

Soft-computing: An Objective Approach in Varied Diabetes Recognition

Obi J.C¹ and Imianvan A.A².

^{1&2} *Department of Computer Science, University of Benin, Benin City, Nigeria.*
triplejo2k2@yahoo.com ; tonyvanni@yahoo.com

ABSTRACT

Diabetes is a chronic disorder caused by elevated glucose within the blood stream. The predominant indicator of diabetes include a glucose level of more 125mg/dl in addition to frequent thirst, unusual thirst, extreme fatigue blurred vision and frequent infection. Existing approach for the recognition of diabetes are to two classes (Type I and Type II) in addition to their subjective approach. This research paper proposed an objective approach utilizing soft-computing techniques for the recognition of five class of diabetes.

Keywords: Fuzzy Logic, Fuzzy Set, Fuzzy Linguistic variables Genetic Algorithm, Neural Network,

1 Introduction

Diabetes; once considered a rare disease in sub-Saharan Africa is one of the most predominate health issue across Africa. In 2010, over 12 million people in sub-Saharan Africa are estimated to have diabetes, and 330,000 people die annually from diabetes-related conditions (DLF, 2011). Over the next 20 years, it is predicted that sub-Saharan Africa will have the highest growth in the number of people with diabetes of any region in the world, which is expected to double in 20 years, reaching 23.9 million by 2030 (DLF, 2011).

In Nigeria about 2million persons are living with diabetes, many people are living with the condition unaware of the seriousness of the disease and its consequences as those diagnosed are often poorly managed due to lack of resources or because the health care professionals vested with such care have limited knowledge about diabetes and how to provide good care. Diabetes might overtake those suffering from Tuberculosis, Malaria, HIV/AIDS, and other terminal diseases by the year 2030 if adequate attention was not paid toward the provision of health education; monitoring, treatment and management are not provided to the masses quickly (Kemi, 2012).

Diabetes is a chronic, debilitating disease requiring life-long management which invariably reduces the risk of serious, long-term complications such as kidney infection, liver disease and glaucoma etc. Offering the long-term monitoring and treatment needed is not easy for the healthcare systems of sub-Saharan Africa in general and Nigeria in particular, which are more focused on treatment and management of acute infection. Awareness of the early symptoms of diabetes is low, even among healthcare professionals. 85% of diabetes cases are undiagnosed, remaining without treatment and increasing the chances of untimely death (DLF, 2011). The professional knowledge in these regions if

DOI: 10.14738/jbemi.15.402

Publication Date: 29th September 2014

URL: <http://dx.doi.org/10.14738/jbemi.15.402>

available is tied mainly to type I and type II diabetes and rest in the hand of senior consultants and physicians. Other forms of diabetes such as gestational, Maturity onset Diabetes of the Young (MODY) and Latent Autoimmune Diabetes in Adulthood (LADA) are really difficult to comprehend from the physician standpoint due to their obsolete knowledge, lack of experience and exposure.

This research paper proposed an objective soft-computing approach utilizing the three fundamental concepts; neural network, genetic algorithm and fuzzy logic as a means for recognition five class of diabetes.

2 Review of Related Literature

The Multiple Knot Spline Smooth Support Vector Machine (MKS-SSVM) proposed by Shanti et al., (2009) is an algorithm for data mining techniques extending the original smoothing techniques introduced by Lee and Mangasarian (2001) to generate a smooth function which approximate the plus function for enhancing the performance, accuracy and result of Smooth Support Vector Machine (SSVM). The plus function which is the modification of the original three order spline function was the focal point of their research which was applied to the diagnosis of Type I diabetes utilizing the Pima Indian Dataset. SSVM and MKS-SSVM were chosen as material and method. The result of their research presented MKS-SSVM as significantly improving performance, accuracy and obtaining promising result to help diagnosis compared to SSVM although their computational time were Identical.

