Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 38 (2012) 1709 – 1718

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Modeling Optimization and Computing

Improving the Prediction rate of Diabetes Diagnosis Using Fuzzy, Neural Network, Case Based (FNC) approach

Mythili Thirugnanam[a],Dr Praveen Kumar[b],S Vignesh Srivatsan[c],Nerlesh C R[c]

[a]*Assistant Professor(Sr) ,School of Computing Science and Engineering,VIT University , Vellore,TamilNadu ,632014,India*
[b]*Assistant civil surgeon, Government Primary Health Centre, T.Palur, Tamil Nadu, 612904, India*
[c]*Bachelor of technology in Information technology,School of Information Technology,VIT University,Vellore,TamilNadu,632014,India*

Abstract

In today's world the diabetes mellitus has become a major health problem among the people of all ages. Most of the researchers have proposed different systems to diagnose diabetes, in spite of which the accuracy of the prediction rate of diabetes is not so significant. Most of the techniques suggested were to detect diabetes associated with the parts of the body such as eye, heart or legs. All the developed systems aimed to identify diabetes by increasing the accuracy of the prediction rate but fail to do so in most of the cases as there were some issues that had not been discussed. The existing systems had number of drawbacks like some were only concentrated for women and who were less than 21 years old [5], some gave greater accuracy if only one data set was used [6], some needed dataset of very good quality [12], some needed to standardize the ontology [10] and some even failed to show the performance of the developed systems [7] .So there is an need to develop a diabetes diagnosis system that improves the accuracy of the prediction rate by considering all the factors into account. This paper presents a novel approach for diagnosis of diabetes which has two stages to predict the diabetes status. The initial prediction stage adopts two computational intelligence and knowledge engineering techniques such as fuzzy logic (F), neural network (N) and case based reasoning (C) as an individual approach (FCN). The final prediction stage applies rule based algorithm to the values obtained from the initial stage. The benefit of applying these stages is that the accuracy of prediction rate will be higher when compared to using only the initial prediction stage done by most of the suggested systems for predicting the occurrence of diabetes mellitus.

Keywords: Diabetes;Fuzzy Logic;Neural Network;Case Based Reasoning

1877-7058 © 2012 Published by Elsevier Ltd.
doi:10.1016/j.proeng.2012.06.208

1. Introduction

Diabetes mellitus in simple terms called diabetes is a disease in which the blood glucose level of a person is high. This high blood sugar level causes symptom such as frequent urination, increase in thirst level and increased hunger if diagnosis is not done at the right time .There are basically three main types of diabetes. They are Type 1 that results if the insulin is failed to be produced by the body, Type 2 thatresults if the cells fail to use insulin properly and Gestational that is presence of high blood sugar level during pregnancy.

Fuzzy logic is a mathematical model that is applied to number of fields like control theory, artificial intelligence and diagnosis of diabetes as it provides approximate values along with its inference.Mostafa Fathi Ganji et al., (2010) used fuzzy ant colony optimization technique for diagnosis of diabetes by extracting a set of fuzzy If-Then rules (FADD) [11]. Neural network is a computation model that has usages in biological and artificial intelligence department for speech recognition, image analysis, and diagnosis of diabetes .It reduces the amount of computation required. K.Rajeshwari et al., (2011) described fuzzy model for diabetic diagnostic decision support in which the procedure includes combination of Fuzzy modeling and artificial neural network architecture [9]. Case Based Reasoning (CBR) is a powerful approach in solving new problems based on the solutions of similar past problems. It helps in identifying diabetes based on earlier occurrences of same problem. Ashraf Mohammed Iqbal et al., (2011) suggested ontology based electronic medical record (EMR) for chronic disease management which helped to achieve semantic interoperability among healthcare information systems and support reasoning for decision support systems [4].The motivation behind proposing this system is the occurrence of diabetes and the fact that many people are unaware of how to diagnose and manage this disease. The proposed system will ease the patients as the need to consult a doctor will not be necessary and also provides information about their level of glucose in the early stages thereby decreasing the number of patients affected by diabetes.

2. Literature Survey

A lot of research has been carried out to diagnose diabetes using computational intelligence and knowledge engineering techniques like fuzzy logic, neural network and CBR and to enhance the diagnosis system and improve the accuracy of prediction rate. Approaches which are discussed in the recent years are briefed below.

