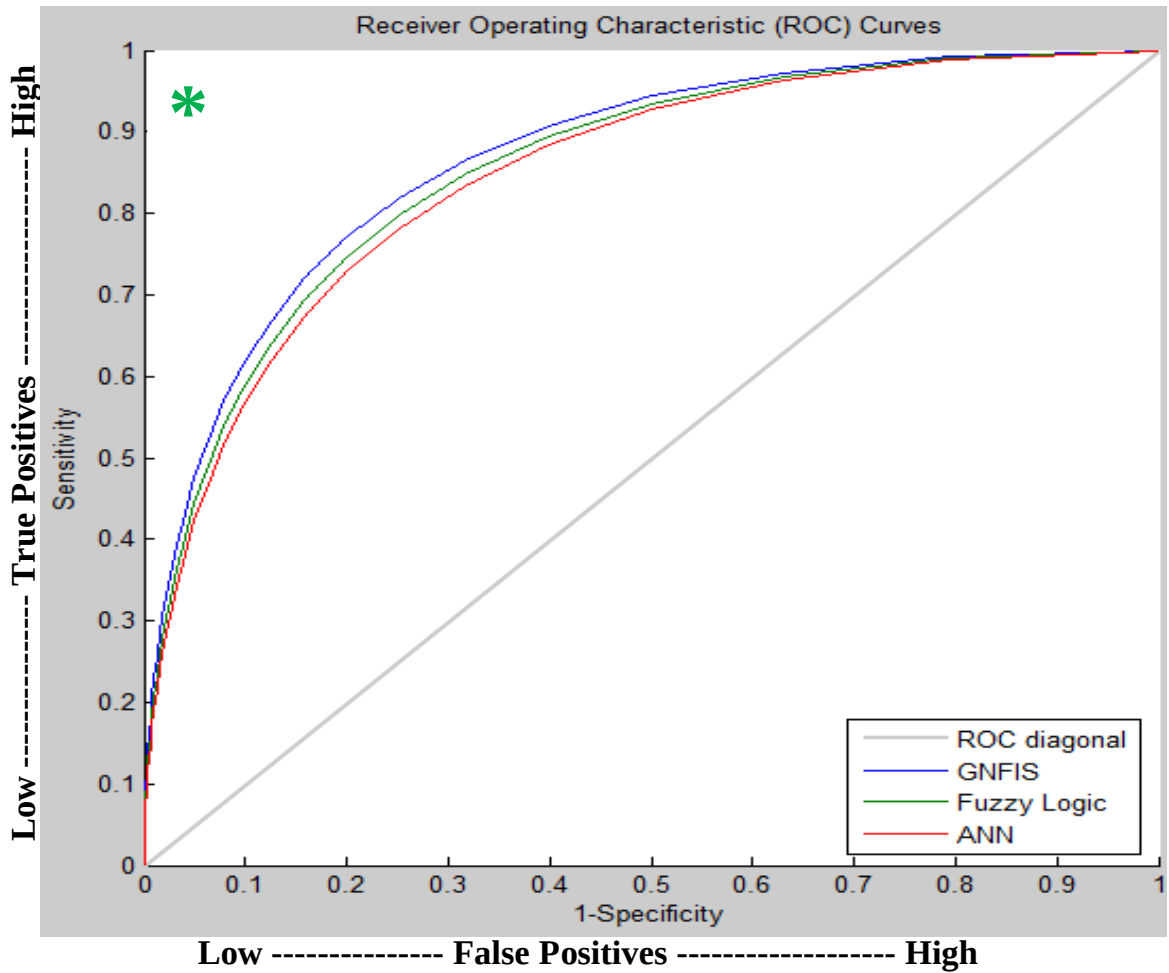


Figure 4.5: ROC curve plot for the check data

The **Blue** line that represent (GNFIS) performed better for diabetes medical diagnosis test than the **Green** and the **Red** line representing fuzzy logic and ANN respectively, because the GNFIS curve comes closer to the **Green** asterisk. The shortest distance from the **Green** asterisk to the **Blue** and **Red** line was pick as the best point along the ROC curve. From the research, **0.98**sensitivity (98 percent) and **0.21** false positive fraction (99 **percent** specificity) was recorded for GNFIS, fuzzy logic gives **0.96** sensitivity (96 percent) and

0.27 false positive fraction (**97 percent** specificity), and ANN gives **0.94**sensitivity (94 percent) and 0.24 false positive fraction (**97 percent** specificity).

The **Green** line representing fuzzy logic gives better medical test than the **Black** line representing neural network, since the curve of the **Green** line comes next to the **Blue** line that is closer to the **Green** asterisk. The best point along the ROC curve was picked looking for the shortest distance from the **Green** asterisk, to the **Red** line.

Table 4.14: Experimental Results for 100 runs.

NTP	PGC	DBP	TSFT	2-HIS	BMI	DPF	AGE	PIMA	FUZZY	ANN	GNFIS
6	148	72	35	0	33.6	0.627	50	1	0.741883	0.734251	0.521435
1	85	66	29	0	26.6	0.351	31	0	0.363793	0.253471	0.181897
8	183	64	0	0	23.3	0.672	32	1	0.712345	0.693421	0.512435
1	89	66	23	94	28.1	0.167	21	0	0.534512	0.734125	0.527341
0	137	40	35	168	43.1	2.288	33	1	0.852278	0.745324	0.534251
5	116	74	0	0	25.6	0.201	30	0	0.113425	0.109876	0.065432
3	78	50	32	88	31	0.248	26	1	0.852278	0.745324	0.534251
10	115	0	0	0	35.3	0.134	29	0	0.108945	0.102832	0.012354
2	197	70	45	543	30.5	0.158	53	1	0.852278	0.745324	0.534251
8	125	96	0	0	0	0.232	54	1	0.852278	0.745324	0.534251
4	110	92	0	0	37.6	0.191	30	0	0.123545	0.134218	0.045321
10	168	74	0	0	38	0.537	34	1	0.827852	0.723745	0.542513
10	139	80	0	0	27.1	1.441	57	0	0.123543	0.156432	0.034521
1	189	60	23	846	30.1	0.398	59	1	0.834521	0.723142	0.564321
5	166	72	19	175	25.8	0.587	51	1	0.832412	0.777342	0.564321
7	100	0	0	0	30	0.484	32	1	0.794321	0.697432	0.555342
0	118	84	47	230	45.8	0.551	31	1	0.421241	0.432152	0.472312
7	107	74	0	0	29.6	0.254	31	1	0.735211	0.632411	0.563212
1	103	30	38	83	43.3	0.183	33	0	0.363542	0.375643	0.453621
1	115	70	30	96	34.6	0.529	32	1	0.694632	0.645321	0.564732
3	126	88	41	235	39.3	0.704	27	0	0.456321	0.401243	0.345261
8	99	84	0	0	35.4	0.388	50	0	0.183425	0.127352	0.045362
7	196	90	0	0	39.8	0.451	41	1	0.657432	0.612346	0.587432
9	119	80	35	0	29	0.263	29	1	0.812456	0.712364	0.586743
11	143	94	33	146	36.6	0.254	51	1	0.867541	0.756432	0.567432
10	125	70	26	115	31.1	0.205	41	1	0.794532	0.745321	0.534211

7	147	76	0	0	39.4	0.257	43	1	0.698432	0.632414	0.523145
1	97	66	15	140	23.2	0.487	22	0	0.462612	0.384632	0.435261
13	145	82	19	110	22.2	0.245	57	0	0.384631	0.352183	0.483521
5	117	92	0	0	34.1	0.337	38	0	0.219352	0.124352	0.046532
5	109	75	26	0	36	0.546	60	0	0.123641	0.128347	0.047631
3	158	76	36	245	31.6	0.851	28	1	0.810324	0.674352	0.567483
3	88	58	11	54	24.8	0.267	22	0	0.456323	0.453217	0.403251
6	92	92	0	0	19.9	0.188	28	0	0.346531	0.234152	0.021465
10	122	78	31	0	27.6	0.512	45	0	0.204352	0.325647	0.126352
4	103	60	33	192	24	0.966	33	0	0.495643	0.456321	0.484362
11	138	76	0	0	33.2	0.42	35	0	0.123549	0.173522	0.045362
9	102	76	37	0	32.9	0.665	46	1	0.756432	0.703425	0.674631
2	90	68	42	0	38.2	0.503	27	1	0.699999	0.689731	0.534211
4	111	72	47	207	37.1	1.39	56	1	0.734216	0.746321	0.674362
3	180	64	25	70	34	0.271	26	0	0.456732	0.342512	0.423142
7	133	84	0	0	40.2	0.696	37	0	0.123545	0.162435	0.093211
7	106	92	18	0	22.7	0.235	48	0	0.201354	0.213541	0.125345
9	171	110	24	240	45.4	0.721	54	1	0.812465	0.801235	0.534251
7	159	64	0	0	27.4	0.294	40	0	0.176523	0.193721	0.0564321
0	180	66	39	0	42	1.893	25	1	0.899432	0.843521	0.503212
1	146	56	0	0	29.7	0.564	29	0	0.231541	0.218321	0.200931
2	71	70	27	0	28	0.586	22	0	0.32173	0.421621	0.264532
7	103	66	32	0	39.1	0.344	31	1	0.634211	0.732412	0.645321
7	105	0	0	0	0	0.305	24	0	0.123541	0.098732	0.003452
1	103	80	11	82	19.4	0.491	22	0	0.435218	0.432154	0.321623
1	101	50	15	36	24.2	0.526	26	0	0.384721	0.495732	0.235421
5	88	66	21	23	24.4	0.342	30	0	0.345212	0.324153	0.234521
8	176	90	34	300	33.7	0.467	58	1	0.835212	0.756432	0.674321

7	150	66	42	342	34.7	0.718	42	0	0.432612	0.463721	0.346215
1	73	50	10	0	23	0.248	21	0	0.212631	0.321732	0.239483
7	187	68	39	304	37.7	0.254	41	1	0.799843	0.735212	0.645321
0	100	88	60	110	46.8	0.962	31	0	0.231521	0.236452	0.125364
0	146	82	0	0	40.5	1.781	44	0	0.203421	0.127342	0.035211
0	105	64	41	142	41.5	0.173	22	0	0.342512	0.312832	0.4532142
2	84	0	0	0	0	0.304	21	0	0.182341	0.129352	0.003421
8	133	72	0	0	32.9	0.27	39	1	0.647321	0.643251	0.512345
5	44	62	0	0	25	0.587	36	0	0.213542	0.312743	0.182411
2	141	58	34	128	25.4	0.699	24	0	0.217211	0.432132	0.423142
7	114	66	0	0	32.8	0.258	42	1	0.624121	0.598765	0.723151
5	99	74	27	0	29	0.203	32	0	0.372141	0.312521	0.352141
0	109	88	30	0	32.5	0.855	38	1	0.624121	0.598765	0.612431
2	109	92	0	0	42.7	0.845	54	0	0.215231	0.213641	0.125341
1	95	66	13	38	19.6	0.334	25	0	0.401243	0.346352	0.432411
4	146	85	27	100	28.9	0.189	27	0	0.321521	0.321821	0.321521
2	100	66	20	90	32.9	0.867	28	1	0.843521	0.675463	0.521421
5	139	64	35	140	28.6	0.411	26	0	0.432161	0.394532	0.473251
13	126	90	0	0	43.4	0.583	42	1	0.602142	0.584765	0.546321
4	129	86	20	270	35.1	0.231	23	0	0.312431	0.374532	0.231523
1	79	75	30	0	32	0.396	22	0	0.401243	0.372153	0.172834
1	0	48	20	0	24.7	0.14	22	0	0.352121	0.382512	0.321621
7	62	78	0	0	32.6	0.391	41	0	0.312536	0.234121	0.162432
5	95	72	33	0	37.7	0.37	27	0	0.312531	0.325121	0.352181
0	131	0	0	0	43.2	0.27	26	1	0.599873	0.585743	0.501243
2	112	66	22	0	25	0.307	24	0	0.432411	0.435211	0.482111
3	113	44	13	0	22.4	0.14	22	0	0.473821	0.483291	0.343621
2	74	0	0	0	0	0.102	22	0	0.201162	0.118112	0.083452

7	83	78	26	71	29.3	0.767	36	0	0.453211	0.490832	0.432416
0	101	65	28	0	24.6	0.237	22	0	0.392512	0.235127	0.324153
5	137	108	0	0	48.8	0.227	37	1	0.623511	0.756432	0.835213
2	110	74	29	125	32.4	0.698	27	0	0.374532	0.312431	0.435211
13	106	72	54	0	36.6	0.178	45	0	0.372153	0.401243	0.483291
2	100	68	25	71	38.5	0.324	26	0	0.382512	0.352121	0.118112
15	136	70	32	110	37.1	0.153	43	1	0.656431	0.756432	0.564321
1	107	68	19	0	26.5	0.165	24	0	0.372153	0.215231	0.473821
1	80	55	0	0	19.1	0.258	21	0	0.382512	0.401243	0.201162
4	123	80	15	176	32	0.443	34	0	0.234121	0.321521	0.453211
7	81	78	40	48	46.7	0.261	42	0	0.325121	0.325163	0.392512
4	134	72	0	0	23.8	0.277	60	1	0.685643	0.656432	0.546321
2	142	82	18	64	24.7	0.761	21	0	0.482111	0.345261	0.435211
6	144	72	27	228	33.9	0.255	40	0	0.343621	0.483721	0.483291
2	92	62	28	0	31.6	0.13	24	0	0.083452	0.125341	0.118112
1	71	48	18	76	20.4	0.323	22	0	0.432416	0.345596	0.490832
6	93	50	30	64	28.7	0.356	23	0	0.324153	0.235123	0.192354
1	122	90	51	220	49.7	0.325	31	1	0.846321	0.846532	0.635211

From the output of the 100 samples runs using the Pima Indians Diabetes dataset presented in Table 4.14, any result that is ≥ 0.5 the system diagnosed the patient to have diabetes (1) and if the result is ≤ 0.4 the patient is diagnosed to be negative (0). GNFIS expert system shown diagnosis accuracy of 98%. Out of 100 input data, 98 showed correct diagnosis. The input data which falls between points 0.1 and 0.4 could be taken to be non-

diabetes and low risk diabetic respectively with fuzzy logic does approximate reasoning.

