

# Diagnosis of Diabetes Mellitus based on Risk Factors

Prof.Sumathy, Prof.Mythili  
Thirugnanam  
VIT University  
Vellore-632014  
Tamil Nadu, India

Dr.Praveen Kumar  
MBBS  
K.N.Clinic  
Walaja  
Vellore-632513  
Tamil Nadu, India

Jishnujit T M, K Ranjith Kumar  
M S Software Engineering  
VIT University  
Vellore-632014  
Tamil Nadu, India

## ABSTRACT

Diabetes mellitus, in simple terms called as diabetes, is a metabolic disease, where a person is affected with high blood glucose level. Diabetes is a metabolic disorder caused due to the failure of body to produce insulin or to properly utilize insulin. This condition arises when the body does not produce enough insulin, or because the cells do not respond to the insulin that is produced. Blood glucose test is the crucial method for diagnosing diabetes. Also, there have been many computerized methods proposed for diagnosis of diabetes. All these methods have some input values which would be the result of different tests that should be carried out in hospitals. This paper proposes a methodology that aims to ease the patients undergoing various medical tests, which most of them consider as a tedious task and time consuming. The parameters identified for diagnosing diabetes have been designed in such a way that, the user can predict if he is affected with diabetes himself. Back Propagation algorithm is used for diagnosis.

## General Terms

Diabetes diagnosis, Artificial Neural Network

## Keywords

Diabetes, diagnosis, Back Propagation Neural Network

## 1. INTRODUCTION

Diabetes Mellitus has become a common health problem nowadays, which would affect people and lead to various complications like visual impairment, cardio vascular disease, leg amputation and renal failure if diagnosis is not done in the right time. According to W.H.O, by 2025 a total of 300 million of the world population will be affected by diabetes (\* <http://www.diabeteshub.com>). Insulin is a natural hormone secreted in human body to convert sugar, starch and food into simpler molecules, which is then utilized by the cells to generate energy required for daily life. Due to lack of insulin, this conversion is affected and the sugar starts getting accumulated in blood stream and thus increases the blood glucose level, and as a result the person develops diabetes mellitus. The condition is divided into two types such as Type I and II based on the age at onset. Among this, Type I diabetes is usually diagnosed in children and Type II is most common form which affects adults.

Artificial Neural Networks (ANN) is a mathematical model widely used for classification and diagnosis in various thirst areas like effective decision making in medical field, signal processing and so on. Nahla H.Barakat, et.al, utilized support vector

machines [4]. Gregory Hastings, et.al, has used fuzzy estimator logic for monitoring and detecting hypoglycemia in diabetic patients [3]. A five layer fuzzy ontology has been developed in fuzzy expert system by Chang-Shing Lee, et.al, for diabetes decision support application [1]. The framework developed by Pinky Dua, et.al, proposes a method for controlling blood glucose using parametric programming algorithms obtaining optimal insulin delivery rate, which reduces online computation [5]. The proposed work uses architecture of ANN, a supervised multilayer forward network with back propagation learning algorithm has been used. As far as awareness is concerned, many patients have self question like whether I am having diabetes mellitus or hypertension. But how frequent the patient will go and check blood sugar is a question. This would be beneficial for these patients.

## 2. LITERATURE SURVEY

A lot of research has been done using ANN in diagnosing diabetes mellitus and few approaches are discussed below.

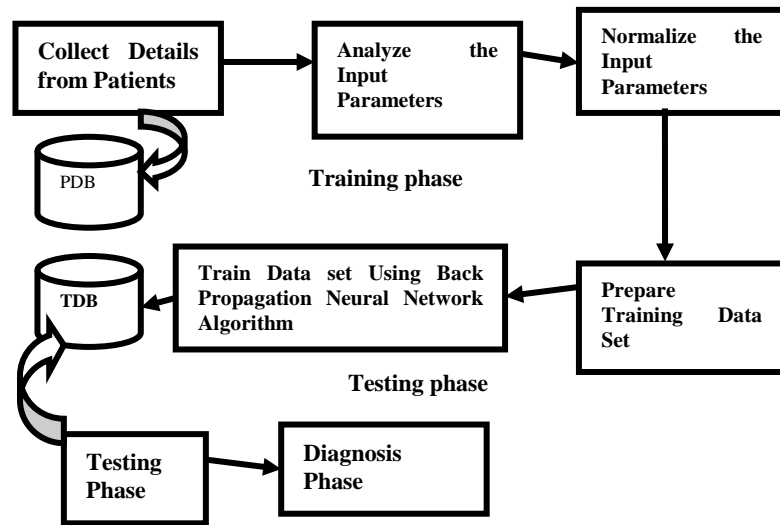
Siti Farhanah Bt Jaffar and Dannawaty Mohd Ali [7] proposed a method for diagnosing diabetes. The diagnosis is accomplished using back propagation neural network algorithm. The inputs to the system are plasma glucose concentration, blood pressure, triceps skin fold, serum insulin, Body Mass Index (BMI), diabetes pedigree function number of times a person was pregnant and age. The biggest challenge was the missing values in the data set. This system has been modified and presented by T.Jayalakshmi and Dr.A.Santhakumaran [8]. They have proposed an idea to overcome the missing values that was not addressed in [7], by constructing the data sets with reconstructed missing values, thus increasing the classification accuracy. [8] Has also proposed an alternate method to overcome missing value by performing data preprocessing, which also speeds up the training process by reducing the actual learning time. Various missing value techniques and pre-processing methods are analyzed. This shows that the results improved and achieved classification accuracy of 99%. [6] Rajeeb Dey, et. al., proposed a method to predict diabetes mellitus using back propagation algorithm of Artificial Neural Network (ANN). The problem of diagnosing diabetes has been treated as a binary classification, i.e., those predicted to be diabetic falls under category 1 and others falls under category 0. The basic architecture of ANN used for accomplishing this classification task is a supervised multilayer feed-forward network with back propagation learning algorithm. The parameters considered in this system to diagnose diabetes are Random Blood Sugar test result, Fasting Blood Sugar test result, Post Plasma