R.Vijayamadheswaran et al., (2011) proposed radial basis function to detect diabetes in retinopathy images. The technique includes a combination of two methods such as contextual clustering segmentation method for checking the presence of exudates in the eye and classification of the exudates by radial basis function network. For performing the automation hundred samples of the fundus image were taken. The effectiveness of RBF was around 96%, which shows that it was a better method when compared to using CC alone. The stated RBF technique classifies the segmented information of the image into hard exudates provided that the fundus images were transformed to a template condition and they were of good quality to improve accuracy [12]. K.Rajeshwari et al., (2011) described fuzzy model for diabetic diagnostic decision support. The procedure includes combination of fuzzy modeling and artificial neural network architecture. This approach was used for the patients who were already diabetic, who had new symptoms of diabetes or patients who were normal. The parameters to be modeled for fuzzy logic were obtained by interviewing the patients. The parameters were normalized and classified by artificial neural network as TYPE 2 diabetic or not. The algorithm was performed on samples that had 600 cases. This method was proven to be very efficient and the accuracy was also very good as it works on real time data set

[9].Mostafa Fathi Ganji et al., (2010) proposed fuzzy ant colony optimization technique for diagnosis of diabetes by extracting a set of fuzzy If-Then rules (FADD) This method uses data sets which contain 768 instances, 8 integer-valued attributes and 2 classes. On evaluating the classification system results show that FADD can detect diabetes with notable accuracy by a new framework for learning the rules independently, controlling the influence of pheromone values to improve the quality of rule, make better decisions in next iterations and finally competition and cooperation [11]. Asma A. AlJarullah (2011) developed a decision tree method that predicts patients who were in the initial stage of diabetes. This method had two stages include in it, they were data preprocessing and diabetes prediction model construction using the decision tree method. The software used was WEKA. The dataset was collected from 768 random females who were at least 21 years old. The accuracy obtained was 78.1768% when WEKA decision tree classifier was applied to dataset to construct the decision tree model. The quality of data was improved in the preprocessing phase that includes attribute identification and selection, handling missing values and numerical discretization [5]. Eider Sanchez et al., (2011) developed a knowledge-based decision support system for the diagnosis of Alzheimer disease(AD) .They proposed a decision support system which consist of three ontologies, such as the MIND , SWAN and SNOMED CT that was supported in process of constructing the knowledge base and semantic reasoning for early detection of AD with the help of multidisciplinary knowledge gathered, inference and reasoning over the underlying knowledge bases which were provided by domain experts. The approach carried out a clinical trial of over 350 patients in 3 hospitals but failed to show the performance of the developed decision support system [7]. Ashraf Mohammed Iqbal et al., (2011) suggested ontology based electronic medical record (EMR) for Chronic Disease Management. The proposed EMR helped to achieve semantic interoperability among healthcare information systems and support reasoning for decision support systems. The ontology ensured the data entry using SNOMED-CT vocabularies, which successfully mapped the proposed ontology onto HL7 RIM to ensure that the clinical messages would be captured. The ontology was implemented in OWL-DL and decision support systems were implemented through description logic representation. The evaluation results showed that the ontology had the capability to capture clinical records [4]. Liping Zheng et al., (2011) proposed a medical aided diagnosis system and maxillofacial diseases (OMD-MADS) for diagnosis of oral and maxillofacial disease. The designed system was based on VTK that was an object-oriented designed, powerful class library for the visualization of graph, image and uses Add-In tree designing method. It mainly included a medical image 3D reconstruction system and oral maxillofacial diseases ontology. In OMD-MADS, users could input medical image, 3D reconstruct, information inquiry and accomplish other operations. This system helped doctor in clinical diagnosis and also taken as an instruction tool in medicine. The system had many functions but some functions were not perfect [10]. Chang-Shing Lee et al., (2011) developed novel fuzzy expert system which concentrated on both data and knowledge problems. A fuzzy layer that included five concepts was used such as fuzzy knowledge layer, fuzzy group relation layer, fuzzy group domain layer, fuzzy personal relation layer and fuzzy personal domain layer. Fuzzy diabetes ontology (FDO) is a knowledge based application which is involved in diabetic decision making .The structure of the FDO constructed the fuzzy concepts and relations. The semantic description of medical staff was simulated based on the FDO and the fuzzy ontology. The accuracy value was obtained between 73.5%- 91.2%. The proposed method works more effectively for diabetes application provided there was only one data set [6].Adem Karahoca et al., (2009) suggested adaptive neuro fuzzy inference system (ANFIS) to estimate diabetes risk which is based on variable inputs such as age, total cholesterol, gender or shape of the body. ANFIS was an estimation method which had fuzzy input and output parameters whereas Multinomial Logistic Regression (MLR) was a non-linear regression method. For the diagnosis of diabetes ANFIS and MLR were executed for comparison. MATLAB v7.1 was used for calculating ANFIS. The dataset consist of 4 variables of 470 subjects. Standard error of ANFIS was smaller than MLR which showed that ANFIS was a better and