Madhavi and Bamnote (2012) proposed a predictive modeling of clinical data using a hybrid of neuro-fuzzy (soft-computing techniques) and data mining techniques for type I and type II diabetes recognition. The first phase of the system is preprocessing of diabetes data which is achieved utilizing soft-computing techniques. The second phase applies the different classification algorithm to the preprocessed data. The classification algorithm aims at identifying the characteristics that indicate the group to which each case belong. Decision-tree, neural network, multivariate adaptive regression spline (MARS), rule induction and K-Nearest are some of the mining techniques utilized. The performance of the proposed predictive model is measured utilizing Accuracy and Receiver Operator Characteristic Curve (AROC). In conclusion effective predictive model is achievable using preprocessing and data mining techniques for diabetes recognition.

Artificial Neural Networks (ANNs) constitute a class of flexible nonlinear models designed to mimic biological neural systems. An ANN is a mathematical model or computational model based on biological neural networks (Gutiérrez, 2011), as an interconnected group of artificial neurons, which carries out computation using a connectionist approach. Typically, a biological neural system consists of several layers, each with a large number of neural units (neurons) that can process the information in a parallel manner. The models with these features are known as ANN models (Robert, 2000). ANNs have been widely applied to solve many difficult problems in different areas, including pattern recognition (matching), signal processing, language learning, electronic medical record processing, tele-diagnosis and computer networking (Robert, 2000). Neural network utilize dataset. The data set is divided into three distinct sets: training, testing and validation sets. The training set is the largest set and is used by neural network to learn patterns present in the data. The testing set is used to evaluate the generalization ability of a supposedly trained network. A final check on the performance of the trained

network is made using validation set. Learning methods in neural networks can be broadly classified into three basic types Supervised, unsupervised and reinforced learning (Diogo et al. 2008):

The theory of fuzzy logic provides a mathematical strength to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning. In standard set theory, an object does or does not belong to a set. There is no middle ground. In such bivalent systems, an object cannot belong to both its set and its compliment set or to neither of them. This principle preserves the structure of the logic and avoids the contradiction of object that both is and is not a thing at the same time (Zadeh, 1965). However, fuzzy logic is highly abstract and employs heuristic (experiment) requiring human experts to discover rules about data relationship (Angel and Rocio, 2011).

Fuzzy classification assumes the boundary between two neighboring classes as a continuous, overlapping area within which an object has partial membership in each class (Kuang et al., 2011). Fuzzy logic highlights the significant of most applications in which categories have fuzzy boundaries, but also provides a simple representation of the potentially complex partition of the feature space. (Sun and Jang, 1993 and Ahmad, 2011) Conventional approaches of pattern classification involve clustering training samples and associating clusters to given categories. The complexity and limitations of previous mechanisms are largely due to the lack of an effective way of defining the boundaries among clusters. This problem becomes more intractable when the number of features used for classification increases (Christos and Dimitros, 2008).

The Genetic Algorithm (GA) is a search and optimization technique based on the principles of genetics and natural selection. They represent processes in nature that are remarkably successful at optimizing natural phenomena. They are capable of solving other types of problems, using genetic operators abstracted from nature; they form a mechanism suitable for a variety of search problems. These algorithms encode a potential solution to a specific problem on a simple chromosome-like data structure and apply recombination operators to these structures so as to preserve critical information. Genetic algorithms are often viewed as function optimizer. The main idea is survival of the fittest (natural selection). Genetic algorithm is composed of three main genetic operators namely; **Selection**: is a way for the genetic algorithm to move toward promising regions in the search space. **Mutation**: is a genetic operator that changes one or more gene values in a chromosome. The mutation process helps to overcome trapping at local maxima. **Crossover**: Exchanging Chromosomes portions of genetic materials.