The approximate reasoning of this system is an example of handling uncertain data.

Table 4.15: The Performance of Three Systems for Diabetes Mellitus Incidence.

True classification in each dataset		Fuzzy Logic		Artificial neural network		GNFIS	
		+ ^b	- ^c	+ ^b	- ^c	+ ^b	- ^c
Training Dataset							
True ^a	+	250	62	260	52	255	56
	-	89	239	118	210	83	246
Accuracy (%)		75.95		73.52		78.27	
Sensitivity (%)		79.68		83.47		81.87	
Specificity (%)		72.40		64.08		74.86	
Test Dataset							
True ^a	+	114	30	118	29	112	31
	-	36	102	43	92	34	105
Accuracy (%)		76.54		72.59		76.97	
Sensitivity (%)		79.40		79.40		78.11	
Specificity (%)		73.54		65.47		75.78	
Total Dataset							
True ^a	+	255	73	257	16	260	10
	-	13	427	11	484	9	489
Accuracy (%)		96.14		95.14		99.34	
Sensitivity (%)		95.14		94.14		97.12	
Specificity (%)		97.40		97.39		98.33	

“True +” and “True -”represents the total sum of true positive and true negative.

+^brepresents the total sum of predictive positive.

-^crepresents the total sum of predictive negative.

The Research Findings are:

1. The GNFIS uses 5 attributes as the minimum and 8 as the maximum attributes.
2. GNFIS recorded 98.33% for the minimum and maximum diagnosis accuracy and 98.76% respectively and the average accuracy both is 98.55 %.
3. From the Pima diabetes dataset, out of 768 instances, 511 samples were correctly classified by the expert system (GNFIS) and 257 samples were detected as outliers. The outlier detection percentage is 33.46.
4. Age, diabetes pedigree function, plasma glucose concentration, BMI and 2-Hour Serum Insulin are considered as critical attributes for Pima diabetes dataset.

Table 4.16: Comparison of the Developed GNFIS with the Existing Systems

S/N	Author (Year)	Method	Accuracy (%)
1.	Ahmad F. et al (2013)	Improved GA	80.4
2.	Ravi et al (2014)	GA + SVM	77.3
3.	Nihat Yilmaz et al (2014)	K-Means + SVM	93.65
4.	Tejashri & Satish (2015)	ANN1(FNN) fed 2nd ANN2	84.24
5.	Santhanam & Padmavathi (2015)	K-Means+GA+SVM	98.82
6.	Oladipo (2017)	GNFIS	99

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This chapter presents the summary, conclusion drawn from the research conducted, recommendations, suggestion for future research and contribution to knowledge.

5.1 Summary

With rapid increase of diabetes in developed countries, the disease can be described as epidemic. More than 246 million people worldwide were being treated for this chronic disease according to information from International Diabetes Foundation in 2005.

Artificial intelligence techniques in medical diagnosis cannot be underestimated as it aided an increase in medical diagnosis accuracy and they have greatly imparted in medical diagnosis. This research, therefore developed and enhanced a diagnosis for diabetes mellitus using Genetic-Neuro-Fuzzy Inferential System (GNFIS).

In this work direct rating method was used for acquiring data. This method is the most direct means of collecting data which is more efficient for constructing a membership function. Also, the data (subject) were presented in a series of data (objects) to the domain experts and asked to rate the membership function. In order to construct a membership function, the responses taken from several physicians were aggregated. The lowest and highest values were considered to form the ranges for membership function calculation. The lowest was used as the minimum and highest as the maximum values.

5.2 Conclusion

The results of the research indicated that the GNFIS is the best method, with 96% accuracy on the whole dataset used. The fuzzy logic model gave accuracy of 94%. The ANNs model recorded the least accuracy with 92% accuracy of the three models. From the sensitivity analysis, as shown in Table 4.13, age, family history of diabetes, BMI and preference for salty food played an important role in the incidence of diabetes dataset, and the results provide evidence for the prevention of diabetes through community interventions. This research has implemented Genetic Algorithms for dimensionality reduction and Neural Network and Fuzzy Logic to diagnose the diabetes data set.

5.3 Recommendations

The GNFIS diagnose accuracy of the proposed system has an increase of 2.08% over neural network and fuzzy logic. However, the accuracy of the system can be increased using other combinations of other machine learning algorithms. In addition, this application runs on stand- alone system. In future work, the system can be applied on high performance client server and mobile applications, thus client server application installed can request and receive information over the network which will be easier to access by all users.

5.4 Suggestion for Future Research

The study recommended that further Data mining techniques can be effectively used to help healthcare professionals in diagnosing and providing suitable treatments for diabetes disease. Also, it was recommended that GNFIS approach should be applied to other complex real-world diagnosis problems.

5.5 Contribution to Knowledge

The contribution to knowledge of this research is as follows:

- (i) Hybridization of more than one classifier was carried out which shows an enhanced result with GNFIS having an accuracy of 96%
- (ii) Generic Algorithm which was used to obtain most relevant attribute from Pima Indians Diabetes Dataset in the Neuro – Fuzzy Inferential System to diagnose diabetes mellitus improves the accuracy of the developed system
- (iii) Comparison of GNFIS with existing system showed that GNFIS had a better performance.

REFERENCES

- Abbod, M. F., Keyserlingk, D. G., Linkens, D. A. & Mahfouf, M. (2001). Survey of Utilization of Fuzzy Technology in Medicine and Healthcare. *Fuzzy Sets and Systems*, 120, 331-349.
- Abdulla, M. & Bakhsh, M. (2011). An Expert System of Determining Diabetes Treatment Based on Cloud Computing Platforms. (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, 2(5), 1982-1987.
- Abebe, A., Solomatine, D. & Venneker, R. (2000). Application of Adaptive Fuzzy Rule-Based Models for Reconstruction of Missing Precipitation Events. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, 45(3), 425-436.
- Acar, E., Özerdem, M. S. & Akpolat, V. (2001). "Diabetes Mellitus Forecast Using Various Types of Artificial Neural Network," *International Advanced Technologies Symposium (IATS'11)*, 1-6.
- Adeli, A. & Mehdi, N. (2010). "A Fuzzy Expert System for Heart Disease Diagnosis". *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, 1, 1-6.
- Adlassnig, K. P. (2001). Fuzzy Systems in Medicine. *Proceedings of the 2nd International Conference in Fuzzy Logic and Technology*, Leicester, United Kingdom, September 5-7, 2-5, Web URL, http://www.eusflat.org/proceedings/EUSFLAT_2001/papers/002_Adlassnig.pdf.
- Agboizebeta, I. A. & Chukwuyeni, O. J. (2012). Cognitive Neuro-Fuzzy Expert System for Hypotension Control. *Computer Engineering and Intelligent Systems*. 3(6), 21-31, www.iiste.org.
- Ahmad, F., Isa, N. A., Hussain Z. & Osman, M. K. (2013). Intelligent Medical Disease Diagnosis Using Improved Hybrid Genetic Algorithm-Multilayer Perceptron Network. *J Med System*, 37(2), 9934-9945
- Akinjinmi, A. A., Adeyooye, O. A, Akingbade, O. A. & Okerentugba, P. O. (2014). Prevalence of Diabetes Mellitus in Abeokuta, Ogun State, Nigeria. *Researcher*, 6 (2), 73-75 <http://www.sciencepub.net/researcher>.
- Akter, M., Uddin Shorif, M. & Haque, A. (2008). "Diagnosis and Management of Diabetes Mellitus through a Knowledge-Based System", *13th International Conference on Biomedical Engineering ICBME*, 1000-1003.

- Ali, K. (2008). *Developing an Expert-System for Diabetics by Supporting with ANFIS*. Unpublished thesis, Institute of Science and Computer Engineering. T.C. Bahçeşehir Üniversitesi
- Al-Shayea, Q. K. (2011). Artificial Neural Networks in Medical Diagnosis. *IJCSI International Journal of Computer Science*, 8(2), 150–154.
- Ambica, A., Satyanarayana, G. & Amarendra, K. (2013). An Efficient Expert System for Diabetes by Naïve Bayesian Classifier. *International Journal of Engineering Trends and Technology (IJETT)* 4(10), 4634 – 4639.
- Ambilwade, R. P., Manza, R. R. & Gaikwad, P. B. (2014). Medical Expert Systems for Diabetes Diagnosis. A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*. 4(11), 647 – 652. www.ijarcse.com
- American Diabetes Associate (2006), "Diagnosis and classification of diabetes mellitus", *Diabetes care*, Suppl 1. S43-8.
- American Diabetes Association (2014), Diagnosis and Classification of Diabetes Mellitus, *Diabetes Care*, vol 27, Supplement 1. S63-9.
- Arcanjo, F. L., Gisele, L., Pappa, P. V., Bicalho, W. M. & Altigran, S. da Silva (2011). Semi-supervised Genetic Programming for Classification, *GECCO'11*, July 12–16. Dublin, Ireland.
- Aruna P., Puviarasan, N., & Palaniappan, B. (2005). An Investigation of Neuro-Fuzzy System in Psychosomatic Disorders. *Expert Systems with Applications*. 28(4), 673 - 679
- Baha, B. Y. & Wajiga, G. M. (2012). Artificial Neural Networks to Detect Risk of Type 2 Diabetes. *ORIND* 10 (2), 120 - 125 www.transcampus.org/journals, www.ajol.info/journals/jorind
- Bojarczuk, C. C. Lopes, H. S. & Freitas, A. A. (2000). Genetic Programming for Knowledge Discovery in Chest Pain Diagnosis, *IEEE Eng.Med. Mag.*, 19(4), 38–44.
- Chan, C. K. Y., Ramadas, A., Quak, K. F. & Oldenburg, B. (2011). Web-based Interventions for Management of Type 2 Diabetes Mellitus: A Systematic Review of Recent Evidence *International Journal of Medical Informatics*. 80(6), 389 – 405
- Chang-Shing, L. & Mei-Hui (2011). A Fuzzy Expert System for Diabetes Decision Support Application, *IEEE Transactions on Systems, Man and Cybernetics*, Part B: Cybernetics 41(n.i), 139 – 153

- Chen, C. W., Chen, P. C., & Chiang, W. L. (2011). Stabilization of Adaptive Neural Network Controllers for Nonlinear Structural Systems Using a Singular Perturbation Approach, *Journal of Vibration and Control* 17(8), 1241–1252.
- Cheung, Y. B., Wee, H.L., Li, S.C., Fong, K.Y. & Thumboo, J. (2005). The Impact of Diabetes Mellitus and other Chronic Medical Conditions on Health-Related Quality of Life: Is The Whole Greater Than The Sum of Its Parts? *Health Qual Life Outcomes* 3(2), 15 – 28, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC548134>
- Chien, B.C., Lin, J. Y. & Hong, T. P. (2002). Learning Discriminant Functions with Fuzzy Attributes for Classification Using Genetic Programming, *Expert Syst. Applicat.*, 23, 31–37.
- Chioma, Obinna (2014, 19th November), Nigeria Diabetic Patients Face Treatment Challenges in Nigeria, *Vanguard Newspaper Media Ltd.*, 9,
- Dazzi, F., Simpson, E. & Goldman, J. M. (2001). Minor Antigen Solves Major Problem. *Nature Medicine*. Vol: 7. www.webofknowledge.com
- De Falco, I., Della Cioppa, A. & Tarantino, E. (2002). Discovering Interesting Classification Rules with Genetic Programming, *Appl. Soft Comput.*, 23, 1–13.
- Delen, D., Walker, G., & Kadam, A. (2005). Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods. *Artificial Intelligence in Medicine* 34(2), 113–127.
- Duckstein, L., Blinowska, A. & Verroust, J. (1995). Fuzzy Classification of Patient State with Application to Electrodiagnosis of Peripheral Polyneuropathy. *IEEE Trans Biomed Eng.* 42(8), 786–92.
- Elson, R. C. (2002). Inhibitor Synchronization of Bursting in Biological Neurons: Dependence on Synaptic Time Constant. *AJP-JN Physiol.* 88 (3), 1166-1175.
- Engelmore, R. S. & Feigenbaum, E. (2007). *Expert Systems and Artificial Intelligence*.http://www.wtec.org/loyola/kb/c1_s1.htm
- Ephzibah, E. P. (2011). Cost Effective Approach on Feature Selection Using Genetic Algorithms and Fuzzy Logic for Diabetes Diagnosis. *International Journal on Soft Computing (IJSC)*, 2(1), 1-10.
- Galina S., Mariusz D., Wioletta S., Monika P. & Tomasz, K. (2011). Artificial Intelligence Approach to Diabetes Diagnosis, *Journal of Society for Bio-Diabetes Research*, 252-262.
- Harun, U. (2010). “A Biomedical System Based on Artificial Neural Network and Principal Component Analysis for Diagnosis of the Heart Valve Diseases”, *Springer Science+Business Media, LLC*. 61-72.