faster learning method than MLR [2]. A.M. Aibinu et al., (2010) proposed modeling techniques for automatic diagnosis and classification of diabetes. It used complex-valued neural networks (CVNN) and real-valued neural network (RVNN) which were based on parametric modeling approach. CVNN converts real valued data into complex valued data and learn the relationship between input and output phase with the use of multilayered network. The dataset contains 768 instances with a two class distribution on female subjects .The accuracy obtained was 72.83% [1].Ersin Kaya et al., (2011) suggested a fuzzy rule based system for congenital heart disease. This system consists of database, rule base and reasoning method. Fuzzy rules were generated by the help of medical dataset and weights of fuzzy rules were calculated. The medical dataset contains 297 measurements with 8 condition attributes and 1 decision attribute. Reasoning methods such as weighted vote method and singles winner method were applied to calculate accuracy. The weighted accuracy value is more than the classification of congenital heart disease [8].Ali Adeli et al., (2010) developed a fuzzy expert system for diagnosis of heart disease. The system takes 13 input fields and one output field. Input values were chest pain type, blood pressure, cholesterol, resting blood sugar, maximum heart rate, resting electrocardiography (ECG), exercise, old peak (ST depression induced by exercise relative to rest), thallium scan, sex and age. Output field were integer valued that ranges between 0-4 which denotes the presence of heart disease. The medical dataset contains 76 attributes and 303 cases of patients. This system was developed in Matlab Software. The results were compared with the databases such as V.A. Medical Centre, Long Beach and Cleveland Clinic Foundation. The results obtained were logical and accuracy value obtained was 94 % [3]. Sumathy et al., (2010) proposed a method that diagnosed diabetes based on the risk factors. The proposed technology eased patients who went through a number of medical tests as they could perform the tests by themselves thereby saving time. The suggested system used Artificial Neural Network (ANN) architecture for classification which had a supervised multilayer feed forward network with back propagation learning algorithm. The inputs to the system were the symptoms which appeared in diabetic patients. Around 75 datasets were used for the work. The ANN technique gave better results than other existing techniques [13].

The above survey clearly indicates that the developed systems had number of issues unaddressed and also the prediction rate was not very appreciable. Some authors concentrated only on certain aspects of the occurrence of diabetes leaving few cases unattended. The suggested systems had many disadvantages such as some systems needed dataset of high quality, some gave high prediction rate only if one dataset was used, some were developed only for females, some only for females who were lesser then 21 years of age, some failed to predict the performance of the predicted system, some couldn't differentiate certain types of diabetes with the other types and some functions of the system were not perfect. In order to overcome these shortcomings a new system is proposed that improves the prediction rate and at the same time considers various issues into account.

3. Proposed Framework

The proposed frame work consists of various steps.The first step is the login page where the user gets access to the system by providing their username and password. On completion of the first step, it automatically directs to the second step which is to obtain inputs from the user regarding diabetes. The next step is the analysis of the input parameter. The analyzed parameters are then normalized. For the given dataset fuzzy logic, neural network and CBR are applied as an individual approach. The results obtained are stored in their respective databases. In the last step rule based algorithm is applied to derive the results. Finally diagnosis report is generated. The detailed description of each step is discussed below.

