Medical Information System for Classification of Diabetes Mellitus Using Layered Neural Network

R. Sonia*, B. Uma Maheswari, M.P. Rajakumar**, J. Ramya****

**Department of Computer Applications,*
B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai 600 048,
Tamilnadu, India (e-mail: sonia.j25@gmail.com)
***Department of Computer Science and Engineering,*
St. Joseph's College of Engineering, Chennai 600 119, Tamilnadu, India.
(e-mail: mahespal2002@gmail.com, rajranjhu@gmail.com, ramsharsha@gmail.com)

Abstract: Diabetes Mellitus, the chronic illness condition is correlated by high blood sugar levels and troubled conditions rapidly increasing over the decades. Type 1 Diabetes occurs due to less secretion of insulin by beta cells. But beta cells produce required insulin in Type 2 diabetes where the body cannot use it. The significant factor for diabetes is insulin deficiency. Diabetes during the trimesters of pregnancy is known as Type 3 diabetes called gestational diabetes. Gestational diabetes automatically disappears after the delivery or may continue as type 2 diabetes. The research work proposes an information system for classifying all three types of diabetes Miletus using a layered neural network. The medical information system contains two major phases, Training phase, and testing phase. The architecture of the layered neural network improves the efficiency of the classification. Also, the experimental analysis and performances of diabetes classifications are obtained using the confusion matrix through sensitivity, specificity, and accuracy. The highest value of specificity and sensitivity achieved for the proposed layered neural network is 0.96 and 0.98, respectively. The experimental results of diabetes mellitus classification exhibit 98% of high accuracy.

Keywords: diabetes, gestational, classification, layered neural network, sensitivity, specificity, accuracy.

1. INTRODUCTION

Diabetes is a metabolic illness associated with high blood sugar(hyperglycaemia) associated with symptoms and other health complications like the failure of organs (American Diabetes Association: Standards of medical care in diabetes, 2006). Insufficient production of insulin or does not use the generated insulin effectively, which leads to diabetes. Computer programs developed helps as a tool for physicians to design an Information system for the diagnosis of diseases based on clinical data (Short life, 1987). The three types of diabetes are compared based on genetics and diagnostic criteria. Also on the region's epidemic, nature is reviewed with the highest occurrence of diabetes Miletus in adults (Kharroubi and Darwish, 2015). American diabetes association's diabetes is diagnosed based on pre-diabetes factors IFG or IGT. Also, diabetes is diagnosed when the glucose level is more than 200mg/dL based on HbA1c, OGTT, and FPG (Kharroubi and Darwish, 2015; American Diabetes Association: Standards of medical care in diabetes, 2006).

Type 1 diabetes which is dependent on insulin is otherwise called as juvenile diabetes. The human body is not able to produce adequate insulin by β-cells it leads to Type 1 diabetes. Also, the beta cells in this category keep varying in infants and adults (American Diabetes Association: Standards of medical care in diabetes, 2019). When the glucose levels are not regulated properly to all parts of the body, other complications like heart disease, nerve problems, and kidney-related complications might occur. Type 1 diabetes usually occurs under the age of 30, and the patients are insulin-dependent, by pumping insulin for their body (Janez et al., 2020; Lian et al., 2018, American Diabetes Association: Standards of medical care in diabetes, 2019).

Type 2 diabetes mellitus is non-insulin-dependent, which otherwise identified as adult diabetes. It occurs due to loss of secretion of insulin by β cells. It increases due to genetics, obesity, and the different lifestyle of peoples. Furthermore, type 2 diabetes occurs in the middle age (Olokoba et al., 2012). Also, it happens before gestational diabetes in ladies and ethnic populations. Furthermore, it affects children and adolescents (American Diabetes Association: Classification and Diagnosis of Diabe tes, 2015).

Gestational diabetes prediction is dependent on the maternal factors during II trimester, III trimester of pregnancy, and biomarkers during the gestation period (Surabhi et al., 2011). Recent study groups have analysed that the pregnant women are under high risk of diabetes on different ethnic races annually (American Diabetes Association: Detection and Diagnosis of GDM, 2013).

Modern technology stores a massive amount of data hence performed using machine learning (Kononenko, 2001). Doctors analyse patient's disease interactively using clinical metrics like BP, glucose, and temperature. Treatments after the iterative analysis and refinement are performed. Furthermore, Artificial Intelligence plays an important role in fuzzy-based classification and diagnosis of disease using neural networks.

Computer-aided. Comprehensibility is a significant factor that makes the situation worse, and the Artificial NN ensemble is used for better diagnosis of disease in the medical world (Karabegovic et al., 2006; Moein et al., 2008; Steimann and Adlassnig, 2005).

The intelligent system plays a dynamic part in medical analysis using Machine learning techniques, which helps the experts to improve accuracy and efficiency (Zhi-Hua and Yuan, 2015). The knowledge fed into an artificial neural network that extracts the rules using fuzzy. ANN ensemble can also solve rule-based diagnosis (George et al., 2001; Andrews et al., 1995; Cunningham et al., 2000).

This study intends to propose a medical information system for diagnosing gestational, type 1, and type 2 diabetes using a layered neural network. The input data (patient data) contain four primary patient's category, such as normal, type 1 diabetes, type 2 diabetes, and gestational diabetes. Initially, the relevant attributes are identified through the help of the attribute selection process. The three neural networks are trained individually in a layered manner, initially with normal and type 1 diabetic patient, then with normal and type 2 diabetic patients and finally, with healthy and gestational diabetes. A trained layered neural network used to perform the diagnosis process whether the patient belongs to normal, type 1, type 2, or gestational categories.

The main contributions of the paper are:

- Diabetes classification technique is applied for three types of diabetes, like type1, type2, and gestational diabetes.
- A specialized dataset is created for the experimentation collected over the internet.
- Classification is performed using a layered neural network.

In the 2nd section, the review of the associated work is discussed. The 3rd section and 4th section include details about the datasets and proposed layered neural network for classification. The 5th section consists of the result and discussion about the proposed method, and the final section includes a conclusion of the proposed method.

2. RELATED EARLIER WORKS

The literature review presents a few related kinds of research that automatically work for the diagnosis of diabetes. Here, we review the works based on machine learning available in the literature recently.