- Hasan, T., Nijat, Y. & Feyzullah, T. (2009) "A Comparative Study on Diabetes Disease Using Neural Networks", *Expert System with Applications*, 36, 8610-8615.
- Hopfield, J. J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *In Proceedings of the National Academy of Sciences of the United States of America*, 79, 2554-2558.
- Hu, F. B. (2011). Globalization of Diabetes: The role of diet, lifestyle, and Genes genes. *Diabetes Care*, 34(6), 1249–1257. <http://doi.org/10.2337/dc11-0442>
- Ibrahim, M., Ahmed, M., Alfonse, M., & Abdel-Badeeh, M. S. (2015). Reasoning Techniques for Diabetics Expert Systems. *Science Direct. Procedia Computer Science*, 65, 813 – 820
- International Diabetes Federation (2010). IDF Member Association Consultation on Diabetes Priorities for the UN Summit on NCDs. Brussels.
- Jang, J. S. R., Sun, C. T. & Mizutani, E. (1996). *Neuro-Fuzzy and Soft Computing*. Prentice Hall, Upper Engineering, MIT Press, Cambridge. MA
- Jonh, R. I., & Innocent, P. R. (2005). Modeling Uncertainty In Clinical Diagnosis Using Fuzzy Logic, *IEEE Transaction System Man and Cybernetic*, 35(6), 1340-1350
- Kahramanli, H., & Allahverdi, N. (2008). Design of a Hybrid System for the Diabetes and Heart Diseases, *Expert Systems with Applications*, 35(1-2), 82-89.
- Kalpana, M. & Senthil Kumar, A. V. (2012). Similarity Measure Between Fuzzy Set, Fuzzy Numbers and Fuzzy Rules Using T Fuzzy Assessment Methodology. *Australian Journal of Basic and Applied Sciences*, 7(8), 72-80,
- Kaur, P. & Chakraborty, S. (2007). A New Approach to Vendor Selection Problem with Impact Factor as an Indirect Measure of Quality. *Journal of Modern Mathematics and Statistics*, 1(1), 8-14.
- Kavitha, S. & Duraiswamy, K. (2011). Adaptive Neuro-Fuzzy Inference System Approach for the Automatic Screening of Diabetic Retinopathy in Fundus Images. *Journal of Computer Science*, 7(7), 1020-1026
- Kishore, J. K., Patnaik, L. M., Mani, V. & Agrawal, V. K. (2000). Application of Genetic Programming for Multicategory Pattern Classification, *IEEE Trans. Evol. Comput.*, 4, 242–258.
- Kumar, K.A. & Punithavalli, M. (2011). Efficient Cancer Classification Using Fast Adaptive Neuro-Fuzzy Inference System (FANFIS) Based on Statistical Techniques, *International Journal of Advanced Computer Science and Applications*, Special Issue on Artificial Intelligence, 132–137.

- Kumar, S. & Kumaravel, A. (2013). "Diabetes Diagnosis Using Artificial Neural Network", *International Journal of Engineering Sciences & Research Technology*, 2(6),1642-1644.
- Kuncheva, L. I. & Steimann, F. (1999). Editorial - Fuzzy Diagnosis. *Artificial Intelligence in Medicine*, 16, 121–128
- Laleh, K., Ali, F., Shahriar, G. & Fatollah, M. (2008). Application of Fuzzy Logic Controller for Intensive Insulin Therapy in Type 1 Diabetic Mellitus Patients by Subcutaneous Route. *WSEAS transactions on systems and control*, 3(9), 712-721
- Lalka, N. & Jain, S. (2015). Comparative Study of Type-1 Fuzzy Logic and Type-2 Fuzzy Logic. *International Journal of Computer Applications*, 124(16), 14-21
- Lavanya, K., Durai, M. A. & N. Ch. Sriman, N. I., (2011), "Fuzzy Rule Based Inference System for Detection and Diagnosis of Lung Cancer", *International Journal of Latest Trends in Computing*, 2(1), 165-171
- Ledley, R. S. (1960): *Digital Computer and Control Engineering*, New York, Toronto, London: McGraw-Hill.
- Lehmann, D. E. (2004). "Computerised Decision-Support Tools in Diabetes Care: Hurdles to Implementation", *Diabetes Technology & Therapeutic*, 6(3), 15-31.
- Liao, S. H. (2005). Expert System Methodologies and Applications: A Decade Review From 1995-2004. *Expert Systems with Applications*, 28(1), 93-103.
- Lim, T. S., Loh, W.-Y. & Shih, Y. S. (2000). A Comparison of Prediction Accuracy, Complexity and Training Time of Thirty Three Old and New Classification Algorithms, *Mach. Learning J.*, 40, 203–228.
- Lin, R. (2009). An Intelligent Model for Liver Disease Diagnosis. *Artificial Intelligence in Medicine*. 47(1), 53–62.
- Mahfouf, M., Abbod, M. F., & Linkens, D. A. (2001). A Survey of Fuzzy Logic Monitoring and Control Utilization in Medicine. *Artificial Intelligence in Medicine*, 21, 27-42.
- Manda, R. N., Rao, G. R. Srishar, K. M., & Allam, A. R. (2010). A Clinical Decision Support System Using Multi-Layer Perception Neural Network to Predict Quality of Life in Diabetes. *Diabetes & Metabolic Syndrome: Clinical Research & Review* 4, 57 – 59
- Marco, A. F. Pimentel, D. A. Clifton, Peter J. W. & Lionel, T. (2013). Modeling Physiological Deterioration in Post-Operative Patient Vital-Sign Data. *Springer Med BiolEng Comput*. 869-877.

- Mastorocostas, P. & Theocharis, J. (2006). A Stable Learning Algorithm for Block-Diagonal Recurrent Neural Networks: Application to the Analysis of Lung Sounds. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(2), 242-254.
- Mendes, R. R. F., Voznika, F. B., Freitas, A. A. & Nievola, J. C. (2001). Discovering Fuzzy Classification Rules with Genetic Programming and Co-Evolution, in *Lecture Notes in Artificial Intelligence*, 2168, *Proc. 5th Eur. Conf. PKDD*, 314–325.
- Mohammad, F., Azam Q., Javad, H., Javad, V. & Hadi, V. (2013). Introduction of a Method to Diabetes Diagnosis According to Optimum Rules in Fuzzy Systems based on Combination of Data Mining Algorithm (D-T), Evolutionary Algorithms (ACO) and Artificial Neural Networks (NN). *Journal of Mathematics and Computer Science* 6, 272 – 285
- Muni, D. P., Pal, N. R. & Das, J. (2004). A Novel Approach to Design Classifiers Using Genetic Programming, *IEEE transactions on evolutionary computation*, 8(2), 183-196.
- Nidhi, M. & Jha, P. (2014). A Review on the Applications of Fuzzy Expert System for Disease Diagnosis. *International Journal of Advanced Research in Engineering and Applied Sciences*. 3(12), 28-43
- Nonso N., Farath A., David E., Jiten V. & James N (2015). Model for Type 2 Diabetes Management: A Tool for Regimen Alterations. *Journal of Computer Sciences and Applications*, 3(3), 40-45
- Novruz, A. (2014). Design of Fuzzy Expert Systems and Its Applications in Some Medical Areas. *International Journal of Applied Mathematics, Electronics and Computers Advanced Technology and Science*, 2(1), 1–8
- Neagoe,V. E. Latin, L.F. and Grunwald, S. (2003). A neuro-fuzzy approach to classification of ECG signals for ischemic heart disease diagnosis, in: *AMIA Annual Symposium Proceedings*, 494–498.
- Nihat, Y., Onur, I. & Mustafa, S. U. (2014). A New Data Preparation Method Based on Clustering Algorithms for Diagnosis Systems of Heart and Diabetes Diseases. Springer: Transaction Processing Systems: *J Med Syst*; 38:48.
- Omisore, M. O., Samuel, O. W. & Atajeromavwo, E. J. (2017). A Genetic-Neuro-Fuzzy inferential model for diagnosis of tuberculosis. *Saudi Computer Society, King Saud University Applied Computing and Informatics*, 13, 27–37.

- Parker, S. G., Gorry, G. A., Kassiner, J. P. & Schwartz, W. B. (1976): Towards the Simulation of Clinical Cognition: Taking a Present Illness by Computer. *American Journal of Medicine*, 60, 981-996.
- Pedrycy, W. & Gomide, F. (1998). *An Introduction to Fuzzy Sets, Analysis and Design*, MIT Press, Cambridge Massachusetts.
- Perez-Ojeda, A. (1976). *Medical Knowledge Network. A Database for Computer Aided Diagnosis*. Unpublished Dissertation, Department of Industrial Engineering, University of Toronto.
- Phoung, N. H., & Kreinovich, V. (2001). Fuzzy logic and its applications in medicine. *International of Medical Informatics*, 62, 2-3
- Polat K. & Gunes, S. (2007). An expert system approach based on principal component analysis and adaptive neurofuzzy inference system to diagnosis of diabetes disease, *Dig. Signal Process.*, 17(4), 702- 710
- Pople, H. E., Myers, J. D. & Miller, R. A. (1975). DIALOG: A Model of Diagnostic Logic for Internal Medicine. 4. *International Joint Conference on Artificial Intelligence, Tblis: USSR, Sept. 3-8*, 848-855
- Prasath, N. Lakshmi, M. Nathiya, N. B. & Neetha N.P (2013) A Survey on the Applications of Fuzzy Logic in Medical Diagnosis. *International Journal of Scientific & Engineering Research*, 4(4), 1199-1203
- Priti, S. S. (2017). Computer Aided Development of Fuzzy, Neural and Neuro-Fuzzy System. *International Journal of Engineering and Applied Computer Science (IJEACS)* 2(1), 10 – 17.
- Pradhan, M. A., Rahman, A., Acharya, P., Gawade, R. & Pateria, A. (2011). Design of Classifier for Detection of Diabetes using Genetic Programming, *International Conference on Computer Science and Information Technology (ICCSIT'2011) Pattaya*, 125-130.
- Radha, I.R. & Rajagopalan, Z.S.P. (2007). "Fuzzy logic approach for diagnosis of diabetic". *Information Technology Journal*, 6 (1), 96-102.
- Rahman, A., Jamil, H., Muhammad, H. S., Maqbool, H., & Sungyoung, L. (2015). H2RM: A Hybrid Rough Set Reasoning Model for Prediction and Management of Diabetes Mellitus. *Sensors* 2015, 15, 15921-15951
- Rajeswari, K. & Vaithyanathan, V. (2011). Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network. *International Journal of Computer Science and Network Security*, 11(4), 126-130

- Rama, D. E., & Nagaveni D. N (2010) Design Methodology of a Fuzzy Knowledgebase System to predict the risk of Diabetic Nephropathy. *International Journal of Computer Science Issues*, 7(5), 1694-0814.
- Rashedur M. & Rahman, F. A. (2013). Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis. *Journal of Software Engineering and Applications*. 85-97
- Roudsari, A. (2004), A web-based diabetes management system, *Transaction of the Institute of Measurement and Control*, 26 (3), 201-222.
- Rao, P. N., Srinivas D., Lakshmi, T. S. & Suneetha, S. (2002). Bacterial index of granuloma and its relevance compared to BI of skin smears. *Lepr Rev.*, 73, 79-80.
- Ravi, G. K., Ramachandra, G. A. & Nagamani, K. (2014). An Efficient Feature Selection System to Integrating SVM with Genetic Algorithm for Large Medical Datasets. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(2), 272-277
- Ryman, D. H. (1992). Computer Assisted Medical Diagnosis Problems, and Methods to Minimize their Effects. *Naval Health Research Center*. San Diego, CA 92186-5122
- Santosh, K. & Kumaravel, A. (2013). Diabetes Diagnosis Using Artificial Neural Network. *International Journal of Engineering Science & Research Technology* 2(6), 1642 – 1644.
- Saritas, I. A. & Sert. U. (2003), A Fuzzy expert system design for diagnosis of prostate cancer. *Proceedings of International conference on Computer Systems and Technologies, Konya, Turkey*.
- Seyedeh, Tawfik, & Yousef (2013). Developing an Expert System for Diabetics Treatment Advices. *International Journal of Hospital Research*. 2(3), 155-162.
- Shortliffe, E. H. & Blois, M. S. (2001) The computer meets medicine and biology: emergence of a discipline. In: Shortliffe EH, Perreault LE, Wiederhold G, et al, eds. *Medical Informatics computer applications in health care and biomedicine*. New York: Springer-Verlag: 3–40.
- Santhanam, M. & Padmavathi, M. S. (2015). An Efficient Model by Applying Genetic Algorithms for Outlier Detection in Classifying Medical Datasets, *Australian Journal of Basic and Applied Sciences*, 9(27), 583-591.
- Sikchi, S. S., Sikchi, S. & Ali M. S. (2013). Fuzzy Expert Systems (FES) for Medical Diagnosis. *International Journal of Computer Applications*, 6(11), 7-16

- Singh, N., Rahul, M. N. S. & Yaduvir, S. (2011). Genetic algorithms: concepts, design for optimization of process controllers. *Canadian Center of Science and Education, Computer and Information Science*, 4(2), 39-54
- Singh, A. B., Cariappa, M. P., Gupta, R. K. & Gupta, R. (2014). “Importance of Periodic Health check up for Indian Women”. *Journal of Evolution of Medical and Dental Sciences*, 3(28), 39-56
- Singh, R. J., & Singh, W. (2014). Data Mining in Healthcare for Diabetes Mellitus, *International Journal of Science and Research*, 3(7), 1993 - 1998.
- Soha, A. (2015). *Genetic Programming for Biomarker Detection in Classification of Mass Spectrometry Data*. Unpublished thesis submitted to the Department of Computer Science, Victoria University of Wellington, New Zealand.
- Sreedevi, E. & Padmavathamma, M. (2015). A Threshold Genetic Algorithm for Diagnosis of Diabetes using Minkowski Distance Method. *International Journal of Innovative Research in Science, Engineering and Technology* 4(7), 5596-5601
- Stelfox, P. C., Scurlock, E.J. & Bates, D.W. (2006). The to err is human report and the patient safety literature. *Quality and Safety in Health Care*, 15(3): 174–178
- Sumathi, O. M. & Mudigonda, M. (2015). A Research Review on Multisensor Data Fusion Techniques in Intelligent Health Care Monitoring. *International Journal of Advanced Engineering Research and Applications*, 1(4), 157-162
- Sumathy, M., Praveen, K., Jishnujit, T. M. & Ranjith, K. (2010). Diagnosis of Diabetes Mellitus based on Risk Factors, *International Journal of Computer Applications*, 10(4), 1- 4
- Tadic, D. Cvjetkovic, V. & Milovanovic, D. (2009). Determining and Monitoring of the Therapy Procedures by Application of the Artificial Intelligence Methods Relevant for Acquiring of the Quality Excellence in the Processes of the Medical Treatment, *International Journal for Quality Research*, 3(3), 1-7.
- Tadić, D., Popović, P. & Đukić, A. (2010). A Fuzzy Approach to Evaluation and Management of Therapeutic Procedure in Diabetes Mellitus Treatment. *Yugoslav Journal of Operations Research*, 20(1), 99-116
- Tan, A., Quek, C., & Yow, K, C. (2005). Maximizing Winning Trades Using a Novel RSPOP Fuzzy Neural Network Intelligent Stock Trading System. *Appl Intell*, 29, 116–128.
- Tawfik, S. Z., Mohammad, V. M. &Yousef, A. S. (2012): An Expert System for Diabetes Diagnosis, *American Academic & Scholarly Research Journal*, 4(5), 1-13.