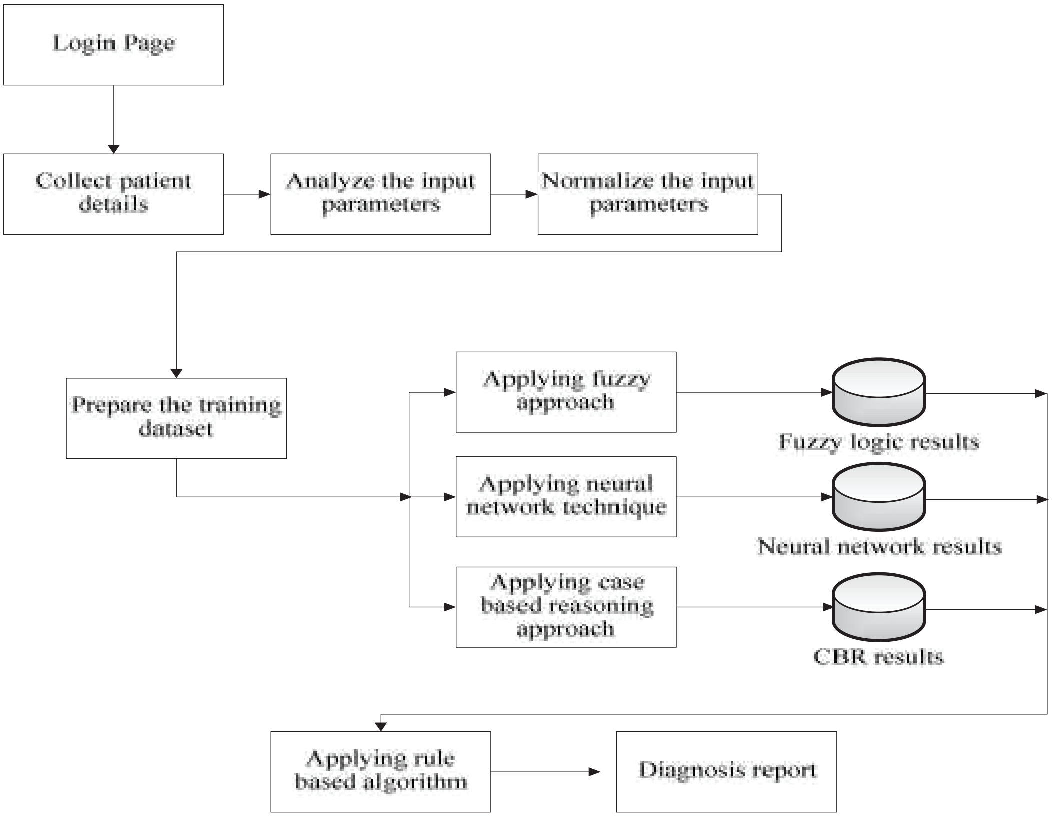

Fig 1: Proposed Framework for Diagnosis of Diabetes using Rule based algorithm.

3.1. Login Page

This is the starting step of the proposed diagnosis system. Here each patient who wants to access the system is given a username and password. For each patient the username and password has to match the one that is already stored in the database. Once the username and the corresponding password match, the patient is given the authentication to use the system for further processes.

3.2. Collect, analyze and normalize the input parameters

Once the first step has been completed, it directs to the next step that is getting the input parameters from the patient which are based on the most common symptoms of diabetes that a patient would be diagnosed with. For this purpose data sets were collected through a survey for the fulfillment of the proposed work. The input parameters that are taken for proposed system include age, gender, background information about the family, having medicines for high blood pressure, found to have high blood glucose during illness, smoking or using tobacco products, quantity of vegetable and fruit intake, physical activity, body mass index, waist hip ratio, frequency of urination, increased hunger, thirst level, poor wound healing, life style (labor class, sedentary work, retired persons and house wife's), gestational diabetes, frequent intake of non-vegetarian food, and itching all over the body. All the collected information is stored in the database for the analysis of the parameters. The sample data sheet is enclosed in the appendix A.