(Hang et al., 2007) proposed a classification technique for diabetic patients of type 2 category**.** The patient's clinical datasets for the classification are collected from UCHT, and knowledge discovery was performed using data mining techniques. A feature selection method defined uniquely produces better efficiency. Finally, classification techniques such as C4.5, etc., predict type 2 diabetic patient's conditions.

(Abhari et al., 2019) focuses on machine learning algorithms, which is used for the classification of advanced and pre-diabetes. The techniques proposed for type 2 diabetes are SVM, ANN, etc. SVM outperforms for better classification and predicting type 2 diabetes.

(Deepti and Dilip, 2018) designed a model for predicting diabetes using three classification algorithms, and the Naïve Bayes algorithm produces the highest accuracy.

(Wang et al., 2017) proposed a novel classification method for diabetes diagnosis and classification. CGM is used to record and monitor the patient's glucose level at specific intervals. Clinical datasets are collected from the People's hospital of China and performed analysis on the Chinese ethnic race. Initially, the 17 features are extracted through GSM and later variant algorithms of AdaBoost is used as a novel indicator to diagnose and classify diabetes types with experimental results of 90.3%. Performance indicators such as ACC and MCC are used to evaluate the results.

(Allalou et al., 2016) determines the gestational diabetes, which is a significant parameter to develop type 2 diabetes after the delivery. The significant risk factors include an increase in BMI during pregnancy, ethnic race, and age. A comparative analysis is achieved on GDM and non-GDM. Also, a Statistical analysis is done using ANOVA and SPSS. The predictive model was constructed using WEKA software with techniques such as Naïve Bayes classifier and J48. Clinical related parameters were analyzed using Pearson correlation coefficients. This work shows the metabolomics signature transformation from gestational diabetes to type 2 diabetes.

(Yu et al., 2010) suggest Support Vector Machine, a machine learning algorithm for classification diabetes. SVM uses large multidimensional space for transforming the input by using the kernel functions. Various 14 Attributes are used for the two types of classification schemes based on diagnosed diabetes, undiagnosed diabetes, pre-diabetes, and no diabetes. Performance analysis is carried out by ROC and other cross-validation functions. RBF and linear kernel functions outperform best for classification schemes.

(Sneha and Gangil, 2019) proposed an analysis on diabetes mellitus prediction at the early stage using machine learning algorithms. Statistical analysis is achieved by using different health organizations. Supervised learning algorithms used for classifications and comparative study was performed by using the attributes. A new modified approach was applied by selecting optimal features, and few attributes were eliminated due to less correlation value and mapped with the algorithms. Hence Decision tree and Random forest algorithm outperform with the high specificity more than 98%.

(Lukmantoa, R et al., 2019) proposed an early diagnosis of diabetes using Fuzzy SVM. A dataset with eight attributes was collected from PID. Data pre-processing applied to all eight attributes, and only six attributes are considered for diagnosis. Feature selection F-score is used to optimize the best features by removing the inappropriate features. Finally, the classification is performed using Fuzzy SVM for the prediction of diabetes. A Logical regression model of machine learning is proposed for the analysis HbA1in diabetes patients with clinical datasets for type II diabetes.

(Taghiyev et al., 2019) performed a study on analysis of type 2 diabetes using logical regression and datamining methods

3. DATASET PREPARATION

The main area to be concentrated on the research is to build the dataset for the three types of diabetes. For evaluating the proposed classification algorithm, two datasets are constructed based on all three categories of diabetes.

3.1 Type 1 diabetes

Type 1 diabetes, juvenile diabetes, which arises in both children and adolescents. The food taken in form carbohydrate absorbs energy and retain the glucose. The pancreas is used to generate a vital hormone called insulin. This insulin transforms glucose into body cells. The human immune system attacks beta cells, which leads to type 1 diabetes. The main objective of the proposed approach is to extract a set of data, which contains diabetes infected patients. The dataset for the classification of types of diabetes is obtained from the UCI online library (Dua and Graff, 2019). Data repository has a type 1 diabetes dataset by eight attributes and class labels ranging from 0 to 1 which is given below in the Table 1. The data selected is full of the female to relate the type 1 dataset with gestational diabetes, since gestational diabetes affects only during pregnancy. The attributes are the symptoms and causes of type 1 category. The dataset contains totally 768 datasets with 500 record of positive class and 268 negative class for diabetes.

Table 1. Attributes of Type 1 Diabetes.

Attributes	Units
Number of pregnancies	-
Glucose	mg/dL
Diastolic BP value	mmHg
Skin thickness	mm
Insulin level	mu/ml
BMI	Kg/m2
Family history of diabetes	-
Age	-

3.2 Type 2 diabetes

Type 2 diabetes is non-dependent on insulin. The needed insulin is formed from beta cells in the pancreas, but it's not used by the body or doesn't provide the required insulin. The risk parameters related to type 2 diabetes are age, hereditary, and ethnicity. Type 2 diabetes is diagnosed using the average blood glucose level tested, Fasting plasma, and OGTT. Blood sugar causes health complications in the heart, kidney, eyes, brain, pregnancy, etc. Thus, according to the proposed approach, the dataset regarding type 2 diabetes is also collected. The attribute set regarding type2 diabetes is listed below in Table 2.

Type II diabetes dataset consists of 19 attributes contributed by Dr. Schorling from virgina school of medicine was used. The above dataset consists of 1046 datasets from which only 403 datasets were considered based on the attributes. The type II dataset are compared with type I dataset and finally 8 significant parameters are chosen for the classification of diabetes which is given in the table 2. As type II diabetes risk is associated with hypertension the parameters Blood pressure and cholesterol plays a vital role and has been selected from the attribute list.

Table 2. Attributes of Type 2 Diabetes.

Attributes	Units
Number of pregnancies	-
Cholesterol level	mmol/L
HDL	mg/dL
Cholesterol/ HDL ratio	-
Age	-
BMI	Kg/m2
Family history of diabetes	-
Gender	-
Blood pressure	mmHg

3.3 Gestational diabetes

The above type 1 and type 2 data sets are considered and only female subjects are considered for gestational diabetes. Gestational diabetes can be considered as type3 diabetes, but unlike the other type of diabetes, it only happens for pregnant females. Gestational diabetes is caused by high sugar concentration in blood, high blood pressure, excess amount of amniotic fluid, etc. In the proposed approach, the datasets for gestational diabetes are extracted from type I and type II datasets after analysing similar attributes. The attribute regarding gestational diabetes is obtained concerning the characteristics of type I and type II.