- Tejashri, N. Giri & Satish, R. Todmal (2015). Prognosis of Diabetes using Neural Network, Fuzzy Logic, Gaussian Kernel Method. *International Journal of Computer Applications* (0975 – 8887) 124(10), 21-34
- Temurtas, F. (2009). A Comparative Study on Thyroid Disease Diagnosis Using Neural Networks. *Expert Systems with Applications*, 36, 944–949
- Tudor, R. S., Hovorka, R., Cavan, D. A., Meeking, D., Hejlesen, O. K. & Andreassen, S. (1998). Dias-niddm-a Model-Based Decision Support System for Insulin Dose Adjustment in Insulin-Treated Subjects with Niddm. *Computer Methods and Programs in Biomedicine* 56: 175
- Überla, K. (1965): Zur Verwendung der Faktorenanalyse in der Medizinischen Diagnostik. *Methods of Information in Medicine*, 2. 89-92.
- Vijaya, L. K. & Padmavathamma, M. (2014) A Model to Predict Diabetes Based on Chromosomes Using Genetic Algorithm. *International Journal Of Innovative Research & Development*. 3(7), 345-367. www.ijird.com
- Vithyatheri, G., Vimala, B. & Huck-Soo, L. (2012). Using Rule-Based Reasoning and Object-Oriented Methodologies to Diagnose Diabetes. *Journal of Social Sciences* 8 (1): 66-73.
- Wang, X., Amy, C. & Noémie, E. (2008). Automated Knowledge Acquisition from Clinical Narrative Reports *AMIA Symposium Proceedings*, 783-787
- Woodbury, M. A. (1963). The Inexplicabilities of Bayes Theorem to Diagnosis. *Proc. Fifth Int. Conf. on Medical Electronics*. Liege, Belgium. *Springfield, Ill., Charles C. Thomas*. 860-868.
- Wrong, H.J, Legnini, M.W., & Whitmore, H.H. (2000), "The Diffusion of Decision Support Systems in Healthcare: Are we there Yet?", *Journal Healtch management*, 45. 249-253.
- Yue Wu, Biaobiao, Z., Jiabin, Lu & K. L. Du (2011). Fuzzy Logic and Neuro-Fuzzy System; A systematic introduction. *International Journal of Artificial Intelligence and Expert Systems (IJAE)*, 2(2) 47 – 80
- Zhang G, Yan P, Zhao H, & Zhang X (2008). A Computer Aided Diagnosis System in Mammography Using Artificial Neural Networks. *International Conference on BioMedical Engineering and Informatics, Sanya, China*. 2, 823–826.
- Zimmermann, H. J. (1996), *Fuzzy set Theory and its Applications*, Kluwer Nijhoff Publishing, USA.

APPENDIX A

PIMA DIABETICS DATASET:

NATIONAL INSTITUTE OF DIABETES AND DIGESTIVE AND KIDNEY DISEASES

NTP	PGC	DBP	TSFT	2-HIS	BMI	DPF	AGE	PIMA
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31.0	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0.0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38.0	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30.0	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0
1	115	70	30	96	34.6	0.529	32	1
3	126	88	41	235	39.3	0.704	27	0
8	99	84	0	0	35.4	0.388	50	0
7	196	90	0	0	39.8	0.451	41	1
9	119	80	35	0	29.0	0.263	29	1
11	143	94	33	146	36.6	0.254	51	1
10	125	70	26	115	31.1	0.205	41	1
7	147	76	0	0	39.4	0.257	43	1
1	97	66	15	140	23.2	0.487	22	0
13	145	82	19	110	22.2	0.245	57	0
5	117	92	0	0	34.1	0.337	38	0
5	109	75	26	0	36.0	0.546	60	0
3	158	76	36	245	31.6	0.851	28	1
3	88	58	11	54	24.8	0.267	22	0
6	92	92	0	0	19.9	0.188	28	0
10	122	78	31	0	27.6	0.512	45	0
4	103	60	33	192	24.0	0.966	33	0
11	138	76	0	0	33.2	0.420	35	0

9	102	76	37	0	32.9	0.665	46	1
2	90	68	42	0	38.2	0.503	27	1
4	111	72	47	207	37.1	1.390	56	1
3	180	64	25	70	34.0	0.271	26	0
7	133	84	0	0	40.2	0.696	37	0
7	106	92	18	0	22.7	0.235	48	0
9	171	110	24	240	45.4	0.721	54	1
7	159	64	0	0	27.4	0.294	40	0
0	180	66	39	0	42.0	1.893	25	1
1	146	56	0	0	29.7	0.564	29	0
2	71	70	27	0	28.0	0.586	22	0
7	103	66	32	0	39.1	0.344	31	1
7	105	0	0	0	0.0	0.305	24	0
1	103	80	11	82	19.4	0.491	22	0
1	101	50	15	36	24.2	0.526	26	0
5	88	66	21	23	24.4	0.342	30	0
8	176	90	34	300	33.7	0.467	58	1
7	150	66	42	342	34.7	0.718	42	0
1	73	50	10	0	23.0	0.248	21	0
7	187	68	39	304	37.7	0.254	41	1
0	100	88	60	110	46.8	0.962	31	0
0	146	82	0	0	40.5	1.781	44	0
0	105	64	41	142	41.5	0.173	22	0
2	84	0	0	0	0.0	0.304	21	0
8	133	72	0	0	32.9	0.270	39	1
5	44	62	0	0	25.0	0.587	36	0
2	141	58	34	128	25.4	0.699	24	0
7	114	66	0	0	32.8	0.258	42	1
5	99	74	27	0	29.0	0.203	32	0
0	109	88	30	0	32.5	0.855	38	1
2	109	92	0	0	42.7	0.845	54	0
1	95	66	13	38	19.6	0.334	25	0
4	146	85	27	100	28.9	0.189	27	0
2	100	66	20	90	32.9	0.867	28	1
5	139	64	35	140	28.6	0.411	26	0
13	126	90	0	0	43.4	0.583	42	1
4	129	86	20	270	35.1	0.231	23	0
1	79	75	30	0	32.0	0.396	22	0
1	0	48	20	0	24.7	0.140	22	0
7	62	78	0	0	32.6	0.391	41	0
5	95	72	33	0	37.7	0.370	27	0
0	131	0	0	0	43.2	0.270	26	1
2	112	66	22	0	25.0	0.307	24	0
3	113	44	13	0	22.4	0.140	22	0
2	74	0	0	0	0.0	0.102	22	0

7	83	78	26	71	29.3	0.767	36	0
0	101	65	28	0	24.6	0.237	22	0
5	137	108	0	0	48.8	0.227	37	1
2	110	74	29	125	32.4	0.698	27	0
13	106	72	54	0	36.6	0.178	45	0
2	100	68	25	71	38.5	0.324	26	0
15	136	70	32	110	37.1	0.153	43	1
1	107	68	19	0	26.5	0.165	24	0
1	80	55	0	0	19.1	0.258	21	0
4	123	80	15	176	32.0	0.443	34	0
7	81	78	40	48	46.7	0.261	42	0
4	134	72	0	0	23.8	0.277	60	1
2	142	82	18	64	24.7	0.761	21	0
6	144	72	27	228	33.9	0.255	40	0
2	92	62	28	0	31.6	0.130	24	0
1	71	48	18	76	20.4	0.323	22	0
6	93	50	30	64	28.7	0.356	23	0
1	122	90	51	220	49.7	0.325	31	1
1	163	72	0	0	39.0	1.222	33	1
1	151	60	0	0	26.1	0.179	22	0
0	125	96	0	0	22.5	0.262	21	0
1	81	72	18	40	26.6	0.283	24	0
2	85	65	0	0	39.6	0.930	27	0
1	126	56	29	152	28.7	0.801	21	0
1	96	122	0	0	22.4	0.207	27	0
4	144	58	28	140	29.5	0.287	37	0
3	83	58	31	18	34.3	0.336	25	0
0	95	85	25	36	37.4	0.247	24	1
3	171	72	33	135	33.3	0.199	24	1
8	155	62	26	495	34.0	0.543	46	1
1	89	76	34	37	31.2	0.192	23	0
4	76	62	0	0	34.0	0.391	25	0
7	160	54	32	175	30.5	0.588	39	1
4	146	92	0	0	31.2	0.539	61	1
5	124	74	0	0	34.0	0.220	38	1
5	78	48	0	0	33.7	0.654	25	0
4	97	60	23	0	28.2	0.443	22	0
4	99	76	15	51	23.2	0.223	21	0
0	162	76	56	100	53.2	0.759	25	1
6	111	64	39	0	34.2	0.260	24	0
2	107	74	30	100	33.6	0.404	23	0
5	132	80	0	0	26.8	0.186	69	0
0	113	76	0	0	33.3	0.278	23	1
1	88	30	42	99	55.0	0.496	26	1
3	120	70	30	135	42.9	0.452	30	0

1	118	58	36	94	33.3	0.261	23	0
1	117	88	24	145	34.5	0.403	40	1
0	105	84	0	0	27.9	0.741	62	1
4	173	70	14	168	29.7	0.361	33	1
9	122	56	0	0	33.3	1.114	33	1
3	170	64	37	225	34.5	0.356	30	1
8	84	74	31	0	38.3	0.457	39	0
2	96	68	13	49	21.1	0.647	26	0
2	125	60	20	140	33.8	0.088	31	0
0	100	70	26	50	30.8	0.597	21	0
0	93	60	25	92	28.7	0.532	22	0
0	129	80	0	0	31.2	0.703	29	0
5	105	72	29	325	36.9	0.159	28	0
3	128	78	0	0	21.1	0.268	55	0
5	106	82	30	0	39.5	0.286	38	0
2	108	52	26	63	32.5	0.318	22	0
10	108	66	0	0	32.4	0.272	42	1
4	154	62	31	284	32.8	0.237	23	0
0	102	75	23	0	0.0	0.572	21	0
9	57	80	37	0	32.8	0.096	41	0
2	106	64	35	119	30.5	1.400	34	0
5	147	78	0	0	33.7	0.218	65	0
2	90	70	17	0	27.3	0.085	22	0
1	136	74	50	204	37.4	0.399	24	0
4	114	65	0	0	21.9	0.432	37	0
9	156	86	28	155	34.3	1.189	42	1
1	153	82	42	485	40.6	0.687	23	0
8	188	78	0	0	47.9	0.137	43	1
7	152	88	44	0	50.0	0.337	36	1
2	99	52	15	94	24.6	0.637	21	0
1	109	56	21	135	25.2	0.833	23	0
2	88	74	19	53	29.0	0.229	22	0
17	163	72	41	114	40.9	0.817	47	1
4	151	90	38	0	29.7	0.294	36	0
7	102	74	40	105	37.2	0.204	45	0
0	114	80	34	285	44.2	0.167	27	0
2	100	64	23	0	29.7	0.368	21	0
0	131	88	0	0	31.6	0.743	32	1
6	104	74	18	156	29.9	0.722	41	1
3	148	66	25	0	32.5	0.256	22	0
4	120	68	0	0	29.6	0.709	34	0
4	110	66	0	0	31.9	0.471	29	0
3	111	90	12	78	28.4	0.495	29	0
6	102	82	0	0	30.8	0.180	36	1
6	134	70	23	130	35.4	0.542	29	1

2	87	0	23	0	28.9	0.773	25	0
1	79	60	42	48	43.5	0.678	23	0
2	75	64	24	55	29.7	0.370	33	0
8	179	72	42	130	32.7	0.719	36	1
6	85	78	0	0	31.2	0.382	42	0
0	129	110	46	130	67.1	0.319	26	1
5	143	78	0	0	45.0	0.190	47	0
5	130	82	0	0	39.1	0.956	37	1
6	87	80	0	0	23.2	0.084	32	0
0	119	64	18	92	34.9	0.725	23	0
1	0	74	20	23	27.7	0.299	21	0
5	73	60	0	0	26.8	0.268	27	0
4	141	74	0	0	27.6	0.244	40	0
7	194	68	28	0	35.9	0.745	41	1
8	181	68	36	495	30.1	0.615	60	1
1	128	98	41	58	32.0	1.321	33	1
8	109	76	39	114	27.9	0.640	31	1
5	139	80	35	160	31.6	0.361	25	1
3	111	62	0	0	22.6	0.142	21	0
9	123	70	44	94	33.1	0.374	40	0
7	159	66	0	0	30.4	0.383	36	1
11	135	0	0	0	52.3	0.578	40	1
8	85	55	20	0	24.4	0.136	42	0
5	158	84	41	210	39.4	0.395	29	1
1	105	58	0	0	24.3	0.187	21	0
3	107	62	13	48	22.9	0.678	23	1
4	109	64	44	99	34.8	0.905	26	1
4	148	60	27	318	30.9	0.150	29	1
0	113	80	16	0	31.0	0.874	21	0
1	138	82	0	0	40.1	0.236	28	0
0	108	68	20	0	27.3	0.787	32	0
2	99	70	16	44	20.4	0.235	27	0
6	103	72	32	190	37.7	0.324	55	0
5	111	72	28	0	23.9	0.407	27	0
8	196	76	29	280	37.5	0.605	57	1
5	162	104	0	0	37.7	0.151	52	1
1	96	64	27	87	33.2	0.289	21	0
7	184	84	33	0	35.5	0.355	41	1
2	81	60	22	0	27.7	0.290	25	0
0	147	85	54	0	42.8	0.375	24	0
7	179	95	31	0	34.2	0.164	60	0
0	140	65	26	130	42.6	0.431	24	1
9	112	82	32	175	34.2	0.260	36	1
12	151	70	40	271	41.8	0.742	38	1
5	109	62	41	129	35.8	0.514	25	1