3 Methodology, Design and Result

Carefully observing these reviewed literatures propagated these drawbacks:

- a) *Existing diabetes models are tied to two classes; type I and type II, with none of these existing models capable to diagnose the three current classes of diabetes namely; Gestational diabetes, Maturity Onset Diabetes of the Young (MODY) and Latent Autoimmune Diabetes in Adulthood (LADA).*
- b) *The fuzzy scaled membership function boundary was in most cases subjective (assumed) and not based on objective (scientific) approach which is achievable utilizing a search and optimization techniques which has invariably lead to flaw medical diagnosis in most cases.*

3.1 The Proposed Soft-computing Model

The proposed model is a Hybrid architectural framework which harnessing the three fundamental soft-computing techniques with the aim of provide class distinction for the differential recognition of varied diabetes identification utilizing the predominant fuzzy set, in addition with the occurrence factors thereby establishing a conclusive boundary. Unlike the current approaches, in which success or failure are based on the wills and experiences of relevant personnel designing and administrating the approach in other to elicit relevant recognition points. This model is artificial intelligence based; therefore success and failure are not dependent on human intuitions, but success, is closely linked within tuned-up approaches within the system components.

The Dataset present in Table 3.1 was obtained through a research survey, utilizing questionnaires as the research tool. The quantitative and qualitative questionnaires comprises of two phase. The first phase contains demographic information's while the second phase tele-medical information. A total of fifty questionnaires were constructed and distributed to various expert diabetes medical professional spread across eight teaching hospitals within five Geo-political regions in Nigeria. All questionnaires administrated were retrieved without mutilation. In other to generate a Fuzzy Linguistic variable dataset, all questionnaires responds were tuned utilizing the proposed Adapted Fuzzy Cluster Mean Equation:

$$AFCME = \sum (A, B, C, D, E)*X \quad (3.1)$$

Where *A, B, C, D and E = Picked Questionnaires Questions Options (PQQO)*

X (0.02) = Assigned Question Option Fuzzy Range Value (AQOFRV)

Unpicked options = 0.00

Utilizing the Genetic Algorithm Procedures proposed by (obi and Imainavan, 2013), the generated membership function derived from our tuned questionnaires were optimized to obtain the fuzzy membership function boundary of 0.53 specified in Table 1, which is invariably attached to the linguistic variable represented as **Serious**, while **Moderate** is 0.4 and 0.3 and below **Minor**

Table 1: Optimized Dataset for the Varied Form of Diabetes (Scale 0.00 - 1.00)

Symp. Codes	Symptoms/ Fuzzy Set (Parameters)	Degree of Membership Function for the Varied Forms of Diabetes (Scale ranging from 0.00 -1.00)				
		Cluster 1 (Type I)	Cluster 2 (Type II)	Cluster 3 (Gestational)	Cluster 4 (MODY)	Cluster 5 (LADA)
P01	Frequent urination	0.50	0.00	0.00	0.00	0.50
P02	Unusual thirst	0.50	0.00	0.00	0.00	0.50
P03	Extreme hunger	0.50	0.00	0.00	0.00	0.50
P04	Unusual weight loss	0.50	0.00	0.00	0.00	0.50
P05	Extreme fatigue	0.50	0.00	0.00	0.00	0.50
P06	Irritability	0.00	0.00	0.50	0.50	0.00
P07	Frequent infections	0.00	0.00	0.50	0.50	0.00
P08	Blurred vision	0.00	0.50	0.00	0.50	0.00
P09	Slow to heal cuts/bruises	0.00	0.50	0.50	0.00	0.00
P10	Tingling/numbness in hands/feet	0.00	0.50	0.00	0.50	0.00
P11	Regular skin/gum/bladder infections	0.00	0.50	0.50	0.00	0.00
P12	Nausea and vomiting	0.00	0.50	0.50	0.00	0.00
P13	Hemoglobin A1c test (HbA1c) >10	0.20	0.20	0.20	0.20	0.20
P14	Rapid Weight loss	0.00	0.00	0.00	0.50	0.50
P15	Leg Cramp	0.20	0.20	0.20	0.20	0.20

3.2 Fuzzy Integration

The fuzzy partition for each input feature consists of clinical symptoms of for the varied forms of diabetes (frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection, blurred vision, slow to heal cuts/bruises, tingling/numbness in hands/feet, regular skin/bladder/gum infection, nausea/vomiting, hemoglobin A1c test (HbA1c) >10, rapid weight loss and leg cramp). However, it can occur that if the fuzzy partition for the varied forms of diabetes is not set up correctly, or if the number of linguistic terms for the input features is not large enough, then some patterns will be misclassified. The linguistic label serious is the base point for classification.