4. CLASSIFICATION USING LAYERED NEURAL NETWORK AND THE PROPSED ATTRIBUTE SELECTION METHOD

The primary intention of our research work is to construct a medical information system for classifying three types of diabetes mellitus using a layered neural network. The medical information system for diabetes classification contains two significant phases, Training and testing phases. The patient data is given as input to the training phase includes four major classes of patient data. Initially, the relevant attributes are identified through the help of the attribute selection process. Further, the three neural networks are trained individually in a layered manner like the first one with normal and type 1 diabetic patient, the second one, with normal and type 2 diabetic patients and the third one, with normal and gestational diabetes. The trained layered neural network is then used to perform the diagnosis process whether the patient belongs to normal, type 1, type 2, or gestational categories. The input data is given to the attribute selection process during the testing phase, and the selected attributes are given as input to the neural network 1 to identify whether the patient for type 1 diabetes. If the decision is healthy, then the same data is fed to the next layer of the neural network to classify the patient for type 2 diabetes. Else, the corresponding information is supplied to neural network 3 to classify whether the data is related to gestational or normal.

4.1. Proposed techniques for attribute selection

In earlier sections, we have discussed various types of diabetes and the attributes associated with it. Based on this data, the proposed approach creates a separate set of datasets, which is related to the characteristics of all three types of diabetes. Even though there are several attributes for three types of data, the proposed approach uses only the most inter-related attributes.

Discretization: A discretization of attributes is applied to categorize them based on the relationships. To discretize the attributes, their count and range are calculated and categorized mainly into four groups. All classes of diabetes attributes are sorted in ascending order based on the minimum, and maximum value for every attribute as diabetes diagnosis is performed for three types type1, type 2, and gestational diabetes. The interval is predicted on the basis of vector sorted in ascending order. Finally, the data are converted to a discretized value which falls under this interval.

4.1.1. Class association rule

Attribute weightage: For the classification process, the different types of diabetes are considered by the proposed method. The main features associated with the classification process are the attributes that define the characteristics of each type of diabetes. The proposed method put forward a technique in selecting the attributes based on their weightage. An attribute weight calculation method is adopted based on the association of each attribute with the classes derived from the dataset.

The attribute weightage calculation process mainly considers two data matrices, which are formed by the attributes, patient id, and class values. The attributes and the patient list form the initial matrix, and the second matrix is created by the class labels and the attribute lists.

$$\begin{matrix} p_1 & \cdots & p_n \\ a_1 & \cdots & a_n \\ a_n & a_n & a_n \end{matrix} \quad \Rightarrow \quad \begin{matrix} c_1 & \cdots & c_n \\ a_1 & \cdots & a_n \\ a_n & a_n & a_n \end{matrix} \tag{1}$$

The two matrices used for attribute length calculation is given in the above equation (1). The attribute weight calculated by finding the frequency of each attribute in both matrices. The relevance of each attribute regarding the matrices is calculated independently and then dependently. The above process fetches out the importance of an attribute regarding their class. The weight of the attribute is derived mainly by based on the classes, i.e., how an attribute is associated with a particular category, which will help the proposed method to classify the four different categories of diabetes. The initial process of the attribute weight calculation phase is the comparison between the two matrices. Initially the attribute frequencies regarding the classes are calculated for finding the importance of the attributes. Once the class-based frequency is calculated, then the attribute weight is analysed based on the first matrix. The weight of the attribute is associated with its presence and absence in both matrices.

$$c_1 \colon a_1 \;\; \Rightarrow \;\; a_1 \; in \; c_1 \tag{2}$$

The above expression shows the relevance of attribute a1 in class1 based on the presence of attribute a1 in the data matrix one. Similarly, the presence of attributes is calculated based on their appearance in matrix one. The c1: a1 value gives an integer value, which is obtained as the count of a1 present in matrix one in relevant class c1.

$$c_1 \colon a_1 \;\; \Rightarrow \;\; a_1 \; in \; c_1 = Count_{c1}(a_1) \tag{3}$$

The presence and absence of the attribute are calculated consecutively for the calculation of the attribute length. Thus, from the above-calculated values, the weight of the attribute a1 is calculated as per the equation (4),

$$weight(a_1) = \frac{Count_{c1}(a_1)}{Count_{c1}(a_1^{\omega})} \tag{4}$$

The weightage value provides the relevance of Attribute a1 based on class1. Similarly, the weight of each attribute based on each class is calculated, and the attributes having the highest weightage are chosen for improving the efficiency of the proposed classification method.

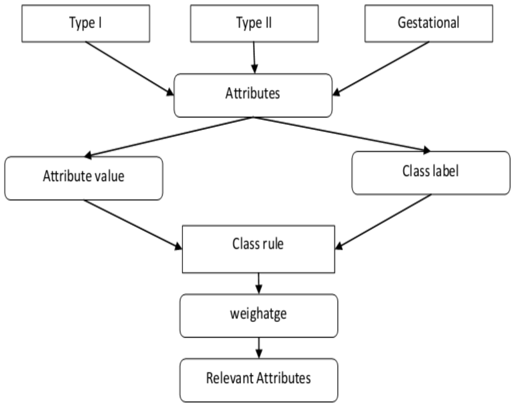

Fig. 1. Attribute Extraction.

4.1.2 Diabetes dataset_1 and dataset_2 use same procedure to calculate attribute strength

1. dataset_1 is given in datast_1.xls file

2. The dataset_1 is normalized
   (i.e., converted to [0 - 1] range) by dividing each element with the maximum value of that column.

3. dataset_1 is discretized by assigning value 1 to those elements for which the value is between 0 and 0.25, value 2 to those elements for which the value is 0.25 to 0.5, value 3 to those elements for which the value is between 0.5 and 0.75 and value 4 to those elements for which the value is between 0.75 and 1.

4. To calculate attribute strength.
   (i) We form a matrix of the following dimension
      number of rows = number of discrete levels X
      number of types of diabetes
      number of columns = number of attributes

   (ii) (a) numerator = find out the number of times value 1 occurred in first column and 1 occurred in same row of last column (i.e., type column).
      (b) denominator = find out the number of times value 1 occurred in the first column and 1 did not occur in same row of last column.
      matrix (1,1) = numerator/denominator
      (c) numerator = find out the number of times value 1 occurred in first column and 2 occurred in same row of last column.