Blood Sugar test, age, sex and their occupation. The performance has been measured in terms of absolute error calculated between network response and desired target. Classification performance achieved using the system is 92.5%. [2] Eng Khaledg Eskaf, Prof. Dr. Osama Badawi and Prof. Dr. Tim Ritchings proposed a method for managing diabetic patients by trying to predict their glucose levels in the near future on the basis of current levels of glucose. The prediction is done using Artificial Neural Network (ANN). Feature extraction procedure was implemented on diabetic blood glucose time series. Blood glucose values of diabetic patients are recorded for 24 hours for about one week with a sampling frequency of 5 minutes. A dynamic model is used as an extraction procedure in order to extract different values from a blood glucose test. ANN was then trained using this knowledge to predict the blood glucose levels of a diabetic patient with reasonable accuracy. [4] Nahla H. Barakat, et.al, proposed a hybrid model for diagnosis of diabetes mellitus by integrating three different data mining techniques using supervised and unsupervised learning algorithm. The inputs were processed using Support Vector Machine (SVM) and rules were extracted using electric approach. Real time data set is taken as input from which rules are formulated to describe relationships between input features and output class labels and thus diagnosis is done based on the rules extracted through electric approach. Of the two rules generated by the electric approach one was found to be inconsistent with generally acceptable medical knowledge in diagnosis of diabetes.

The above survey clearly highlights that ANN technique for diagnosis of diabetes gives much better results than other existing techniques. Also, when considering the inputs to all these systems, there is at least one input value for which the patient should get the help of a doctor or a hospital staff. The proposed system aims to avoid the patients from undergoing blood tests, checking diastolic and systolic blood pressure etc, thus creating a user friendly environment without the need for a doctor or a hospital staff. The inputs designed are based on the symptoms which could appear during the early stages of diabetes and based on the physical conditions.

### 3. PROPOSED FRAMEWORK

Figure 1: Framework for the proposed work



### 4. PROCEDURE FOR DIABETES DIAGNOSIS

#### 4.1.1 Data collection and classification

Around 75 data sets were collected in the survey for the proposed work. The inputs designed for the system are age, gender, family background, taking medication for high blood pressure, found to have high blood glucose in a health examination during illness, smoking or using tobacco products, amount of vegetable and fruit intake, physical activity (30 minutes daily), body mass index, waist hip ratio, increased urination, hunger, thirst, 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. The inputs values have been assigned based on the comparison between two men where one has diabetes and the other a non-diabetic. Based on this comparison, a probability value has been assigned to each of these inputs and these values are fed into the system. These input values range between 0 and 3 based on the condition or physical appearance of that person. The inputs to the system have been designed on the basis of common symptoms of diabetes mellitus and some of the factors which could lead to diabetes like high blood pressure or food habits, etc.

#### 4.1.2 Normalizing data

Before giving the inputs to the network, the input values should be normalized which is based on the activation function used. The activation function used is log sigmoid

$$f(x) = \frac{1}{1 + \exp(-\sigma x)}$$

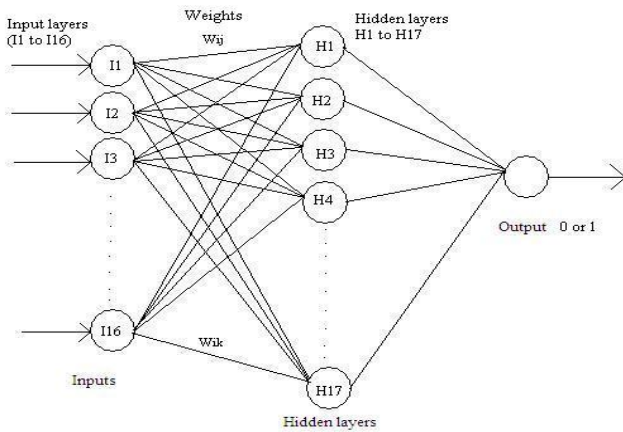
Where  $\sigma x$  input of hidden layer

In the proposed study, the input values have been normalized between the range 0 and 1. This is done by dividing the entire input values by 3.