- a) Not diagnose with a class diabetes (C_1)
- b) Might be diagnose with a class of diabetes (C_2)
- c) Diagnose with a class of diabetes (C_3)

If the patient is exhibiting at least three or less of the symptoms of a class of diabetes THEN (C_1), if the patient is exhibiting exactly four of the symptoms of a class of diabetes THEN (C_2) and if the patients is exhibiting five or more of the symptoms of a class of diabetes THEN (C_3).

The Fuzzy IF-THEN Rules that can be generated from the initial fuzzy partitions of the classification of varied for diabetes is thus:

- R1:** IF the patient exhibit *frequent urination* and its *serious* THEN class C_1 .
- R2:** IF the patient exhibit *frequent urination* and *unusual thirst* and both symptoms are *serious* THEN class C_1 .
- R3:** IF the patient exhibit *frequent urination, unusual thirst and extreme hunger* and these symptoms are *serious* THEN class C_1 .
- R4:** IF the patient exhibit *frequent urination, unusual thirst, extreme hunger* and *unusual weight loss* and these symptoms are *serious* THEN C_2 .
- R5:** IF the patient exhibit *frequent urination, unusual thirst, extreme hunger, unusual weight loss* and *extreme fatigue* and these symptoms are *serious* THEN class C_3 .
- R6:** IF the patient exhibit *frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue* and *irritability* and these symptoms are *serious* THEN class C_3 .
- R7:** IF the patient is exhibits *frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability* and *frequent infection* and these symptoms are *serious* THEN class C_3 .
- R8:** IF the patient exhibits *frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection* and *blurred vision* and these symptoms are *serious* THEN class C_3 .
- R9:** IF the patient exhibit frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection, blurred vision and slow to heal cuts/bruises and these symptoms is serious THEN class C_3 .
- R10:** IF the patient exhibit frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection, blurred vision, slow to heal cuts/bruises and tingling/numbness in hands/feet and these symptoms are serious THEN class C_3 .
- R11:** IF the patient exhibit frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection, blurred vision, slow to heal cuts/bruises, tingling/numbness in hands/feet and regular skin/bladder/gum infection and these symptoms are serious THEN class C_3 .
- R12:** IF the patient exhibit frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection, blurred vision, slow to heal cuts/bruises, tingling/numbness in hands/feet, regular skin/bladder/gum infection and nausea/vomiting, *and these symptoms are serious* THEN class C_3 .
- R13:** IF the patient exhibit frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection, blurred vision, slow to heal cuts/bruises, tingling/numbness in hands/feet, regular skin/bladder/gum infection, nausea/vomiting, and *hemoglobin A1c test (HbA1c) >10* and these symptoms are serious THEN class C_3 .

R14: IF the patient exhibits frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection, blurred vision, slow to heal cuts/bruises, tingling/numbness in hands/feet, regular skin/bladder/gum infection, nausea/vomiting, *hemoglobin A1c test (HbA1c) >10* and *rapid weight loss and these symptoms are serious* THEN class C₃.

R15: IF the patient exhibits frequent urination, unusual thirst, extreme hunger, unusual weight loss, extreme fatigue, irritability, frequent infection, blurred vision, slow to heal cuts/bruises, tingling/numbness in hands/feet, regular skin/bladder/gum infection, nausea/vomiting, *hemoglobin A1c test (HbA1c) >10, rapid weight loss and Leg Cramp* and these symptoms are serious THEN class C₃.