(d) denominator = find out the number of times value 1 occurred in the first column and 2 did not occur in same row of last column.

matrix (2,1) = numerator/denominator

(e) numerator = find out the number of times value 1 occurred in first column and 3 occurred in same row of last column (i.e., type column).

(f) denominator = find out the number of times value 1 occurred in the first column and 3 did not occur in same row of last column.

matrix (3,1) = numerator/denominator

(g) numerator = find out the number of times value 1 occurred in first column and 4 occurred in same row of last column

(h) denominator = find out the number of times value 1 occurred in the first column and 4 did not occur in same row of last column.

matrix (4,1) = numerator/denominator

(i) numerator = find out the number of times value 2 occurred in first column and 1 occurred in same row of last column

(j) denominator = find out the number of times value 2 occurred in the first column and 1 did not occur in same row of last column.

matrix (5,1) = numerator/denominator

(k). numerator = find out the number of times value 2 occurred in first column and 2 occurred in same row of last column.

(l). denominator = find out the number of times value 2 occurred in the first column and 2 did not occur in same row of last column.

matrix (5,1) = numerator/denominator.

(iii) Repeat the procedure in step (ii) taking second column instead of first column to fill the second column of the matrix.

5. Find out the maximum value in each column of the resulting matrix. This is the attribute strength of that attribute which correspond to that column.

6. Discard two columns for which the attribute strength is the lowest.

7. Take 80% of the rows at random from the resulting dataset for training.

8. Take the remaining 20% rows of the dataset for testing.

4.2 Diabetes classification

The proposed approach concentrates on classifying the different types of diabetes by associating different attributes possessed by the different datasets. As discussed in the above section, the proposed method creates four sets of data by associating the attributes. The main objective of those datasets is the reduction of an attribute. The most appropriate attribute among them is selected according to their priority and frequency. Once the characteristics are optimized, then the process of classifying the different types has to be subjected. The proposed approaches consider neural network method for classification. The classification phase includes mainly two processes, the training phase, and the testing phase.

4.2.1 Classification of diabetes types using layered neural network

The significant objective of the proposed approach is to categorize diabetes types based on the characteristics. The characteristics are marked as attributes in the proposed approach. As defined above the proposed approach uses three different sets of neural networks based on the four different data types. The proposed approach adopts four different sets of data from the data sets extracted from various data repositories, which are mentioned in the above sections. Like the entire neural network, we also have the training phase and testing phase. There are three groups of data that are being considered for the neural network engines are normal_type1, normal_type2, and normal gestational. The proposed method uses a typical neural network for effective diabetes classification. A layered neural network is subjected here to categorize the three types of different diabetes.

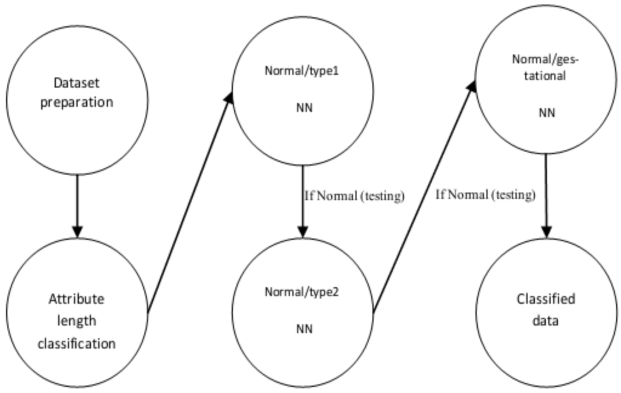

Fig. 2. Proposed layered Neural Network (NN).

The Fig.2 illustrates the overall structure of layered neural network to classify the different types of diabetes. The attributes obtained after the attribute weight calculation phase is given for training with the three neural networks. After the training phase, the data to be tested are subjected initially to the normal / type1 NN. If normal data has resulted, then it is subject to the normal / type 2 NN, if again normal, the data is subject to the normal/gestational diabetes. The layered architecture of the neural network will provide improved efficiency for the classification process.

4.2.2 Training phase

The primary phase in the classification process is the training phase, where all the diabetes data groups are trained for finding their relevant class. The Multilayer perceptron consists of input layer, which is a known value, a hidden layer, which is an unknown value and an output layer, which is a perceived value. The input layer comprises of the eight attributes of Pima Indian diabetes dataset. The output layer of the training phase contains two values for each of the three neural networks. The class values of each of the four data types will range as 1 for normal patient without diabetes, 2 for type1 diabetes patients, 3 for type2 diabetes, and 4 for gestational diabetes.

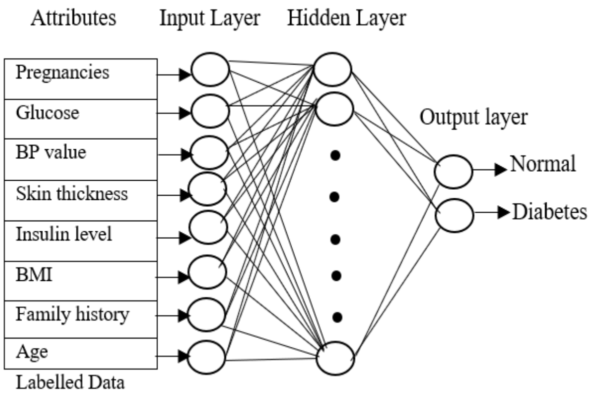

Fig. 3. Neural Network Training phase.

Fig. 3 shows the neural network with 8 attributes of type1 diabetes and the same network can be used by the proposed approach for classifying the diabetes for type 2 diabetes with pre-processed attributes specified in table 2. The main objective of the training phase is to calculate generalized values for the hidden layers. All the layers are connected to each other to form a network. The neuron is activated by transmitted signal and also has an activation function which act as a threshold. The weights are found through the activation function. The hidden layer gets stable only when the neural network derives the least error value from the training dataset. A fast-supervised learning algorithm scaled conjugate gradient (SCG) reduces the error value in the network. The SCG will continue up to the level in which the hidden layers get stable value. Similarly, all three neural networks are trained with the related diabetes dataset and all the hidden layer values are calculated.