6	125	68	30	120	30.0	0.464	32	0
5	85	74	22	0	29.0	1.224	32	1
5	112	66	0	0	37.8	0.261	41	1
0	177	60	29	478	34.6	1.072	21	1
2	158	90	0	0	31.6	0.805	66	1
7	119	0	0	0	25.2	0.209	37	0
7	142	60	33	190	28.8	0.687	61	0
1	100	66	15	56	23.6	0.666	26	0
1	87	78	27	32	34.6	0.101	22	0
0	101	76	0	0	35.7	0.198	26	0
3	162	52	38	0	37.2	0.652	24	1
4	197	70	39	744	36.7	2.329	31	0
0	117	80	31	53	45.2	0.089	24	0
4	142	86	0	0	44.0	0.645	22	1
6	134	80	37	370	46.2	0.238	46	1
1	79	80	25	37	25.4	0.583	22	0
4	122	68	0	0	35.0	0.394	29	0
3	74	68	28	45	29.7	0.293	23	0
4	171	72	0	0	43.6	0.479	26	1
7	181	84	21	192	35.9	0.586	51	1
0	179	90	27	0	44.1	0.686	23	1
9	164	84	21	0	30.8	0.831	32	1
0	104	76	0	0	18.4	0.582	27	0
1	91	64	24	0	29.2	0.192	21	0
4	91	70	32	88	33.1	0.446	22	0
3	139	54	0	0	25.6	0.402	22	1
6	119	50	22	176	27.1	1.318	33	1
2	146	76	35	194	38.2	0.329	29	0
9	184	85	15	0	30.0	1.213	49	1
10	122	68	0	0	31.2	0.258	41	0
0	165	90	33	680	52.3	0.427	23	0
9	124	70	33	402	35.4	0.282	34	0
1	111	86	19	0	30.1	0.143	23	0
9	106	52	0	0	31.2	0.380	42	0
2	129	84	0	0	28.0	0.284	27	0
2	90	80	14	55	24.4	0.249	24	0
0	86	68	32	0	35.8	0.238	25	0
12	92	62	7	258	27.6	0.926	44	1
1	113	64	35	0	33.6	0.543	21	1
3	111	56	39	0	30.1	0.557	30	0
2	114	68	22	0	28.7	0.092	25	0
1	193	50	16	375	25.9	0.655	24	0
11	155	76	28	150	33.3	1.353	51	1
3	191	68	15	130	30.9	0.299	34	0
3	141	0	0	0	30.0	0.761	27	1

4	95	70	32	0	32.1	0.612	24	0
3	142	80	15	0	32.4	0.200	63	0
4	123	62	0	0	32.0	0.226	35	1
5	96	74	18	67	33.6	0.997	43	0
0	138	0	0	0	36.3	0.933	25	1
2	128	64	42	0	40.0	1.101	24	0
0	102	52	0	0	25.1	0.078	21	0
2	146	0	0	0	27.5	0.240	28	1
10	101	86	37	0	45.6	1.136	38	1
2	108	62	32	56	25.2	0.128	21	0
3	122	78	0	0	23.0	0.254	40	0
1	71	78	50	45	33.2	0.422	21	0
13	106	70	0	0	34.2	0.251	52	0
2	100	70	52	57	40.5	0.677	25	0
7	106	60	24	0	26.5	0.296	29	1
0	104	64	23	116	27.8	0.454	23	0
5	114	74	0	0	24.9	0.744	57	0
2	108	62	10	278	25.3	0.881	22	0
0	146	70	0	0	37.9	0.334	28	1
10	129	76	28	122	35.9	0.280	39	0
7	133	88	15	155	32.4	0.262	37	0
7	161	86	0	0	30.4	0.165	47	1
2	108	80	0	0	27.0	0.259	52	1
7	136	74	26	135	26.0	0.647	51	0
5	155	84	44	545	38.7	0.619	34	0
1	119	86	39	220	45.6	0.808	29	1
4	96	56	17	49	20.8	0.340	26	0
5	108	72	43	75	36.1	0.263	33	0
0	78	88	29	40	36.9	0.434	21	0
0	107	62	30	74	36.6	0.757	25	1
2	128	78	37	182	43.3	1.224	31	1
1	128	48	45	194	40.5	0.613	24	1
0	161	50	0	0	21.9	0.254	65	0
6	151	62	31	120	35.5	0.692	28	0
2	146	70	38	360	28.0	0.337	29	1
0	126	84	29	215	30.7	0.520	24	0
14	100	78	25	184	36.6	0.412	46	1
8	112	72	0	0	23.6	0.840	58	0
0	167	0	0	0	32.3	0.839	30	1
2	144	58	33	135	31.6	0.422	25	1
5	77	82	41	42	35.8	0.156	35	0
5	115	98	0	0	52.9	0.209	28	1
3	150	76	0	0	21.0	0.207	37	0
2	120	76	37	105	39.7	0.215	29	0
10	161	68	23	132	25.5	0.326	47	1

0	137	68	14	148	24.8	0.143	21	0
0	128	68	19	180	30.5	1.391	25	1
2	124	68	28	205	32.9	0.875	30	1
6	80	66	30	0	26.2	0.313	41	0
0	106	70	37	148	39.4	0.605	22	0
2	155	74	17	96	26.6	0.433	27	1
3	113	50	10	85	29.5	0.626	25	0
7	109	80	31	0	35.9	1.127	43	1
2	112	68	22	94	34.1	0.315	26	0
3	99	80	11	64	19.3	0.284	30	0
3	182	74	0	0	30.5	0.345	29	1
3	115	66	39	140	38.1	0.150	28	0
6	194	78	0	0	23.5	0.129	59	1
4	129	60	12	231	27.5	0.527	31	0
3	112	74	30	0	31.6	0.197	25	1
0	124	70	20	0	27.4	0.254	36	1
13	152	90	33	29	26.8	0.731	43	1
2	112	75	32	0	35.7	0.148	21	0
1	157	72	21	168	25.6	0.123	24	0
1	122	64	32	156	35.1	0.692	30	1
10	179	70	0	0	35.1	0.200	37	0
2	102	86	36	120	45.5	0.127	23	1
6	105	70	32	68	30.8	0.122	37	0
8	118	72	19	0	23.1	1.476	46	0
2	87	58	16	52	32.7	0.166	25	0
1	180	0	0	0	43.3	0.282	41	1
12	106	80	0	0	23.6	0.137	44	0
1	95	60	18	58	23.9	0.260	22	0
0	165	76	43	255	47.9	0.259	26	0
0	117	0	0	0	33.8	0.932	44	0
5	115	76	0	0	31.2	0.343	44	1
9	152	78	34	171	34.2	0.893	33	1
7	178	84	0	0	39.9	0.331	41	1
1	130	70	13	105	25.9	0.472	22	0
1	95	74	21	73	25.9	0.673	36	0
1	0	68	35	0	32.0	0.389	22	0
5	122	86	0	0	34.7	0.290	33	0
8	95	72	0	0	36.8	0.485	57	0
8	126	88	36	108	38.5	0.349	49	0
1	139	46	19	83	28.7	0.654	22	0
3	116	0	0	0	23.5	0.187	23	0
3	99	62	19	74	21.8	0.279	26	0
5	0	80	32	0	41.0	0.346	37	1
4	92	80	0	0	42.2	0.237	29	0
4	137	84	0	0	31.2	0.252	30	0

3	61	82	28	0	34.4	0.243	46	0
1	90	62	12	43	27.2	0.580	24	0
3	90	78	0	0	42.7	0.559	21	0
9	165	88	0	0	30.4	0.302	49	1
1	125	50	40	167	33.3	0.962	28	1
13	129	0	30	0	39.9	0.569	44	1
12	88	74	40	54	35.3	0.378	48	0
1	196	76	36	249	36.5	0.875	29	1
5	189	64	33	325	31.2	0.583	29	1
5	158	70	0	0	29.8	0.207	63	0
5	103	108	37	0	39.2	0.305	65	0
4	146	78	0	0	38.5	0.520	67	1
4	147	74	25	293	34.9	0.385	30	0
5	99	54	28	83	34.0	0.499	30	0
6	124	72	0	0	27.6	0.368	29	1
0	101	64	17	0	21.0	0.252	21	0
3	81	86	16	66	27.5	0.306	22	0
1	133	102	28	140	32.8	0.234	45	1
3	173	82	48	465	38.4	2.137	25	1
0	118	64	23	89	0.0	1.731	21	0
0	84	64	22	66	35.8	0.545	21	0
2	105	58	40	94	34.9	0.225	25	0
2	122	52	43	158	36.2	0.816	28	0
12	140	82	43	325	39.2	0.528	58	1
0	98	82	15	84	25.2	0.299	22	0
1	87	60	37	75	37.2	0.509	22	0
4	156	75	0	0	48.3	0.238	32	1
0	93	100	39	72	43.4	1.021	35	0
1	107	72	30	82	30.8	0.821	24	0
0	105	68	22	0	20.0	0.236	22	0
1	109	60	8	182	25.4	0.947	21	0
1	90	62	18	59	25.1	1.268	25	0
1	125	70	24	110	24.3	0.221	25	0
1	119	54	13	50	22.3	0.205	24	0
5	116	74	29	0	32.3	0.660	35	1
8	105	100	36	0	43.3	0.239	45	1
5	144	82	26	285	32.0	0.452	58	1
3	100	68	23	81	31.6	0.949	28	0
1	100	66	29	196	32.0	0.444	42	0
5	166	76	0	0	45.7	0.340	27	1
1	131	64	14	415	23.7	0.389	21	0
4	116	72	12	87	22.1	0.463	37	0
4	158	78	0	0	32.9	0.803	31	1
2	127	58	24	275	27.7	1.600	25	0
3	96	56	34	115	24.7	0.944	39	0

0	131	66	40	0	34.3	0.196	22	1
3	82	70	0	0	21.1	0.389	25	0
3	193	70	31	0	34.9	0.241	25	1
4	95	64	0	0	32.0	0.161	31	1
6	137	61	0	0	24.2	0.151	55	0
5	136	84	41	88	35.0	0.286	35	1
9	72	78	25	0	31.6	0.280	38	0
5	168	64	0	0	32.9	0.135	41	1
2	123	48	32	165	42.1	0.520	26	0
4	115	72	0	0	28.9	0.376	46	1
0	101	62	0	0	21.9	0.336	25	0
8	197	74	0	0	25.9	1.191	39	1
1	172	68	49	579	42.4	0.702	28	1
6	102	90	39	0	35.7	0.674	28	0
1	112	72	30	176	34.4	0.528	25	0
1	143	84	23	310	42.4	1.076	22	0
1	143	74	22	61	26.2	0.256	21	0
0	138	60	35	167	34.6	0.534	21	1
3	173	84	33	474	35.7	0.258	22	1
1	97	68	21	0	27.2	1.095	22	0
4	144	82	32	0	38.5	0.554	37	1
1	83	68	0	0	18.2	0.624	27	0
3	129	64	29	115	26.4	0.219	28	1
1	119	88	41	170	45.3	0.507	26	0
2	94	68	18	76	26.0	0.561	21	0
0	102	64	46	78	40.6	0.496	21	0
2	115	64	22	0	30.8	0.421	21	0
8	151	78	32	210	42.9	0.516	36	1
4	184	78	39	277	37.0	0.264	31	1
0	94	0	0	0	0.0	0.256	25	0
1	181	64	30	180	34.1	0.328	38	1
0	135	94	46	145	40.6	0.284	26	0
1	95	82	25	180	35.0	0.233	43	1
2	99	0	0	0	22.2	0.108	23	0
3	89	74	16	85	30.4	0.551	38	0
1	80	74	11	60	30.0	0.527	22	0
2	139	75	0	0	25.6	0.167	29	0
1	90	68	8	0	24.5	1.138	36	0
0	141	0	0	0	42.4	0.205	29	1
12	140	85	33	0	37.4	0.244	41	0
5	147	75	0	0	29.9	0.434	28	0
1	97	70	15	0	18.2	0.147	21	0
6	107	88	0	0	36.8	0.727	31	0
0	189	104	25	0	34.3	0.435	41	1
2	83	66	23	50	32.2	0.497	22	0

4	117	64	27	120	33.2	0.230	24	0
8	108	70	0	0	30.5	0.955	33	1
4	117	62	12	0	29.7	0.380	30	1
0	180	78	63	14	59.4	2.420	25	1
1	100	72	12	70	25.3	0.658	28	0
0	95	80	45	92	36.5	0.330	26	0
0	104	64	37	64	33.6	0.510	22	1
0	120	74	18	63	30.5	0.285	26	0
1	82	64	13	95	21.2	0.415	23	0
2	134	70	0	0	28.9	0.542	23	1
0	91	68	32	210	39.9	0.381	25	0
2	119	0	0	0	19.6	0.832	72	0
2	100	54	28	105	37.8	0.498	24	0
14	175	62	30	0	33.6	0.212	38	1
1	135	54	0	0	26.7	0.687	62	0
5	86	68	28	71	30.2	0.364	24	0
10	148	84	48	237	37.6	1.001	51	1
9	134	74	33	60	25.9	0.460	81	0
9	120	72	22	56	20.8	0.733	48	0
1	71	62	0	0	21.8	0.416	26	0
8	74	70	40	49	35.3	0.705	39	0
5	88	78	30	0	27.6	0.258	37	0
10	115	98	0	0	24.0	1.022	34	0
0	124	56	13	105	21.8	0.452	21	0
0	74	52	10	36	27.8	0.269	22	0
0	97	64	36	100	36.8	0.600	25	0
8	120	0	0	0	30.0	0.183	38	1
6	154	78	41	140	46.1	0.571	27	0
1	144	82	40	0	41.3	0.607	28	0
0	137	70	38	0	33.2	0.170	22	0
0	119	66	27	0	38.8	0.259	22	0
7	136	90	0	0	29.9	0.210	50	0
4	114	64	0	0	28.9	0.126	24	0
0	137	84	27	0	27.3	0.231	59	0
2	105	80	45	191	33.7	0.711	29	1
7	114	76	17	110	23.8	0.466	31	0
8	126	74	38	75	25.9	0.162	39	0
4	132	86	31	0	28.0	0.419	63	0
3	158	70	30	328	35.5	0.344	35	1
0	123	88	37	0	35.2	0.197	29	0
4	85	58	22	49	27.8	0.306	28	0
0	84	82	31	125	38.2	0.233	23	0
0	145	0	0	0	44.2	0.630	31	1
0	135	68	42	250	42.3	0.365	24	1
1	139	62	41	480	40.7	0.536	21	0