3.3 Diabetes Diagnosis

If a patient is exhibiting five or more symptoms of diabetes with linguistic label serious in collaboration with a Glucose level $\geq 125\text{mg/dl}$, Age range = *Teenager*, Origin = *Caucasians* and pancreas obliteration = *swift* Then **Type I**

If a patient is exhibiting five or more symptoms of diabetes with linguistic label serious in collaboration with a Glucose level $\geq 125\text{mg/dl}$, Age range = *Post Adult*, Origin = *Blacks* and Pancreas Obliteration = *swift* Then **Type II**

If a patient is exhibiting five or more symptoms of diabetes with linguistic label serious in collaboration with a Glucose level $\geq 125\text{mg/dl}$, Age range = *Pre-manopause*, Origin = *Plus* Then **Gestational**

If a patient is exhibiting five or more symptoms of diabetes with linguistic label serious in collaboration with a Glucose level $\geq 125\text{mg/dl}$, Age range = *Teenager*, and mutated dominant gene = *present* and Origin = *Caucasians* Then **MOYD**

If a patient is exhibiting five or more symptoms of diabetes with linguistic label serious in collaboration with a Glucose level $\geq 125\text{mg/dl}$, Age range = *Pre- Adult*, Origin = *Caucasian* and pancreas obliteration = *Progressive* Then **LADA**

The proposed model is depicted on Figure 1,

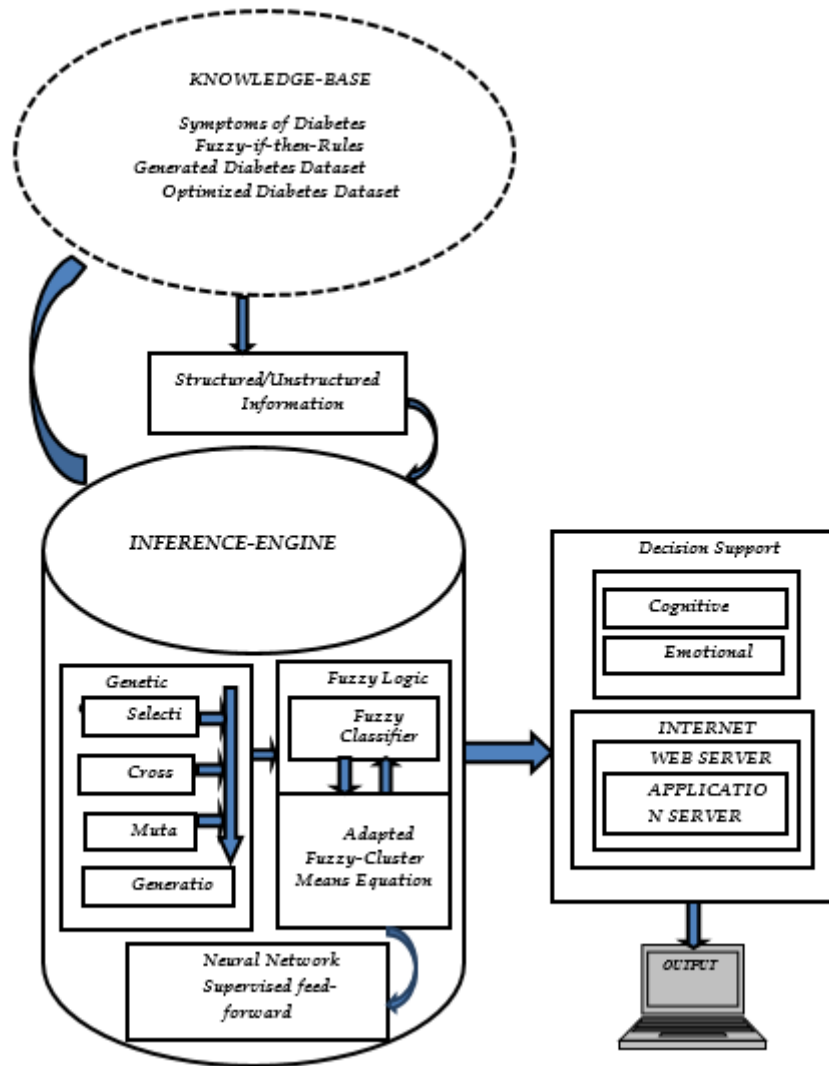


Figure 1: Soft computing Architecture for the Diagnosis of varied forms of Diabetes

3.4 Unified Modeling Language (UML)

Unified Modeling Language (UML) is a standard modeling language used for modeling software systems. The focus of UML is on creating simple, well documented and easy to understand software models. Use case and Sequence Diagram where utilized for depicting the two main view of the proposed system.