4.2.3. Testing phase

The testing phase concentrates on extracting the different types of diabetes from an unknown set of attributes and data. The proposed approach uses a sequential testing procedure for classifying the data based on the different diabetes types. The patient's known data train the layered neural network used in the testing phase. The trained neural network uses three different layers, which are similar to the neural training network, but with minor changes. The input layer in the testing phase is identical to the training phase. The hidden layers go through significant changes because it obtains the known values from training phase. On the other hand, the output layer is empty here in the testing phase.

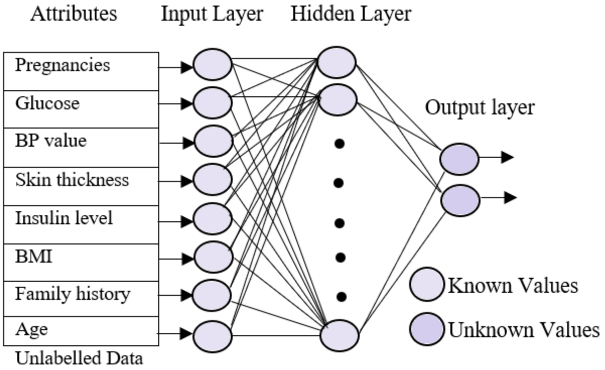

Fig. 4. Neural Network Testing phase.

Fig. 4 represents the processing of the testing. The aim behind the testing phase is to classify the unlabelled data into different groups based on the diabetes types. The testing phases use three neural networks, which are trained for different types of diabetes, namely normal, type1, type2, and gestational diabetes. Initially, the unlabelled data is tested for type1 diabetes using the first neural network. The resultant data which is not listed under diabetes, i.e., the normal data is selected and subjected to the second neural network for finding the possibility of diabetes type II in the data. Gestational diabetes is tested separately during the gestation period because it affects only pregnant ladies.

5. RESULT AND DISCUSSIONS

The experimental diagnosis of the research work for the diabetes classification method is discussed. The analysis is performed based on two different types of diabetes dataset with attributes collected over the internet. The experiments are implemented in MATLAB and the toolkit used is neural network toolbox. To reduce the error value Scaled conjugate gradient is used. We have used MATLAB’s trainscg with mean square as the performance function. Fuzzy logic tool box is used for the classification using sugeno fuzzy Inference system.

5.1. Dataset description

The proposed approach uses two different datasets for the classification process according to the various categories of diabetes types, such as type 1, type 2, and gestational diabetes. Data sets used for the type 1 diabetes classification is obtained from UCI machine learning repository. This Pima Indian Diabetes dataset is formerly from the National Institute of Diabetes and digestive and kidney diseases, which contains data of 768 diabetes data set with 500 positive subjects and 268 negative subjects. Dataset 1 includes three different diabetes types with all common attributes. Whereas dataset 2 includes three diabetes types with unique attributes for classification. The type 2 diabetes dataset of 1046 data with 19 original attributes contributed by Dr. schorling from virgina school of medicine is considered. As gestational diabetes is also included in classification only 403 datasets are taken with few attributes for both type 2 and gestational diabetes by comparing it with the 8 attributes of type 1 diabetes.

5.2. Neural network parameters, Fuzzy parameters and training algorithm

The neurons are activated through a transmitted signal and also has an activation function which acts as a threshold. The training algorithm used for multi-layer perception neural network (MLPN) is scaled conjugate gradient algorithm (SCG). The learning rate are set so that the slopes do not diverge. The neural networks were iterated for epochs with learning rate. From the table 3, we can identify those five hidden layers are utilized along with input and output layer. The parameters related to neural network and fuzzy are given in the Table 3 and 4 respectively. Also, Table 4 gives the fuzzy parameters taken for fuzzy classifier. Here, we have used genfis classifier for fuzzy classification as there are more than six inputs.

5.3 Performance Analysis

This section includes the performance analysis of the proposed classification of diabetes types. The section uses the above listed parameters namely specificity, sensitivity and accuracy. The evaluation parameters are calculated by testing it with the collected datasets. The performance evaluation is conducted on two datasets which are processed according to the proposed approach. The data sets are tested using the methods of the proposed approach for analysing the specificity, sensitivity and accuracy of the proposed approach. The performance evaluation conducted in this section also uses a comparison with an existing work, which is a classification process based on fuzzy method. The classification is based on three classes; class 2 represents the type 1 diabetes; class 3 represents the type 2 diabetes and class 4 represents gestational diabetes. The analysis based on the proposed classification method is shown in the following graphs.

Table 3. Neural Network parameters.

Neural Network	Epoch	No of iterations	Error	Training algorithm	Number of hidden layers	Nodes
NN1 (normal / type1 NN)	88	108	0.090677	scaled conjugate gradient algorithm	5	[10:50:25:5:1]
NN2 (normal / type2 NN)	88	108	0.018476	scaled conjugate gradient algorithm	5	[10:50:25:5:1]
NN3 (normal / gestational NN)	198	218	0.001328	scaled conjugate gradient algorithm	5	[10:50:25:5:1]

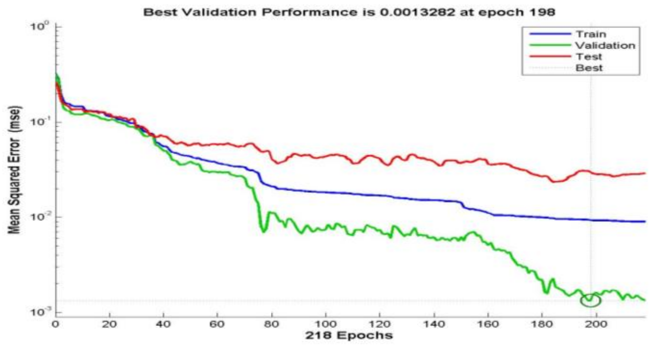

Fig. 5. Best validation performance.

The networks performance is measured by Mean square error. The error values are taken they are squared and their mean value is calculated. The above fig. 5 shows the lower value is obtained in epoch 198 with the best performance validation of 0.0013 for gestational diabetes. Also, table 5 given below shows the Comparison of diagnosis and classification results with performance accuracy by different authors.