### 4.1.3 Network Architecture

The network architecture refers to the number of layers, the number of nodes in each layer, and the number of hidden layers in the network. In the proposed system, 16 inputs have been chosen for the diagnosis (fig.2). So there will be 16 input layers for the network. N+1 hidden layer have been designed for the network. The output layer of the network has one neuron which gives the value either 0 or 1. The value 0 represents the user is not affected with diabetes and the value 1 represents the user is suffering from diabetes. The inputs designed are based on the symptoms of diabetes as mentioned above in section 3.1.

**Figure 2: Proposed Neural Network Architecture**



### 4.1.4 Back Propagation Algorithm

Back propagation is a systematic method for training multi layer artificial neural network. It is a multi-layer feed forward network that uses extent gradient-descent based back propagation rule. During the training phase, the training data is fed to the input layer. The data is propagated through the hidden layer and then to the output layer, which is called forward pass. During this phase, each node in the hidden layer gets input from all the input layer nodes, which are then multiplied with the randomly assigned weights and then summed up. Similarly the output layer node receives inputs from all nodes of the hidden layer, which are then multiplied with the randomly assigned weights and then summed up. This forms the output of the output layer.

The output values are now compared with the target value, and the error between the output value and the target output value is calculated and it is propagated back to the hidden layer, which is called backward pass. This error is used to update all the weights between the input and hidden layers as well as the weights between hidden and output layers. The training continues until the error is minimized.

In the proposed work, sixteen input layers have been designed for the network and the weights have been assigned randomly. The

randomly assigned weights are in the range between 0 and 1. The input to each hidden layer is calculated by

$$Y_{in} = \sum \text{weight} * \text{input value}$$

Output of the hidden layer is calculated by using the activation function

$$f(x) = \frac{1}{1 + \exp(-\sigma x)}$$

Error rates are calculated using

$$\text{Error} = f(x) * [1 - f(x)] [\text{target value} - f(x)]$$

Which is then back propagated to the hidden layers, weights are updated and the process continues until error rate is minimum, which forms the trained data.

The new weights are calculated based on,

$$\text{Weight}_{\text{new}} = \text{weight} + \eta * \text{error} * f(x)$$

Where,  $\eta$  is the learning rate.

After completing above process, we guarantee maximum accuracy and minimum error rate.

## 5. CONCLUSION AND FUTURE WORK

With the literature survey performed and the analysis carried over, the proposed methodology would serve as a better tool for diabetes diagnosis. To conclude, we propose a new method of diagnosing diabetes by using newly designed inputs parameters. Back propagation algorithm has been used for diagnosis with a newly designed set of inputs, i.e. based on the basic symptoms which appear in diabetic patients and the physical conditions. The proposed work also reduces the cost for different medical tests, also help patients to take precautionary measures well in advance. With the successful implementation of this proposed work, the same method could also be applied in diagnosing other diseases like, coronary artery disease, hypertension etc.

## 6. REFERENCES

- [1] Chang-Shing Lee, Mei-Hui Wang, "A fuzzy expert system for diabetes decision support application", IEEE transactions on systems, man and cybernetics 2010.
- [2] Eng Khaled Eskaf, Prof. Dr. Osama Badawi and Prof. Dr. Tim Ritchings, "Predicting blood glucose levels in diabetes using feature extraction and artificial neural networks"
- [3] Gregory Hastings, Nejhddeh Ghevondian, "A self-organizing estimator for hypoglycemia monitoring in diabetic patients", 20<sup>th</sup> annual international conference of IEEE engineering in medicine and biology society, Vol. 20, No 3, 1998.

- [4] Nahla H. Barakat, Andrew P. Bradley, and Mohamed Nabil H. Barakat, "Intelligible support vector machines for diagnosis of diabetes mellitus", IEEE transaction on information technology in Biomedicine, Vol. 14, No. 4, July 2010
- [5] Pinky Dua, Francis J. Doyle, and Efstratios N. Pistikopoulos, "Model based blood glucose control for Type I diabetes via parametric programming, IEEE International Transactions on biomedical engineering, vol.53, no.8, August 2006.
- [6] Rajeeb Dey and Vaibhav Bajpai and Gagan Gandhi and Barnali Dey, "Application of artificial neural network technique for diagnosing diabetes mellitus", 2008 IEEE Region 10 Colloquium and the Third ICIS, Kharagpur, INDIA December 8-10.
- [7] Siti Farhanah Bt Jaafar and Dannawaty Mohd Ali, "Diabetes mellitus forecast using artificial neural networks", Asian conference of paramedical research proceedings, 5-7, September, 2005, Kuala Lumpur, MALAYSIA
- [8] T.Jayalakshmi and Dr.A.Santhakumaran, "A novel classification method for classification of diabetes mellitus using artificial neural networks". 2010 International Conference on Data Storage and Data Engineering.