Figure 2 and 3 shows the user case model for varied forms of diabetes of our proposed system.

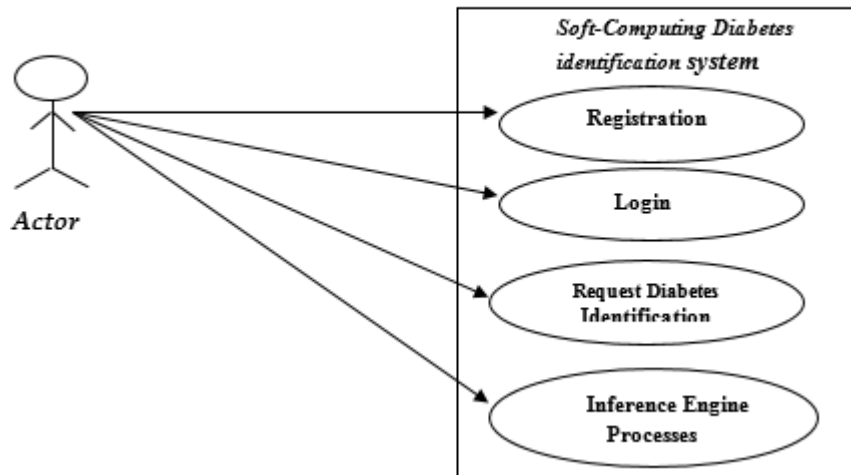


Figure 2: Use Case Diagram modeling Diabetes software System

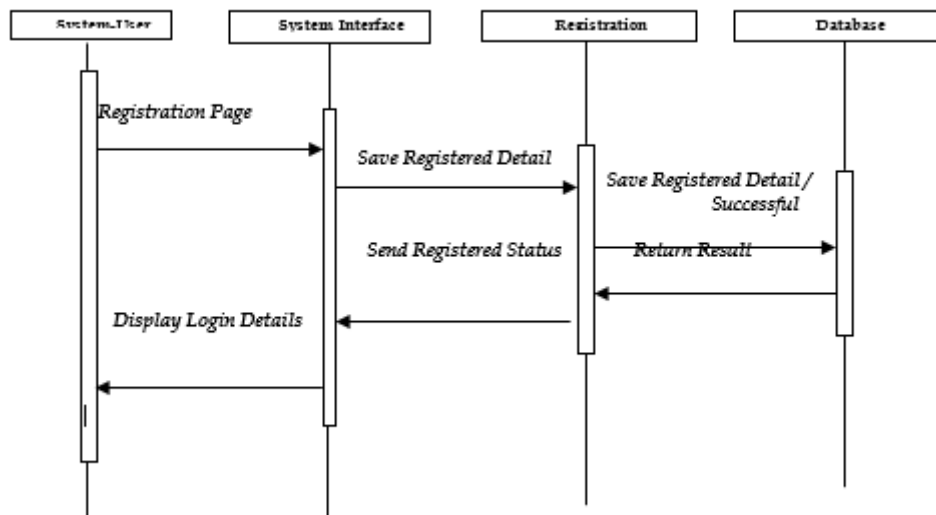


Figure 3: 18 Sequence Diagram modeling Registration Process

4 Implementation and Results

Matrix Laboratory (MATLAB) serves as our simulation tool in achieving the our results because of its interactive environment for algorithm development, data visualization, data analysis, and numerical approach which was relevant to our numerical dataset which was more appropriate than with spreadsheets or traditional programming languages, such as C/C++ or Java.

4.1 Results

The system was successfully installed and tested on the following operating systems: Microsoft Windows (XP, Vista & 7), Ubuntu 9 and Solaris 10. The system was validated with online records of Diabetes cases with slight deviation from results, corrections were effected. University of Benin

Teaching Hospital (UBTH) provided a communicative framework for testing the system on a one-on-one basis. Our simulation results are presented on Table 2 and 3 respectively.