The Fuzzy inference system maps the given input to output using fuzzy rules and operators. In our proposed system the logical operator AND and OR method is used. The main workings of the system are fuzzification, Inference and defuzzification. There are two components of FIS, the Mamdani and Sugeno type. We have chosen the Sugeno model for our work as the output depends on the input variable. As we have 8 multiple attributes as input the gaussian membership function is used with Crisp value for all the variables individually is fed into the FIS editor. Genfis model is used when there are more than 6 inputs. To define the membership function genfis uses the subtractive clustering and grid partitioning. The genfis returns single output sugeno fuzzy inference system. The sample model with three input and single output, rule evaluation and rule viewer for diabetes is given below in fig.6, fig.7 and fig.8 respectively.

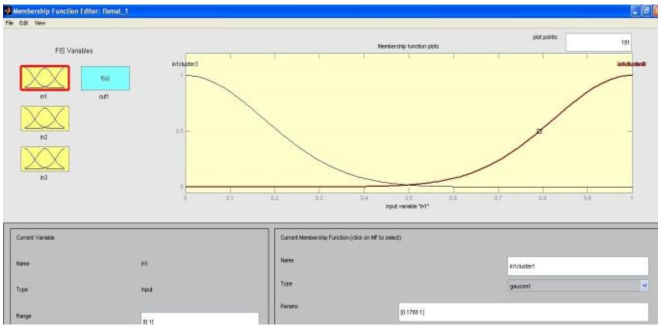

Fig. 6. FIS editor with Input and output variables.

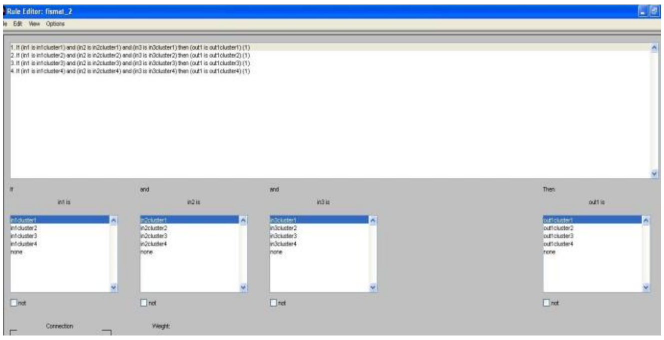

Fig. 7. Rule Evaluation.

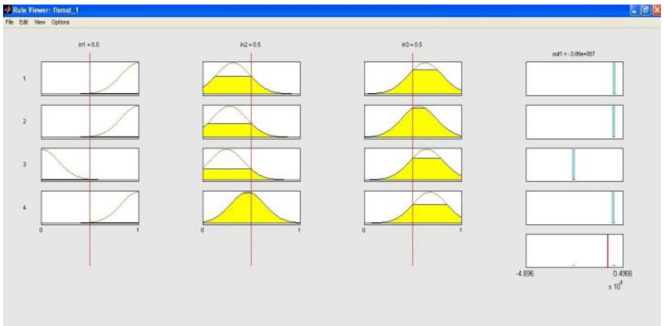

Fig. 8. Rule viewer of the FIS.

Table 4. Fuzzy parameters.

FIS Type	sugeno
AND method and Implication	prod
OR method and Aggregation	probor
Membership function	gaussian
Defuzzification	Wtaver

The Diabetes diagnosis and classification are performed either for type 1, type2 or both type and type. Our Neural network architecture focuses on the classification of Type 1, typ2 and gestational diabetes. Hence a layered neural network architecture is proposed.

Table 5. Comparison of diagnosis and classification results by different authors.

Attributes	Diabetes Diagnosis	Classifier	Accuracy %
Hang et al.,	Type 2	Naïve bayes SVM	95 76.30
Deepti et al.,	Type 2	Naïve bayes Decision tree	65.10 73.82
Sneha et al.,	Type1 & Type 2	Naïve bayes	82
Yu et al.,	Type 2	SVM	83.5
Lukmantoa, R et al.,	DM	Fuzzy SVM	89
Allalou, A. et al.,	GDM	Decision Tree	83

5.4. Evaluation matrices

Sensitivity is the process of classifying positive cases correctly, and specificity is a process to classify the negative instances correctly. The confusion matrix for classification is designed by True positive (TP), True negative (TN), False positive (FP) and False negative (FN). The accuracy is measured with measures of sensitivity and specificity. Measures are performed on both the data sets and graphically represented in bar Charts. The sensitivity, specificity, and accuracy are defined as,

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

$$Accuracy = \frac{TP+TN}{TN+FP+TP+FN} \quad (7)$$

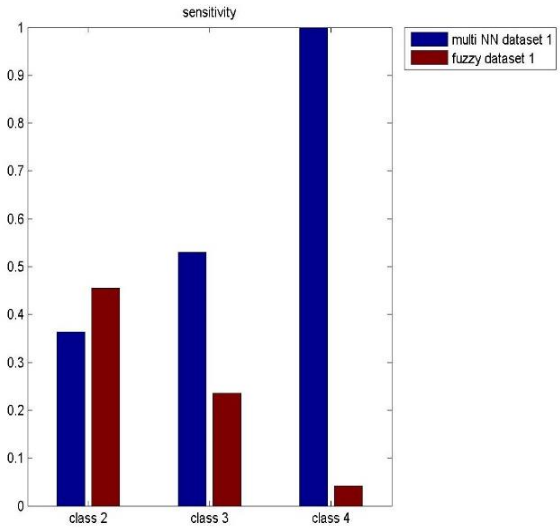

Fig. 9. Sensitivity of dataset 1.

Fig. 9 and Fig.10 represent the analysis of the proposed approach based on parameter sensitivity. The study from the graphs shows that in both cases, the sensitivity value varying in nature. In the case of dataset 1, the sensitivity Variation shows an incremental nature, while in dataset 2, it shows a zigzag view. The results of sensitivity show that, most of the patients are affected by type 2 diabetes. Since class 3 represents gestational diabetes, it will not be taken into account. The comparison of the sensitivity values with the layered neural network and fuzzy method shows considerable differences. In both the datasets, the proposed method shows more Sensitivity than the fuzzy approach. In this case, dataset 1, the fuzzy method shows slightly better performance for class 2. Also, it differs in specificity with the existing method and significant difference is shown only on the class2 and class3 of dataset 1. In both the datasets, the response of specificity shows an incremental nature shown in Fig. 11 and Fig.12.