3.3. Applying neural network technique

Back propagation algorithm [13] is used to apply Artificial Neural network technique that uses three layers such as input, hidden and an output layer. In this proposed system the raw information is fed into the input layer which consists of n number of neurons .The hidden layer activities are based on the input layer and it consists of n+1 neurons. The output layer represents either 0 or 1 that denotes 0 when the patient is not affected with diabetes and the value 1 as the patient affected with diabetes.

3.4. Applying Fuzzy concepts

Since fuzzy logic is a mathematical model, its concepts are implemented by preparing the fuzzy set, fuzzy expert system and defining decision rules. The concepts are briefed below.

3.4.1. Fuzzy set

We need to define fuzzy sets for the prediction of diabetes. Fuzzy sets are helpful in describing input parameters. Attributes such as age, gender, background information about the family, having medicines for high blood pressure, found to have high blood glucose during illness, smoking or using tobacco products, quantity of vegetable and fruit intake, physical activity, body mass index, waist hip ratio, frequency of urination, increased hunger, thirst level, poor wound healing, life style (labor class, sedentary work, retired persons and house wife's), gestational diabetes, frequent intake of non-vegetarian food, and itching all over the body are made into a single set which are then fed as input to the system.

The fuzzy set for the proposed system is given as follows:
D={ Age, gender, Background information about the family, Having medicines for high blood pressure, Found to have high blood glucose during illness, Smoking or using tobacco products, Quantity of vegetable and fruit intake, Physical activity, body mass index, waist hip ratio, Frequency of urination, Increased hunger, Thirst level, Poor wound healing, Life style (labor class, sedentary work, retired persons and house wife's), Gestational diabetes, Frequent intake of non-vegetarian food, Itching all over the body} where "D" is the fuzzy set.

For this purpose each input parameter is given a value based on the comparison between a patient whose is diagnosed with diabetes and one whose is normal. The values have been assigned within a range of 0 to 2 based on the information provided for each input parameter. On successful completion of this process the analyzed parameters are then normalized.

3.4.2. Fuzzy expert system

For the operation of a normal fuzzy expert system values are assigned to each attribute of the given input parameter. For example age has 0 points for less than 35 years, 1point for 35 to 45 and 2 points for greater than 45 and for gender, male is assigned 2 points and female is assigned 1 points. Similarly all the other input parameters have been assigned their respective points. The trained data set for few attributes such as age, gender are mentioned below.

Age =\{2,0,0,2,2,2,2,2,1,2,0,1,2,2,2,0\}
Gender =\{2,2,2,2,2,2,1,2,2,2,2,1,1,2,2,2,2\}

3.4.3. Define Decision rule

In order to define the decision rules we need to normalize the attribute values. For this purpose the decision parameters are divided into 3 fuzzy set. Fuzzy sets are “low”, “intermediate” and “high”. For the attribute age fuzzy membership expression is shown in table1.

Table 1: Defining decision rule for input parameter

Input Field	Range	Fuzzy Sets
Age	<35 Years	Low
	35-45 Years	Intermediate
	>45 Years	High

Similarly the attribute fuzzy membership expressions for the other parameters are derived.
Some rules for the developing system are as follows:

Rule 1: If low (age, high blood glucose level, increased urination) then low (Diabetes Mellitus).
Rule 2: if low (physical activity, vegetable intake) then high (Diabetes mellitus)
Rule 3: if low (age, itching, vegetable intake) and high (blood glucose) then intermediate (Diabetes mellitus)

These rules are defined in simple language terms. At the same time these rules result in specific and repeatable (same inputs gives same output) results. To relate the observations to the rules, we must know the degree an observation has membership in the fuzzy set associated with each rule. The following memberships need to be evaluated. For logical "and" operations using fuzzy sets the resulting membership functions are defined as the minimum of the values of the memberships on the component sets. If the rule has "or" operations the maximum of the memberships would be used.

3.5. Applying CBR approach

CBR approach is divided into two phases such as data preprocessing phase and query processing phase. The data preprocessing phase has the following processes knowledge acquisition, creation of ontology and trained dataset. The query processing phase has the following processes such as user, query and CBR result. In the information gathering phase we need to create classes and their hierarchies. Next values are assigned to the respective attributes, thereby creating the trained dataset.

In the query processing phase user requests a query which are then compared with trained dataset of CBR .After comparing related results are obtained.