Table 2: Subjective Diagnosis Approach

	Diabetic Patient	Type I	Type II	Gestational	MODY	LADA	Diabetes (Nil)
1.	Patient A	-		-	-	-	-
2.	Patient B	-		-	-	-	-
3.	Patient C	-		-	-	-	-
4.	Patient D	-		-	-	-	-
5.	Patient E	-		-	-	-	-
6.	Patient F	-		-	-	-	-
7.	Patient G	-		-	-	-	-
8.	Patient H	-		-	-	-	-
9.	Patient I	-		-	-	-	-
10.	Patient J	-		-	-	-	-

Table 3: Soft-Computing Expert System

		Varied Classes of Diabetes					
	Diabetic Patient	Type I	Type II	Gestational	MODY	LADA	Diabetes (Nil)
1.	Patient A	-	-	-	-	-	
2.	Patient B	-	-	-	-	-	
3.	Patient C	-		-	-	-	-
4.	Patient D	-		-	-	-	-
5.	Patient E	-		-	-	-	-
6.	Patient F	-		-	-	-	-
7.	Patient G	-		-	-	-	-
8.	Patient H	-	-			-	-
9.	Patient I	-	-			-	-
10.	Patient J	-	-			-	-

5 Conclusion

A soft-computing system has been design, developed, implemented and simulated for the recognition of five class of diabetes diagnosis utilizing the fundamental concepts of soft-computing. Unlike previous approaches in which successfully and failure are tied fully to human intuition, this is an objective system.

REFERENCES

- [27]. Ahmad H. (2011), "Fuzzy approach to Likert Spectrum in Classified levels in surveying researches" retrieved <http://www.tjmcs.com>.
- [28]. Angel C. and Rocio R. (2011), "Documentation management with Ant colony Optimization Meta-heuristic: A Fuzzy Text Clustering Approach Using Pheromone trails" retrieved from soft computing in Industrial applications, Advances in intelligent and soft Computing, vol. 96, 2011, 261-70, DOI: 10.1007/978-3-642-20505-1_23
- [29]. DLF: Diabetes Leadership Forum (2011), "Diabetes: the hidden pandemic and its impact on Sub-Saharan Africa" retrieved online from www.novonordisk.com/...diabetes/.../Sub-Saharan%20Africa%20BB.
- [30]. Diogo F. P., Flávio R.S. O. and Fernando B. L. N (2008), "Multi-objective abilities in the Hybrid Intelligent Suite for decision support" retrieved from http://ieeexplore.ieee.org/xpl/freeabs_all.
- [31]. Gutiérrez P.A. (2011), "Hybrid Artificial Neural Networks: Models", retrieved online from <http://dl.acm.org/citation.cfm?id=20233>
- [32]. Kemi O. (2012), "Three Nigerians Suffering from Diabetes" retrieved online from <http://nationalmirroronline.net/index.php/health/health-news/42251.html>
- [33]. Kuang Y. H.; Ting-H. C. and Ting-Cheng Chang (2011), "Determination of the threshold value β of variable precision rough set by fuzzy algorithms" retrieved from <http://www.sciencedirect.com/science/article/pii/S0888613X11000831>
- [34]. Madhavi G. and Bamnote K. (2012), Predictive Diagnosis Model" Clinical Microbiology Reviews, Vol.13, No.1, Pp. 76-82.
- [35]. Robert F. (2000) "Introduction to Neuro-Fuzzy Systems: Advances in Soft Computing Series", Springer-Verlag, Berlin/Heidelberg, Germany.
- [36]. Shanti W. P., Abdullah E., Jasni M. Z, and Rahayu S.P. (2009), "A New Smoth Vector Machine and Its Applications in Diabetes Disease Diagnosis", Journal of Computer Science Vol. 5(12): 1003-1008
- [37]. Sun C.T. and Jang J.S. (1993) "A neuro-fuzzy classifier and its applications", in: Proc. IEEE Int. Conference on Neural Networks, San Francisco, pp.94-98.
- [38]. Zadeh L.A. (1965), "Fuzzy sets. Information and Control", Vol.8, pp.338-353.