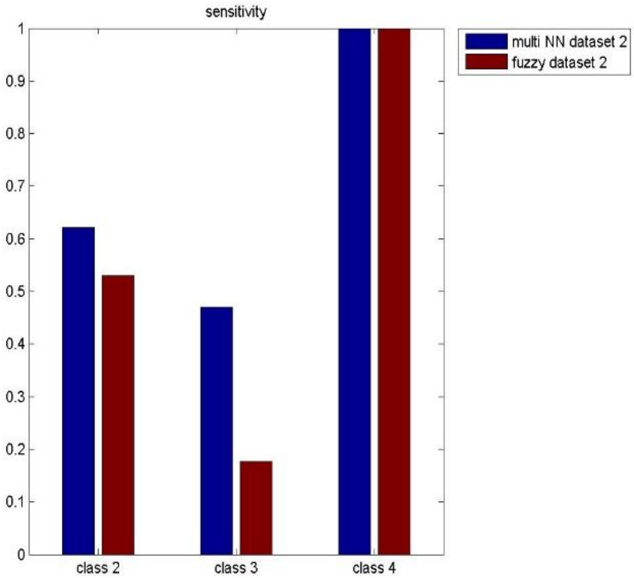

Fig. 10. Sensitivity of dataset 2.

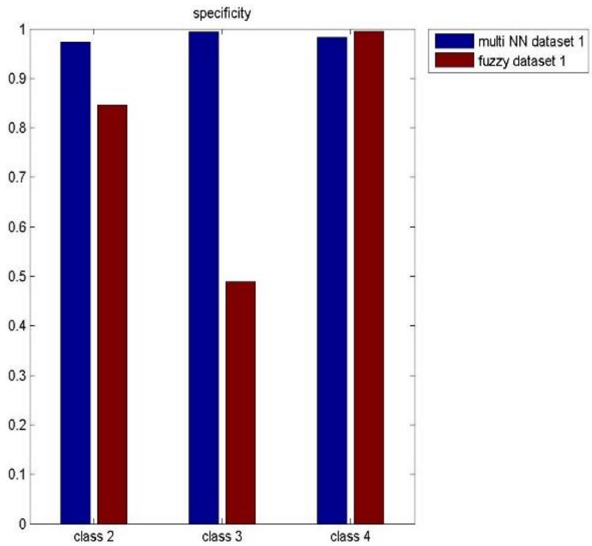

Fig. 11. Specificity of dataset 1.

The accuracy is the combinational result of the parameters specificity and Sensitivity according to their responses in the datasets. Fig. 13 and 14 shows the responses of efficiency based on the datasets. Both of the datasets produce an incremental way of response based on the classes. The analysis showed that most of the patients are related to the diabetes of type 2, which occurs in adult age. The values of different analysis are shown in the below Table 6 and Table 7.

Table 6. Classification of diabetes for dataset 1.

Datasets	Dataset 1		
Parameters	**Class 2**	**Class 3**	**Class 4**
Sensitivity	0.38	0.52	0.95
Specificity	0.94	0.94	0.97
Accuracy	0.86	0.94	0.98

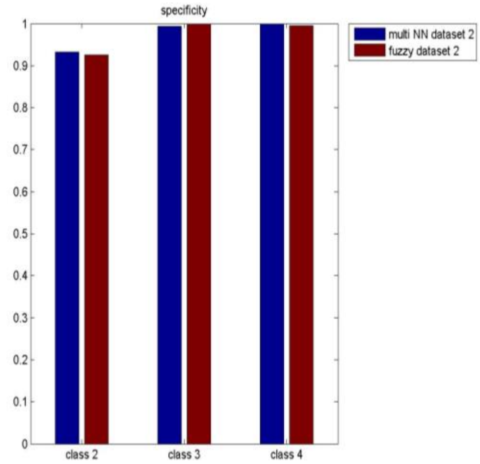

Fig. 12. Specificity of dataset 1.

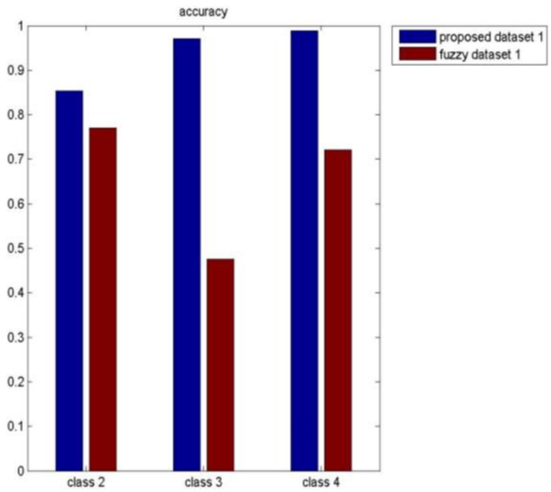

Fig. 13. Accuracy of dataset 1.

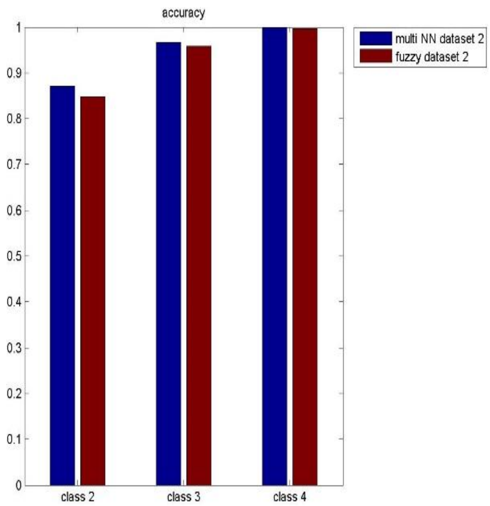

Fig. 14. Accuracy of dataset 2.

Table 7. Classification of diabetes for dataset 2.

Datasets	Dataset 2		
Parameters	**Class 2**	**Class 3**	**Class 4**
Sensitivity	0.62	0.46	0.96
Specificity	0.92	0.98	0.95
Accuracy	0.88	0.96	0.97

6. CONCLUSION

Diabetes mellitus occurs due to beta cells in pancreas which increase glucose levels. In the proposed approach, along with type 1 and type 2 diabetes, we have included gestational diabetes, which is present only for pregnant ladies. The Unknown data used for the testing phase are classified using a proposed layered neural network. The experimental results exhibit high accuracy of 98% of classification for all three categories of diabetes. The future work can be done through including semantic methods and disease diagnosis after classification.