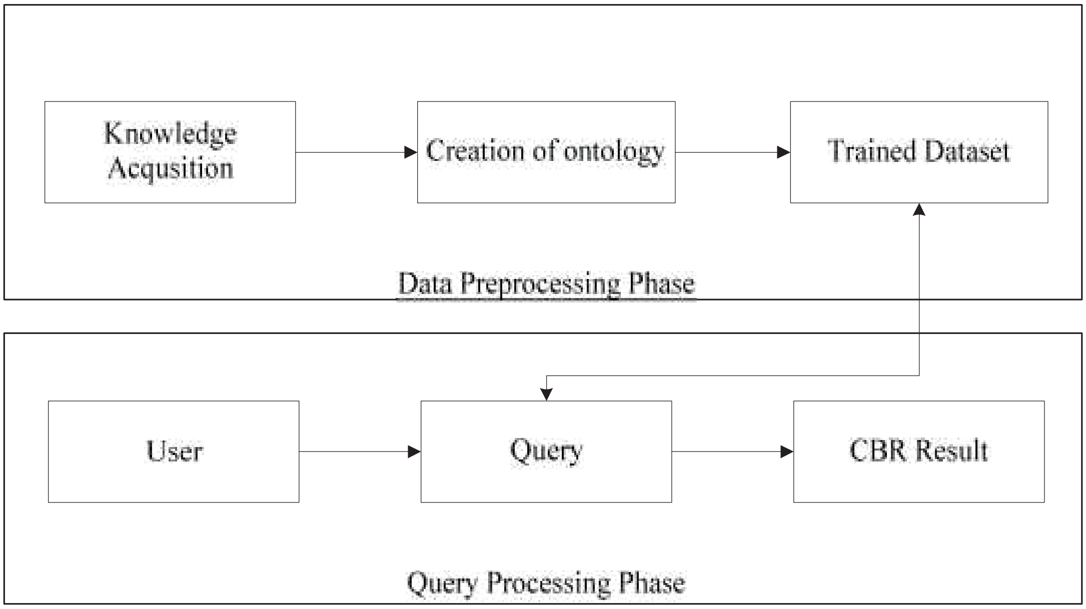

Figure 2: CBR Approach for predicting initial level diabetes

3.6. Applying Rule Based Algorithm

On applying the FNC approach individually we get the prediction rate for the input parameters. The prediction rate can be made more accurate by applying the rule based algorithm. Rule based algorithm is helpful in classifying and predicting both certain and uncertain data. On applying this algorithm to the values obtained we can ensure that the accuracy of prediction will be higher.

3.7. Diagnosis Report

This is the final step in the proposed framework. The expected report generated shows the efficiency of the approaches used and also their accuracy of prediction rate for the diagnosis of diabetes.

4. Conclusion and Future Enhancement

This paper presents a mixture of fuzzy logic, neural network and CBR approaches for prediction of diabetes. The proposed system helps in diagnosing diabetes mellitus. With the literature survey performed and the analysis carried over, the proposed approaches would serve as a better method for diabetes diagnosis. To conclude, we have proposed a new approach called FNC approach for diagnosing diabetes by using newly designed influenced inputs parameters. The system will ease the patients undergoing medical tests for diagnosing this disease without consulting a doctor thus helping the patients to take precautionary measures well in advance. After completing the implementation using the FNC approach, if the accuracy of prediction rate is high we can incorporate this approach for the diagnosis of other diseases.

5. References

[1] A.M. Aibinu, M. J. E. Salami and A. A. Shafie, "Application of Modeling Techniques to Diabetes Diagnosis", IEEE Trans Biomed Engineering, pp 194-198, 2010.

[2] AdemKarahoca, DilekKarahoca and Ali Kara "Diagnosis of Diabetes by using Adaptive Neuro Fuzzy Inference", IEEE Trans, 2009.

[3] Ali.Adeli, Mehdi.Neshat," A Fuzzy Expert System for Heart Disease Diagnosis", IMECS, vol 1, 2010.

[4] Ashraf Mohammed Iqbal, Michael Shepherd and Syed SibteRazaAbidi," An Ontology-Based Electronic Medical Record for Chronic Disease Management", Proceedings of the 44th Hawaii International Conference on System Sciences, pp 1-10, 2011.