0	173	78	32	265	46.5	1.159	58	0
4	99	72	17	0	25.6	0.294	28	0
8	194	80	0	0	26.1	0.551	67	0
2	83	65	28	66	36.8	0.629	24	0
2	89	90	30	0	33.5	0.292	42	0
4	99	68	38	0	32.8	0.145	33	0
4	125	70	18	122	28.9	1.144	45	1
3	80	0	0	0	0.0	0.174	22	0
6	166	74	0	0	26.6	0.304	66	0
5	110	68	0	0	26.0	0.292	30	0
2	81	72	15	76	30.1	0.547	25	0
7	195	70	33	145	25.1	0.163	55	1
6	154	74	32	193	29.3	0.839	39	0
2	117	90	19	71	25.2	0.313	21	0
3	84	72	32	0	37.2	0.267	28	0
6	0	68	41	0	39.0	0.727	41	1
7	94	64	25	79	33.3	0.738	41	0
3	96	78	39	0	37.3	0.238	40	0
10	75	82	0	0	33.3	0.263	38	0
0	180	90	26	90	36.5	0.314	35	1
1	130	60	23	170	28.6	0.692	21	0
2	84	50	23	76	30.4	0.968	21	0
8	120	78	0	0	25.0	0.409	64	0
12	84	72	31	0	29.7	0.297	46	1
0	139	62	17	210	22.1	0.207	21	0
9	91	68	0	0	24.2	0.200	58	0
2	91	62	0	0	27.3	0.525	22	0
3	99	54	19	86	25.6	0.154	24	0
3	163	70	18	105	31.6	0.268	28	1
9	145	88	34	165	30.3	0.771	53	1
7	125	86	0	0	37.6	0.304	51	0
13	76	60	0	0	32.8	0.180	41	0
6	129	90	7	326	19.6	0.582	60	0
2	68	70	32	66	25.0	0.187	25	0
3	124	80	33	130	33.2	0.305	26	0
6	114	0	0	0	0.0	0.189	26	0
9	130	70	0	0	34.2	0.652	45	1
3	125	58	0	0	31.6	0.151	24	0
3	87	60	18	0	21.8	0.444	21	0
1	97	64	19	82	18.2	0.299	21	0
3	116	74	15	105	26.3	0.107	24	0
0	117	66	31	188	30.8	0.493	22	0
0	111	65	0	0	24.6	0.660	31	0
2	122	60	18	106	29.8	0.717	22	0
0	107	76	0	0	45.3	0.686	24	0

1	86	66	52	65	41.3	0.917	29	0
6	91	0	0	0	29.8	0.501	31	0
1	77	56	30	56	33.3	1.251	24	0
4	132	0	0	0	32.9	0.302	23	1
0	105	90	0	0	29.6	0.197	46	0
0	57	60	0	0	21.7	0.735	67	0
0	127	80	37	210	36.3	0.804	23	0
3	129	92	49	155	36.4	0.968	32	1
8	100	74	40	215	39.4	0.661	43	1
3	128	72	25	190	32.4	0.549	27	1
10	90	85	32	0	34.9	0.825	56	1
4	84	90	23	56	39.5	0.159	25	0
1	88	78	29	76	32.0	0.365	29	0
8	186	90	35	225	34.5	0.423	37	1
5	187	76	27	207	43.6	1.034	53	1
4	131	68	21	166	33.1	0.160	28	0
1	164	82	43	67	32.8	0.341	50	0
4	189	110	31	0	28.5	0.680	37	0
1	116	70	28	0	27.4	0.204	21	0
3	84	68	30	106	31.9	0.591	25	0
6	114	88	0	0	27.8	0.247	66	0
1	88	62	24	44	29.9	0.422	23	0
1	84	64	23	115	36.9	0.471	28	0
7	124	70	33	215	25.5	0.161	37	0
1	97	70	40	0	38.1	0.218	30	0
8	110	76	0	0	27.8	0.237	58	0
11	103	68	40	0	46.2	0.126	42	0
11	85	74	0	0	30.1	0.300	35	0
6	125	76	0	0	33.8	0.121	54	1
0	198	66	32	274	41.3	0.502	28	1
1	87	68	34	77	37.6	0.401	24	0
6	99	60	19	54	26.9	0.497	32	0
0	91	80	0	0	32.4	0.601	27	0
2	95	54	14	88	26.1	0.748	22	0
1	99	72	30	18	38.6	0.412	21	0
6	92	62	32	126	32.0	0.085	46	0
4	154	72	29	126	31.3	0.338	37	0
0	121	66	30	165	34.3	0.203	33	1
3	78	70	0	0	32.5	0.270	39	0
2	130	96	0	0	22.6	0.268	21	0
3	111	58	31	44	29.5	0.430	22	0
2	98	60	17	120	34.7	0.198	22	0
1	143	86	30	330	30.1	0.892	23	0
1	119	44	47	63	35.5	0.280	25	0
6	108	44	20	130	24.0	0.813	35	0

2	118	80	0	0	42.9	0.693	21	1
10	133	68	0	0	27.0	0.245	36	0
2	197	70	99	0	34.7	0.575	62	1
0	151	90	46	0	42.1	0.371	21	1
6	109	60	27	0	25.0	0.206	27	0
12	121	78	17	0	26.5	0.259	62	0
8	100	76	0	0	38.7	0.190	42	0
8	124	76	24	600	28.7	0.687	52	1
1	93	56	11	0	22.5	0.417	22	0
8	143	66	0	0	34.9	0.129	41	1
6	103	66	0	0	24.3	0.249	29	0
3	176	86	27	156	33.3	1.154	52	1
0	73	0	0	0	21.1	0.342	25	0
11	111	84	40	0	46.8	0.925	45	1
2	112	78	50	140	39.4	0.175	24	0
3	132	80	0	0	34.4	0.402	44	1
2	82	52	22	115	28.5	1.699	25	0
6	123	72	45	230	33.6	0.733	34	0
0	188	82	14	185	32.0	0.682	22	1
0	67	76	0	0	45.3	0.194	46	0
1	89	24	19	25	27.8	0.559	21	0
1	173	74	0	0	36.8	0.088	38	1
1	109	38	18	120	23.1	0.407	26	0
1	108	88	19	0	27.1	0.400	24	0
6	96	0	0	0	23.7	0.190	28	0
1	124	74	36	0	27.8	0.100	30	0
7	150	78	29	126	35.2	0.692	54	1
4	183	0	0	0	28.4	0.212	36	1
1	124	60	32	0	35.8	0.514	21	0
1	181	78	42	293	40.0	1.258	22	1
1	92	62	25	41	19.5	0.482	25	0
0	152	82	39	272	41.5	0.270	27	0
1	111	62	13	182	24.0	0.138	23	0
3	106	54	21	158	30.9	0.292	24	0
3	174	58	22	194	32.9	0.593	36	1
7	168	88	42	321	38.2	0.787	40	1
6	105	80	28	0	32.5	0.878	26	0
11	138	74	26	144	36.1	0.557	50	1
3	106	72	0	0	25.8	0.207	27	0
6	117	96	0	0	28.7	0.157	30	0
2	68	62	13	15	20.1	0.257	23	0
9	112	82	24	0	28.2	1.282	50	1
0	119	0	0	0	32.4	0.141	24	1
2	112	86	42	160	38.4	0.246	28	0
2	92	76	20	0	24.2	1.698	28	0

6	183	94	0	0	40.8	1.461	45	0
0	94	70	27	115	43.5	0.347	21	0
2	108	64	0	0	30.8	0.158	21	0
4	90	88	47	54	37.7	0.362	29	0
0	125	68	0	0	24.7	0.206	21	0
0	132	78	0	0	32.4	0.393	21	0
5	128	80	0	0	34.6	0.144	45	0
4	94	65	22	0	24.7	0.148	21	0
7	114	64	0	0	27.4	0.732	34	1
0	102	78	40	90	34.5	0.238	24	0
2	111	60	0	0	26.2	0.343	23	0
1	128	82	17	183	27.5	0.115	22	0
10	92	62	0	0	25.9	0.167	31	0
13	104	72	0	0	31.2	0.465	38	1
5	104	74	0	0	28.8	0.153	48	0
2	94	76	18	66	31.6	0.649	23	0
7	97	76	32	91	40.9	0.871	32	1
1	100	74	12	46	19.5	0.149	28	0
0	102	86	17	105	29.3	0.695	27	0
4	128	70	0	0	34.3	0.303	24	0
6	147	80	0	0	29.5	0.178	50	1
4	90	0	0	0	28.0	0.610	31	0
3	103	72	30	152	27.6	0.730	27	0
2	157	74	35	440	39.4	0.134	30	0
1	167	74	17	144	23.4	0.447	33	1
0	179	50	36	159	37.8	0.455	22	1
11	136	84	35	130	28.3	0.260	42	1
0	107	60	25	0	26.4	0.133	23	0
1	91	54	25	100	25.2	0.234	23	0
1	117	60	23	106	33.8	0.466	27	0
5	123	74	40	77	34.1	0.269	28	0
2	120	54	0	0	26.8	0.455	27	0
1	106	70	28	135	34.2	0.142	22	0
2	155	52	27	540	38.7	0.240	25	1
2	101	58	35	90	21.8	0.155	22	0
1	120	80	48	200	38.9	1.162	41	0
11	127	106	0	0	39.0	0.190	51	0
3	80	82	31	70	34.2	1.292	27	1
10	162	84	0	0	27.7	0.182	54	0
1	199	76	43	0	42.9	1.394	22	1
8	167	106	46	231	37.6	0.165	43	1
9	145	80	46	130	37.9	0.637	40	1
6	115	60	39	0	33.7	0.245	40	1
1	112	80	45	132	34.8	0.217	24	0
4	145	82	18	0	32.5	0.235	70	1

10	111	70	27	0	27.5	0.141	40	1
6	98	58	33	190	34.0	0.430	43	0
9	154	78	30	100	30.9	0.164	45	0
6	165	68	26	168	33.6	0.631	49	0
1	99	58	10	0	25.4	0.551	21	0
10	68	106	23	49	35.5	0.285	47	0
3	123	100	35	240	57.3	0.880	22	0
8	91	82	0	0	35.6	0.587	68	0
6	195	70	0	0	30.9	0.328	31	1
9	156	86	0	0	24.8	0.230	53	1
0	93	60	0	0	35.3	0.263	25	0
3	121	52	0	0	36.0	0.127	25	1
2	101	58	17	265	24.2	0.614	23	0
2	56	56	28	45	24.2	0.332	22	0
0	162	76	36	0	49.6	0.364	26	1
0	95	64	39	105	44.6	0.366	22	0
4	125	80	0	0	32.3	0.536	27	1
5	136	82	0	0	0.0	0.640	69	0
2	129	74	26	205	33.2	0.591	25	0
3	130	64	0	0	23.1	0.314	22	0
1	107	50	19	0	28.3	0.181	29	0
1	140	74	26	180	24.1	0.828	23	0
1	144	82	46	180	46.1	0.335	46	1
8	107	80	0	0	24.6	0.856	34	0
13	158	114	0	0	42.3	0.257	44	1
2	121	70	32	95	39.1	0.886	23	0
7	129	68	49	125	38.5	0.439	43	1
2	90	60	0	0	23.5	0.191	25	0
7	142	90	24	480	30.4	0.128	43	1
3	169	74	19	125	29.9	0.268	31	1
0	99	0	0	0	25.0	0.253	22	0
4	127	88	11	155	34.5	0.598	28	0
4	118	70	0	0	44.5	0.904	26	0
2	122	76	27	200	35.9	0.483	26	0
6	125	78	31	0	27.6	0.565	49	1
1	168	88	29	0	35.0	0.905	52	1
2	129	0	0	0	38.5	0.304	41	0
4	110	76	20	100	28.4	0.118	27	0
6	80	80	36	0	39.8	0.177	28	0
10	115	0	0	0	0.0	0.261	30	1
2	127	46	21	335	34.4	0.176	22	0
9	164	78	0	0	32.8	0.148	45	1
2	93	64	32	160	38.0	0.674	23	1
3	158	64	13	387	31.2	0.295	24	0
5	126	78	27	22	29.6	0.439	40	0

10	129	62	36	0	41.2	0.441	38	1
0	134	58	20	291	26.4	0.352	21	0
3	102	74	0	0	29.5	0.121	32	0
7	187	50	33	392	33.9	0.826	34	1
3	173	78	39	185	33.8	0.970	31	1
10	94	72	18	0	23.1	0.595	56	0
1	108	60	46	178	35.5	0.415	24	0
5	97	76	27	0	35.6	0.378	52	1
4	83	86	19	0	29.3	0.317	34	0
1	114	66	36	200	38.1	0.289	21	0
1	149	68	29	127	29.3	0.349	42	1
5	117	86	30	105	39.1	0.251	42	0
1	111	94	0	0	32.8	0.265	45	0
4	112	78	40	0	39.4	0.236	38	0
1	116	78	29	180	36.1	0.496	25	0
0	141	84	26	0	32.4	0.433	22	0
2	175	88	0	0	22.9	0.326	22	0
2	92	52	0	0	30.1	0.141	22	0
3	130	78	23	79	28.4	0.323	34	1
8	120	86	0	0	28.4	0.259	22	1
2	174	88	37	120	44.5	0.646	24	1
2	106	56	27	165	29.0	0.426	22	0
2	105	75	0	0	23.3	0.560	53	0
4	95	60	32	0	35.4	0.284	28	0
0	126	86	27	120	27.4	0.515	21	0
8	65	72	23	0	32.0	0.600	42	0
2	99	60	17	160	36.6	0.453	21	0
1	102	74	0	0	39.5	0.293	42	1
11	120	80	37	150	42.3	0.785	48	1
3	102	44	20	94	30.8	0.400	26	0
1	109	58	18	116	28.5	0.219	22	0
9	140	94	0	0	32.7	0.734	45	1
13	153	88	37	140	40.6	1.174	39	0
12	100	84	33	105	30.0	0.488	46	0
1	147	94	41	0	49.3	0.358	27	1
1	81	74	41	57	46.3	1.096	32	0
3	187	70	22	200	36.4	0.408	36	1
6	162	62	0	0	24.3	0.178	50	1
4	136	70	0	0	31.2	1.182	22	1
1	121	78	39	74	39.0	0.261	28	0
3	108	62	24	0	26.0	0.223	25	0
0	181	88	44	510	43.3	0.222	26	1
8	154	78	32	0	32.4	0.443	45	1
1	128	88	39	110	36.5	1.057	37	1
7	137	90	41	0	32.0	0.391	39	0