REFERENCES

Abhari, S., Niakan Kalhori, S. R., Ebrahimi, M., Hasannejadasl, H., & Garavand, A. (2019). Artificial Intelligence Applications in Type 2 Diabetes Mellitus Care: Focus on Machine Learning Methods. *Healthcare informatics research*, *25*(4), 248–261.

Allalou, A., Nalla, A., Prentice, K. J., Liu, Y., Zhang, M., Dai, F. F., Ning, X., Osborne, L. R., Cox, B. J., Gunderson, E. P., & Wheeler, M. B. (2016). A Predictive Metabolic Signature for the Transition from Gestational Diabetes Mellitus to Type 2 Diabetes. *Diabetes*, vol.65,no.9, pp.2529–2539.

American Diabetes Association: Standards of medical care in diabetes, (2006). *Diabetes Care*, vol.29, suppl 1, S4–S42.

American Diabetes Association: Standards of medical care in diabetes, (2019). *Diabetes care*, vol.42, suppl 1, S1-S3.

American Diabetes Association: Detection and Diagnosis of GDM, (2013). *Diabetes care*, vol.36, Suppl 1, S11-S66.

American Diabetes Association: Classification and Diagnosis of Diabetes (2015). *Diabetes Care,* vol.38, Supp 1, S8–S16.

Andrews, R., Diederich, J., and Tickle, A.B. (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-Based Systems*, vol.8, no.6, pp.373-389.

Chiu, s., (1994). Fuzzy model identification based on cluster estimation, *J. Intell. Fuzzy Syst.*, vol.2, no.3, pp.267–278.

Cunningham, P., Carney, J., & Jacob, S. (2000). Stability problems with artificial neural networks and the ensemble solution. *Artificial intelligence in medicine*, vol.*20*, *no*.3, pp.217–225.

Dua, D., and Graff, C., (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

George D. Magoulas and Andriana Prentza (2001). Machine learning in Medical Applications. *Machine Learning and Its Applications*.pp.300-307.

Huang, Y., McCullagh, P., Black, N., and Harper, R. (2007). Feature selection, and classification model construction on type 2 diabetic patients' data. *Artif Intell Med*, vol.41, no.3, pp.251-262.

Janez, A., Guja, C., Mitrakou, A., Lalic, N., Tankova, T., Czupryniak, L., Tabak, A. G., Prazny, M., Martinka, E., and Smircic-Duvnjak, L. (2020). Insulin Therapy in Adults with Type 1 Diabetes Mellitus: a Narrative Review. *Diabetes therapy: research, treatment, and education of diabetes and related disorders, vol.*11(2), pp.387–409.

Karabegovic, A., Zikrija, A., and Ponjavic, M. (2006). Applications of Fuzzy Logic in Geographic Information Systems for Multiple Criteria Decision Making, pp.235-244.

Kharroubi, A. T., and Darwish, H. M. (2015). Diabetes mellitus: The epidemic of the century. *World journal of diabetes*, vol.6, no.6, pp.850–867.

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med*., vol.23, no.1, pp.89-109.

Lukmantoa, R. B., Suharjitoa, Nugrohoa, A., Habibullah, A. (2019). Early Detection of Diabetes Mellitus using Feature Selection and Fuzzy Support Vector Machine, *4th International Conference on Computer Science and Computational Intelligence*, vol.157, pp.46–54.

Moein, S., Monadjemi, S., and Moallem, P. (2008). A Novel Fuzzy-Neural Based Medical Diagnosis System. World Academy of Science, Engineering and Technology, Open Science Index 13, *International Journal of Computer and Information Engineering*, vol.2, no.1, pp.79 - 83.

Nanda, S., Savvidou, M., Syngelaki, A., Akolekar, R., and Nicolaides, K. H. (2011). Prediction of gestational diabetes mellitus by maternal factors and biomarkers at 11 to 13 weeks. *Prenatal diagnosis*, *31*(2), 135-141.

Olokoba, A. B., Obateru, O. A., and Olokoba, L. B. (2012). Type 2 diabetes mellitus: a review of current trends. *Oman medical journal*, vol.*27,* no.4, pp.269–273.

Short life, E. H (1987). Computer Programs to Support Clinical Decision Making. *Journal of the American Medical Association,* vol.258, no.1, pp.61-66

Sisodia, D. and Sisodia D.S. (2018). Prediction of Diabetes using Classification Algorithms, *Procedia Computer Science*, vol.132, pp.1578-1585, doi:10.1016/j.procs.2018.05.122.

Sneha, N., and Gangil, T. (2019). Analysis of diabetes mellitus for early prediction using optimal features selection. *Journal of Big Data* 6,13.

Steimann, F., and Adlassnig, K.P. (1998). Fuzzy medical diagnosis.pp.1-16. Doi:101887/0750304278/b438c104.

Taghiyev, A., Alpaslan, A. A., Allahverdi, N., and Caglar, S. (2019). A Machine Learning Framework to Identify the Causes of HbA1c in Patients with Type 2 Diabetes Mellitus. Control Engineering and Applied Informatics. Vol.21, no.2, pp.34-42.

Wang Y Liu S, Chen R, Chen Z, Yuan J, Li Q (2017). A Novel Classification Indicator of Type 1 and Type 2 Diabetes in China. *Sci Rep*, vol.7, no.1:17420.

Yi, L., Swensen, A. C., and Qian, W. J. (2018). Serum biomarkers for diagnosis and prediction of type 1 diabetes. *Translational research: the journal of laboratory and clinical medicine*, vol.201, pp.13–25.

Yu, W., Liu, T., Valdez, R., Gwinn, M., and Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC medical informatics and decision making*, *10*:16.

Zhou, Z. H., and Jiang, Y. (2003). Medical diagnosis with C4.5 Rule preceded by artificial neural network ensemble. *IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society*, vol.7, no.1, pp.37-42.

Dataset

http://archive.ics.uci.edu/ml/datasets/Diabetes.

https://github.com/rrichajalota/Pima-Indians-Diabetes-kaggle.git.

http://staff.pubhealth.ku.dk/~tag/Teaching/share/data/Diabetes