[5] Asma A. AlJarullah," Decision Tree Discovery for the Diagnosis of Type II Diabetes", International Conference on Innovations in Information Technology, pp 303-307, 2011.

[6] Chang-Shing Lee, and Mei-Hui Wang, "A Fuzzy Expert System for Diabetes Decision Support Application", IEEE Trans.Cyberrnetics., vol. 41, pp 139-153, 2011.

[7] Eider Sanchez, Carlos Toro , Eduardo Carrasco, Patricia Bonachela , Carlos Parra ,Gloria Bueno and Frank Guijarro,"A Knowledge-based Clinical Decision Support System for the diagnosis of Alzheimer Disease", 13th International Conference on e-Health Networking, Applications and Services, pp.351-357,2011.

[8] Ersin Kaya, Bulent Oran and AhmetArslan," A Diagnostic Fuzzy Rule-Based System for Congenital Heart Disease", World academy of science, Engineering and technology, pp253-256, 2011.

[9] K.Rajeswari and V.Vaithiyanathan," Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network", IJCSNS International Journal of Computer S 126 science and Network Security, vol.11, pp 126-130, 2011.

[10] LipingZheng, Guangyao Li and Junqing Li," The Design and Implementation of Oral Disease Aided Diagnosis System", Journal of Next Generation Information Technology, Volume 2, pp 38-44, 2011.

[11] MostafaFathiGanji and Mohammad SanieeAbadeh," Using fuzzy Ant Colony Optimization for Diagnosis of Diabetes Disease", Proceedings of ICEE, 2010.

[12] R. Vijayamadheswaran, Dr.M.Arthanari and M.Sivakumar," Detection of diabetic retinopathy using radial basis function", International Journal of Innovative Technology & Creative Engineering, vol 1,pp 40-47,2011.

[13] Sumathy, MythiliThirugnanam, Praveen Kumar, Jishnujit T M, K Ranjith Kumar," Diagnosis of Diabetes Mellitus based on Risk Factors", International Journal of Computer Applications, vol 10– No.4, pp 1-4, 2010.

Appendix A

Name : Mr. Kirubakaran

27/M

DIABETES MELLITUS

1) Age

<35 years (0 points)	✓
35 – 45 years (1 point)	
>45 years (2 point)	

2) Gender

Female (1 point)	
Male (2 point)	✓

3) Family background (Whether your parents or brother or sister have been diagnosed as with diabetes)

No (0 point)	✓
Yes (2 point)	

4) Are you currently taking medication for high blood pressure

No (0 point) (O/E BP more than 140/90mm hg – 1)	✓
Yes (2 point)	

5) Have you ever been found to have high blood glucose in a health examination during illness

No (0 point)	✓
Yes (2 point)	

6) Smoking or using any tobacco products

No (0 point)	✓
Yes (1 point)	

7) Vegetable and fruit intake

Every day (0 point)	
Not every day (1 point)	✓

8) Physical activity 30 minutes daily

Yes (0 point)	
No (1 point)	✓

9) BMI (weight/height^2)

Weight (kilograms)	
Height (centimeters)	
<19 (0 point)	
19 – 24 (1 point)	✓
24 – 29 (2 point)	
>29 (3 point)	

10) Waist hip ratio

Waist measurement	
Hip measurement	
<0.8 (0 point)	✓
>0.8 (1 point)	

11) Increased urination, hunger and thirst

No (0 point)	✓
Yes (1 point)	

12) Poor wound healing

No (0 point)	✓
Yes (1 point)	

13) Lifestyle

Labor class (0 point)	
Sedentary work (1 point)	✓
Retired persons and house wife's (2 point)	

14) Gestational diabetes (applicable only for reproductive females)

Weight of the born baby <3.5 kg (0 point)	– NA –
Weight of the born baby >3.5 kg (1 point)	

15) Frequent intake of non vegetarian food (more than 2 times a week), oily item and high calorie food

No (0 point)	
Yes (1 point)	✓

16) Itching all over the body

No (0 point)	✓
Yes (1 point)	

Total points - /24

Patient's blood glucose (GRBS) - 106 mg%