0	123	72	0	0	36.3	0.258	52	1
1	106	76	0	0	37.5	0.197	26	0
6	190	92	0	0	35.5	0.278	66	1
2	88	58	26	16	28.4	0.766	22	0
9	170	74	31	0	44.0	0.403	43	1
9	89	62	0	0	22.5	0.142	33	0
10	101	76	48	180	32.9	0.171	63	0
2	122	70	27	0	36.8	0.340	27	0
5	121	72	23	112	26.2	0.245	30	0
1	126	60	0	0	30.1	0.349	47	1
1	93	70	31	0	30.4	0.315	23	0

APPENDIX B

SOURCE CODE

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package ga.neuro.fuzzy.diabetes.lib;

/**
 *
 * @author JMXCODE
 */
import com.mysql.jdbc.ReplicationDriver;
import java.io.File;
import java.io.FileInputStream;
import java.io.FileNotFoundException;
import java.io.FileOutputStream;
import java.io.InputStream;
import java.io.OutputStream;
import static java.lang.System.exit;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.PreparedStatement;
import java.sql.ResultSet;
import java.sql.SQLException;
import java.sql.Statement;
import java.util.Properties;
import java.util.Scanner;
import javax.swing.JOptionPane;

public class Database {

    private Connection conn;
    private Statement stmt;
    public ResultSet rs;
    public int num_of_rows;
    private ReplicationDriver driver;
    private Properties props;
    private String loginName, loginPass, dbname, server;

    public Database() throws SQLException {

        this.driver = new ReplicationDriver();
        this.props = new Properties();
        this.conn = null;
    }
}
```

```

this.stmt = null;
this.rs = null;
this.num_of_rows = 0;
this.loginName = "root";
this.loginPass = "";
this.dbname = "";
this.server = "";

Scanner input = null;
try {
    File db = new File(System.getProperty("user.dir") + "\\Data\\data.db");
    input = new Scanner(db);
    this.server = input.nextLine();
    this.dbname = input.nextLine();
    this.loginName = input.nextLine();
    this.loginPass = input.nextLine();
} catch (Exception ex) {
    JOptionPane.showMessageDialog(null, "Database Settings Error: "
        + "\n Unable to load Database settings" /*+ ex.toString()*/);
    exit(0);
}

if (input != null) {
    input.close();
}

props.put("autoReconnect", "true");
props.put("roundRobinLoadBalance", "true");
props.put("user", loginName.trim());
props.put("password", loginPass.trim());

}

public int connect() {
    int ret = 0;

    try {
        conn = DriverManager.getConnection("jdbc:mysql://" + this.server + "/" +
this.dbname, props);
        ret = 1;
    } catch (SQLException ex) {
        setupDB();
    }

    return ret;
}

```

```

public void setupDB() {
    String sql = "";
    try {
        conn = DriverManager.getConnection("jdbc:mysql://" + this.server, props);
        Scanner input = null;

        File db = new File(System.getProperty("user.dir") + "\\Data\\db.sql");
        input = new Scanner(db);
        while (input.hasNextLine()) {
            sql += input.nextLine();
        }

        if (input != null) {
            input.close();
        }

        String[] stmtss = sql.toString().split(";");
        for (int i = 0; i < stmtss.length; i++) {
            execScalar(stmtss[i]);
        }
        conn = DriverManager.getConnection("jdbc:mysql://" + this.server + "/" +
this.dbname, props);
    } catch (SQLException e) {
        JOptionPane.showMessageDialog(null, "Database Connection Error: \n" +
e.toString());
        exit(0);
    } catch (FileNotFoundException e) {
        JOptionPane.showMessageDialog(null, "Database Installation Error: \n" +
e.toString());
        exit(0);
    }
}

```

```

public int execScalar(String sql_stmt) {
    int ret = 0;

    if(this.conn == null){ return ret;}
    try {
        stmt = conn.createStatement(java.sql.ResultSet.TYPE_FORWARD_ONLY,
            java.sql.ResultSet.CONCUR_UPDATABLE);
        stmt.executeUpdate(sql_stmt);
        ret = 1;
    } catch (SQLException ex) {
        JOptionPane.showMessageDialog(null, "Database Scalar Error: " + ex.toString());
    } finally {
        if (stmt != null) {
            try {

```

```

        stmt.close();
    } catch (SQLException ex) {
    }
}
}
return ret;
}

public int execQuery(String sql_stmt) {
    int ret = 0;

    if(this.conn == null){ return ret;}
    try {
        stmt = conn.createStatement();
        rs = stmt.executeQuery(sql_stmt);

        if (stmt.execute(sql_stmt)) {
            rs = stmt.getResultSet();
            rs.last();
            num_of_rows = rs.getRow();
            rs.beforeFirst();
            ret = 1;
        }
    } catch (SQLException ex) {
        // handle any errors
        JOptionPane.showMessageDialog(null, "Database Query Error: " + ex.toString());
        /*System.out.println("SQLException: " + ex.getMessage());
        System.out.println("SQLState: " + ex.getSQLState());
        System.out.println("VendorError: " + ex.getErrorCode());*/
    }
    return ret;
}
}

```

```

public int execPrep(String path, String sql, String vals) {
    int ret = 1; int i;
    PreparedStatement pre;

    if(this.conn == null){ return ret;}
    try {
        pre = conn.prepareStatement(sql);
        String valss[] = vals.split("-");
        for (i = 0; i < valss.length; i++) {
            if (valss[i].equals("NULL")) {
                pre.setInt(i + 1, 0);
            } else {
                pre.setString(i + 1, valss[i]);
            }
        }
    }
}

```

```

        }
    }

    File file = new File(path);
    FileInputStream fs;
    try {
        fs = new FileInputStream(file);
        pre.setBinaryStream(i + 1, fs);
    } catch (FileNotFoundException ex) {
        System.out.println(ex.toString());
    }
    pre.executeUpdate();
    pre.close();

} catch (SQLException ex) {
    ret = 0;
    System.out.println(ex.toString());
}

return ret;
}

public int execGetBin(String path, String sql) {
    int ret = 0; int i = 0;

    if(this.conn == null){ return ret;}
    try {
        Statement stmt = conn.createStatement();
        rs = stmt.executeQuery(sql);
        while (rs.next()) {
            InputStream in = rs.getBinaryStream(1);
            OutputStream f = new FileOutputStream(new File(path));
            i++;
            int c = 0;
            while ((c = in.read()) > -1) {
                f.write(c);
            }
            f.close();
            in.close();
            ret = 1;
        }
    } catch (Exception ex) {
        ret = 0;
        System.out.println(ex.getMessage());
    }
    return ret;
}

```

```

public int closAll() {
    int ret = 1;

    if(this.conn == null){ return ret;}
    if (rs != null) {
        try {
            rs.close();
        } catch (SQLException sqlEx) {
            ret = 0;
        } // ignore
        rs = null;
    }
    if (stmt != null) {
        try {
            stmt.close();
        } catch (SQLException sqlEx) {
            ret = 0;
        } // ignore
        stmt = null;
    }
    return ret;
}
}
/*
 * To change this license header, choose License Headers in Project Properties.
 * To change this template file, choose Tools | Templates
 * and open the template in the editor.
 */
package ga.neuro.fuzzy.diabetes.lib;

import ga.neuro.fuzzy.diabetes.lib.fuzzy.*;
import ga.neuro.fuzzy.diabetes.lib.neuro.*;
import java.util.ArrayList;
import java.util.List;

/**
 *
 * @author Oracle1
 */
public class InferenceEngine {

    private NeuralNet net;
    private Defuzzifier def;
    private FuzzySystem fuzzySYS;;

    public InferenceEngine(){

```

```

try {
    this.def = new DiabetesDefuzzifier();
    this.fuzzySYS = new DiabetesFuzzySystem(def);
} catch (Exception ex) {
    System.err.println("Init error: "+ex.toString());
}

try {
    this.net = new NeuralNet(100, 8, 5, 5, 0.01);
    this.net.learn(1);
} catch (Exception ex) {
    System.err.println("Init error: "+ex.toString());
}
}

public double fuzzyInfer(double[] vals) {
    double ret = 0.0;

    for (int pos = 0; pos < vals.length; pos++) {
        if(pos == 0){
            vals[pos] = vals[pos] == 1.0?1.0: vals[pos]==2.0?3.0:
vals[pos]==3.0?5.0:vals[pos];
        }
        else if(pos == 1){
            vals[pos] = vals[pos] == 1.0?50.0: vals[pos]==2.0?100.0:
vals[pos]==3.0?150.0:vals[pos];
        }
        else if(pos == 2){
            vals[pos] = vals[pos] == 1.0?10.0: vals[pos]==2.0?40.0:
vals[pos]==3.0?70.0:vals[pos];
        }
        else if(pos == 3){
            vals[pos] = vals[pos] == 1.0?10.0: vals[pos]==2.0?40.0:
vals[pos]==3.0?70.0:vals[pos];
        }
        else if(pos == 4){
            vals[pos] = vals[pos] == 1.0?10.0: vals[pos]==2.0?70.0:
vals[pos]==3.0?120.0:vals[pos];
        }
        else if(pos == 5){
            vals[pos] = vals[pos] == 1.0?15.0: vals[pos]==2.0?35.0:
vals[pos]==3.0?50.0:vals[pos];
        }
        else if(pos == 6){
            vals[pos] = vals[pos] == 1.0?0.2: vals[pos]==2.0?0.5:
vals[pos]==3.0?0.7:vals[pos];
        }
    }
}

```

```

        else if(pos == 7){
            vals[pos] = vals[pos] == 1.0?15.0: vals[pos]==2.0?35.0:
vals[pos]==3.0?50.0:vals[pos];
        }
    }

    try {
        ret = this.fuzzySYS.deduce(vals);
    } catch (Exception ex) {
        System.err.println("Fuzzy error: "+ex.toString());
    }

    return ret;
}

public double neuroInfer(double[] vals) {
    double ret = 0.0;

    for (int pos = 0; pos < vals.length; pos++) {
        if(pos == 0){
            vals[pos] = vals[pos] == 1.0?1.0: vals[pos]==2.0?3.0:
vals[pos]==3.0?5.0:vals[pos];
        }
        else if(pos == 1){
            vals[pos] = vals[pos] == 1.0?50.0: vals[pos]==2.0?100.0:
vals[pos]==3.0?150.0:vals[pos];
        }
        else if(pos == 2){
            vals[pos] = vals[pos] == 1.0?10.0: vals[pos]==2.0?40.0:
vals[pos]==3.0?70.0:vals[pos];
        }
        else if(pos == 3){
            vals[pos] = vals[pos] == 1.0?10.0: vals[pos]==2.0?40.0:
vals[pos]==3.0?70.0:vals[pos];
        }
        else if(pos == 4){
            vals[pos] = vals[pos] == 1.0?10.0: vals[pos]==2.0?70.0:
vals[pos]==3.0?120.0:vals[pos];
        }
        else if(pos == 5){
            vals[pos] = vals[pos] == 1.0?15.0: vals[pos]==2.0?35.0:
vals[pos]==3.0?50.0:vals[pos];
        }
        else if(pos == 6){
            vals[pos] = vals[pos] == 1.0?0.2: vals[pos]==2.0?0.5:
vals[pos]==3.0?0.7:vals[pos];
        }
    }
}

```

```

        else if(pos == 7){
            vals[pos] = vals[pos] == 1.0?15.0: vals[pos]==2.0?35.0:
vals[pos]==3.0?50.0:vals[pos];
        }

        //System.out.println(vals[pos]);
    }

    List<Double> data = new ArrayList<Double>();
    for (int i = 0; i < vals.length; ++i) {
        data.add(Double.valueOf(vals[i]));
    }

    try {
        this.net.evaluate(data);
        ret = this.net.getResult();
    } catch (Exception ex) {
        System.err.println("Neuro error"+ex.toString());
    }

    return ret;
}

```

```

public double gaNeuroFuzzyInfer(int[] vals) {
    List<Double> data = new ArrayList<Double>();
    for (int i = 0; i < vals.length; ++i) {
        data.add(Double.valueOf(vals[i]));
    }

    try {
        NeuralNet netf = new NeuralNet(100, 8, 5, 4, 0.01);
        netf.learn(1);
        netf.evaluate(data);
        double ret = netf.getResult();
        return ret;
    } catch (Exception ex) {
        return 0.0;
    }
}

```

```

}
/*

```

- * To change this license header, choose License Headers in Project Properties.
- * To change this template file, choose Tools | Templates
- * and open the template in the editor.

```

*/

package ga.neuro.fuzzy.diabetes.lib;

/**
 *
 * @author Oracle1
 */
public class KnowledgeBase {

}

/*
 * To change this license header, choose License Headers in Project Properties.
 * To change this template file, choose Tools | Templates
 * and open the template in the editor.
 */
package ga.neuro.fuzzy.diabetes.lib;

import java.io.BufferedReader;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.IOException;
import java.util.HashMap;
import java.util.Map;

public class ReadCVS {

    public Map<Integer, String[]> data;
    public int noline;

    public ReadCVS() {
        data = new HashMap<Integer, String[]>();
        noline = 0;
    }

    public void read(String Path) {
        noline = 0;
        String csvFile = Path;
        BufferedReader br = null;
        String line = "";
        String cvsSplitBy = ",";

        try {
            br = new BufferedReader(new FileReader(csvFile));
            while ((line = br.readLine()) != null) {
                // use comma as separator
                String[] rline = line.split(cvsSplitBy);
            }
        }
    }
}

```

```

        savedata(rline);
    }

    } catch (FileNotFoundException e) {
        e.printStackTrace();
    } catch (IOException e) {
        e.printStackTrace();
    } finally {
        if (br != null) {
            try {
                br.close();
            } catch (IOException e) {
                e.printStackTrace();
            }
        }
    }
}

//System.out.println("Done");
}

public static boolean isNumeric(String str) {
    return str.matches("[+-]?\\d*(\\.\\d+)?");
}

private void savedata(String[] rline) {
    data.put(nolines, rline);
    nolines++;
    //System.out.println(nolines);
}

}

package ga.neuro.fuzzy.diabetes.lib.fuzzy;

/**
 * Implements a fuzzy set defined by its domain and a reference to a membership function.
 *
 */

public class CalculatedFuzzySet implements IFuzzySet{

    private IDomain domain;
    private IIntUnaryFunction function;

    public CalculatedFuzzySet(IDomain domain, IIntUnaryFunction function) {
        this.domain = domain;
        this.function = function;
    }

```

```

    }

    @Override
    public IDomain getDomain() {
        return domain;
    }

    @Override
    public double getValueAt(DomainElement e) {
        return function.valueAt(domain.indexOfElement(e));
    }

    @Override
    public IFuzzySet cutAt(double mu) {
        MutableFuzzySet ret = new MutableFuzzySet(this.domain);
        for (DomainElement e : this.domain) {
            ret = ret.set(e, this.getValueAt(e) * mu);
        }
        return ret;
    }

    @Override
    public double getMemberI() {
        throw new UnsupportedOperationException("Not supported yet."); //To change body
of generated methods, choose Tools | Templates.
    }
}

package ga.neuro.fuzzy.diabetes.lib.fuzzy;

import java.util.Arrays;
import java.util.Iterator;
import java.util.NoSuchElementException;

/**
 * <code>CompositeDomain</code> class models a cartesian product of multiple
<code>SimpleDomain</code> instances.
 *
 */
public class CompositeDomain extends Domain{

    private SimpleDomain[] components;

    public CompositeDomain(SimpleDomain ... components) {
        super();
        this.components = components;
    }
}

```

```

@Override
public int getCardinality() {
    int ret = 1;
    for (SimpleDomain component : components) {
        ret *= component.getCardinality();
    }
    return ret;
}

@Override
public IDomain getComponent(int idx) {
    try {
        return components[idx];
    } catch (Exception e) {
        System.err.println("Index out of range!");
    }
    return null;
}

@Override
public int getNumberOfComponents() {
    return components.length;
}

@Override
public Iterator<DomainElement> iterator() {
    return new CompositeDomainIterator();
}

private class CompositeDomainIterator implements Iterator<DomainElement> {

    private boolean done;

    private int cnt;
    private double[] cursorValue;

    public CompositeDomainIterator() {
        done = false;
        cnt = CompositeDomain.this.getNumberOfComponents();
        cursorValue = new double[cnt];
        for (int i = 0; i < cnt; ++i) {
            cursorValue[i] = ((SimpleDomain)
CompositeDomain.this.getComponent(i)).getFirst();
        }
    }
}

```

```

        private boolean finished() {
            for (int i = 0; i < cnt; ++i) {
                if (cursorValue[i] < ((SimpleDomain)
CompositeDomain.this.getComponent(i)).getLast() - 1)
                    return false;
            }
            return true;
        }

        @Override
        public boolean hasNext() {
            return !this.done;
        }

        @Override
        public DomainElement next() {
            if (this.hasNext()) {
                double[] oldCursor = new double[cnt];
                oldCursor = Arrays.copyOf(cursorValue, cnt);
                DomainElement ret = DomainElement.of(oldCursor);
                this.done = finished();
                for (int i = cnt - 1; i >= 0; --i) {
                    if (cursorValue[i] == ((SimpleDomain)
CompositeDomain.this.getComponent(i)).getLast() - 1)
                        continue;
                    ++cursorValue[i];
                    for (int j = i + 1; j < cnt; ++j)
                        cursorValue[j] = ((SimpleDomain)
CompositeDomain.this.getComponent(j)).getFirst();
                    break;
                }
                return ret;
            }
            throw new NoSuchElementException();
        }

        @Override
        public void remove() {
            throw new UnsupportedOperationException();
        }
    }

}

package ga.neuro.fuzzy.diabetes.lib.fuzzy;

import java.util.Arrays;

```

```

/**
 * Class which models a single domain element.
 *
 */

public class DomainElement {

    private double[] values;

    public DomainElement(double ... values) {
        super();
        this.values = values;
    }

    public int getNumberOfComponents() {
        return this.values.length;
    }

    public double getComponentValue(int i) throws Exception {
        if (i < 0 || i > this.getNumberOfComponents())
            throw new Exception("Index out of range!");
        return values[i];
    }

    public static DomainElement of(double ... values) {
        return new DomainElement(values);
    }

    public boolean lessThan(DomainElement other) {
        try {
            for (int i = 0; i < this.getNumberOfComponents(); ++i) {
                if (this.getComponentValue(i) !=
other.getComponentValue(i))
                    return this.getComponentValue(i) <
other.getComponentValue(i);
            }
        } catch (Exception e) {
            System.err.println("Non comparable domain elements");
        }
        return false;
    }

    @Override
    public int hashCode() {
        final int prime = 31;

```

```

        int result = 1;
        result = prime * result + Arrays.hashCode(values);
        return result;
    }

    @Override
    public boolean equals(Object obj) {
        if (this == obj)
            return true;
        if (obj == null)
            return false;
        if (getClass() != obj.getClass())
            return false;
        DomainElement other = (DomainElement) obj;
        if (!Arrays.equals(values, other.values))
            return false;
        return true;
    }

    @Override
    public String toString() {

        if (this.getNumberOfComponents() == 1) {
            try {
                return
Double.valueOf(this.getComponentValue(0)).toString();
            } catch (Exception ignorable) {}
        }

        StringBuilder sb = new StringBuilder();
        sb.append('(');

        for (int i = 0; i < values.length; ++i) {
            sb.append(values[i]);
            if (i != values.length - 1)
                sb.append(", ");
        }

        sb.append('');

        return sb.toString();
    }
}

package ga.neuro.fuzzy.diabetes.lib.neuro;

import ga.neuro.fuzzy.diabetes.lib.Database;

```

```

import java.sql.SQLException;
import java.util.ArrayList;
import java.util.List;

/**
 * Class which models a feedforward neural network. Concrete learning
 * algorithms are implemented as extensions of this class.
 */

public class NeuralNet {

    protected int maxIter;
    protected int sampleSize;

    protected double learningRate;

    protected List<Layer> layers;
    protected List<List<Double>> samples;
    protected List<List<Integer>> answers;

    //public NeuralNet(String sampleFilePath, int maxIter, int sampleSize, int
    hiddenLayerSize, int layerCnt, double learningRate) throws Exception {
    public NeuralNet(int maxIter, int sampleSize, int hiddenLayerSize, int layerCnt,
    double learningRate) throws Exception {
        //this.sampleSize = sampleSize;
        this.layers = new ArrayList<>();
        this.maxIter = maxIter;
        this.learningRate = learningRate;

        if (layerCnt < 3) {
            throw new Exception("Network must have at least 3 layers");
        }

        layers.add(new Layer(sampleSize, true, hiddenLayerSize));

        for (int i = 0; i < layerCnt - 3; ++i) {
            layers.add(new Layer(hiddenLayerSize, false, hiddenLayerSize));
        }

        layers.add(new Layer(hiddenLayerSize, false, 1));
        layers.add(new Layer(1, false, 0));

        samples = new ArrayList<List<Double>>();
        answers = new ArrayList<List<Integer>>();

        try {

```

```

        Database con = new Database();
        con.connect();
        int ret = con.executeQuery("SELECT `knowledge_base`.* FROM
`knowledge_base`");
        if ((ret == 1) && (con.num_of_rows > 0)) {
            con.rs.next();
            //txtRules.setText(con.rs.getString(2));

            int count = con.rs.getMetaData().getColumnCount(); //number of column
            this.sampleSize = count-2;

            String tmpStr[] = new String[count];
            while (con.rs.next()) {
                List<Double> sample = new ArrayList<Double>();
                List<Integer> answer = new ArrayList<Integer>();
                int i = 2;
                for (i = 2; i < count; i++) {
                    sample.add(Double.valueOf(con.rs.getString(i)));
                }

                answer.add(Integer.valueOf(con.rs.getString(i)));

                samples.add(sample);
                answers.add(answer);
            }
        } catch (SQLException ex) {

//Logger.getLogger(knowledgeBaseApp.class.getName()).log(Level.SEVERE, null, ex);
        }

    }

    public void evaluate(List<Double> sample) {

        for (int i = 0; i < sample.size(); ++i) {
            layers.get(0).getNeuron(i).setW0(sample.get(i));
        }

        for (int i = 1; i < layers.size(); ++i) {
            layers.get(i).solveLayer(layers.get(i - 1));
        }

    }

    public String getLetter() {
        //String[] letters = {"Alpha", "Beta", "Gamma", "Delta", "Epsilon"};

```

```

        //int pivot = 0;
        Layer last = layers.get(layers.size() - 1);
        //for (int i = 1; i < 5; ++i)
        //    if (last.getNeuron(i).getY() > last.getNeuron(pivot).getY())
        //        pivot = i;
        //return letters[pivot];
        return String.format("%d", last.getNeuron(1).getY());
    }

    public double getResult() {
        Layer last = layers.get(layers.size() - 1);
        return last.getNeuron(0).getY();
    }

    public double getError() {
        double totalError = 0;
        for (int i = 0; i < samples.size(); ++i) {
            this.evaluate(samples.get(i));
            //for (int j = 0; j < 5; ++j) {
            double diff = (double) answers.get(i).get(0) -
layers.get(layers.size() - 1).getNeuron(0).getY();
            totalError += diff * diff;
            //}
        }
        return totalError;
    }

    // public abstract void learn();

    public void updateWeights(List<List<Double>> samples, List<List<Integer>>
answers, int batchSize) {

        int outLayer = layers.size() - 1;

        for (int i = 0; i < samples.size(); ++i) {

            this.evaluate(samples.get(i));

            //for (int j = 0; j < 5; ++j) {
            //double outDelta =
Math.pow(layers.get(outLayer).getNeuron(j).getY() - (double) answers.get(i).get(j), 2);
            double y = layers.get(outLayer).getNeuron(0).getY();
            double outDelta = y * (1 - y) * (answers.get(i).get(0) - y);
            layers.get(outLayer).getNeuron(0).setDelta(outDelta);
            //}

            for (int j = outLayer - 1; j >= 0; --j) {

```

```

        layers.get(j).updateWeights(layers.get(j + 1), learningRate);
        layers.get(j).updateDeltas(layers.get(j + 1));
    }

    if(i % batchSize == 0) {
        for (Layer l : layers)
            l.swapWeights();
    }

}

public void learn(int batchSize) {
    double error = 0.0;
    for(int i = 0; i < this.maxIter; ++i) {
        error = this.getError();
        updateWeights(samples, answers, batchSize);
    }

    System.out.println(batchSize + " error-> " + error);
}

}

package ga.neuro.fuzzy.diabetes.lib.ga;

import java.io.BufferedReader;
import java.io.FileInputStream;
import java.io.IOException;
import java.io.InputStreamReader;
import java.util.ArrayList;
import java.util.List;

/**
 * Class which models experiment measurements.
 *
 */

public class Measurements {

    private List<Double> x, y, fxy;

    public Measurements(String fileName) throws NumberFormatException,
    IOException {

        x = new ArrayList<Double>();

```

```

        y = new ArrayList<Double>();
        fxy = new ArrayList<Double>();

        BufferedReader bf = new BufferedReader(new InputStreamReader(new
FileInputStream(fileName)));

        String line;
        while ((line = bf.readLine()) != null) {
            String words[] = line.split("\\s+");
            x.add(Double.parseDouble(words[0]));
            y.add(Double.parseDouble(words[1]));
            fxy.add(Double.parseDouble(words[2]));
        }

        bf.close();
    }

    public double evaluateOne(double x, double y, Chromosome c) {
        double[] b = c.getBeta();
        return Math.sin(b[0] + b[1] * x) + b[2] * Math.cos(x * (b[3] + y)) / (1 +
Math.exp((x - b[4]) * (x - b[4])));
    }

    public List<Double> evaluateAll(Chromosome c) {
        List<Double> ret = new ArrayList<Double>();
        for (int i = 0; i < x.size(); ++i)
            ret.add(evaluateOne(x.get(i), y.get(i), c));
        return ret;
    }

    public double getError(List<Double> gxy) {
        double ret = 0;
        for (int i = 0; i < fxy.size(); ++i)
            ret += (fxy.get(i) - gxy.get(i)) * (fxy.get(i) - gxy.get(i));
        return ret / (double) fxy.size();
    }
}

package ga.neuro.fuzzy.diabetes.lib.ga;

import java.util.Random;

/**
 * Class which models a single chromosome unit for a genetic algorithm.
 *
 */

```

```

public class Chromosome implements Comparable<Chromosome>{

    private double[] beta;
    private double error;

    private Measurements m;

    public Chromosome(Measurements m) {
        beta = new double[5];
        this.m = m;
    }

    public Chromosome(double[] beta, Measurements m) {
        this.beta = beta;
        this.m = m;
        this.error = m.getError(m.evaluateAll(this));
    }

    public double[] getBeta() {
        return beta;
    }

    public void setBeta(double[] beta) {
        this.beta = beta;
        this.error = m.getError(m.evaluateAll(this));
    }

    public void generate_random() {
        Random r = GeneticConstants.r;
        for (int i = 0; i < 5; ++i)
            beta[i] = -4 + 8 * r.nextDouble();
        error = m.getError(m.evaluateAll(this));
    }

    public static Chromosome crossover(Chromosome c1, Chromosome c2) {
        double[] newBeta = new double[5];
        for (int i = 0; i < 5; ++i)
            newBeta[i] = (c1.getBeta()[i] + c2.getBeta()[i]) / 2;
        return new Chromosome(newBeta, c1.m);
    }

    public void mutate() {
        Random r = GeneticConstants.r;
        for (int i = 0; i < 5; ++i)
            if (r.nextDouble() < GeneticConstants.P_MUTATION)
                beta[i] = -4 + 8 * r.nextDouble();
    }
}

```

```

        this.error = m.getError(m.evaluateAll(this));
    }

    public double getError() {
        return error;
    }

    @Override
    public int compareTo(Chromosome o) {
        double e1 = this.getError(), e2 = o.getError();
        if (e1 < e2) return -1;
        if (e1 > e2) return 1;
        return 0;
    }
}

```