A BRIEF SURVEY ON THE TECHNIQUES USED FOR THE DIAGNOSIS OF DIABETES-MELLITUS

Pragati Agrawal [1], Amit kumar Dewangan [2]

[1] *Research Scholar, Deptt. of Computer Science, Dr. C V Raman University, Kota, Bilaspur (C.G.), India*
[2] *Amit kumar Dewangan, Deptt. of Computer Science, Dr. C V Raman University, Kota, Bilaspur (C.G.), India*

Abstract - *Diabetes-Mellitus refers to the metabolic disorder that happens from misfunction in insulin secretion and action. It is characterized by hyperglycemia. The persistent hyperglycemia of diabetes leads to damage malfunction and failure of different organs such as kidneys, eyes, nerves, blood vessels and heart. In the past decades several techniques have been implemented for the detection of diabetes. The diagnosis of diabetes is very important now a days using various types of techniques.*

Here we summarize various techniques, their classification and implementation using various types of software tools and techniques. The diagnosis of diabetes can be done using Artificial Neural Network, K-fold cross validation and classification, Vector support machine, K-nearest neighbor method, Data Mining Algorithm, etc.

Key Words: Diabetes Mellitus, Data Mining, Neural Network, Accuracy, etc…

1. INTRODUCTION

The term diabetes mellitus corresponds to a metabolic disorder of various symptoms classified by persistent hyperglycaemia with disturbances of carbohydrates, fats and proteins metabolism laeding from problems in insulin secretion, insulin action, or both. The affect of diabetes mellitus consists of chronic malfunction, damage and failure of several organs. Diabetes mellitus arises with classified symptoms such as blur vision, polyuria, and excessive weight loss. Most of the cases of diabetes are categorized into two main etiopathogenic categories:

- Type 1
- Type 2

1. TYPE-1 DIABETES

It occurs due to complete lacking of insulin secretion. It is detected by serological corroboration of an autoimmune pathology process occurring in islets of pancreas and by genetic markers.

2. TYPE-2 DIABETES

It occurs due to combination of resistance to insulin action and deficient compensatory insulin secretory response.

These are the types of diabetes. [1]

Several pathogenetic steps are included in the occurence of diabetes. These include steps like destruction of beta cells of the pancreas with consequent insulin insufficiency, and rest that leads to problems in insulin action. The abnormal action of carbohydrates, fats and proteins metabolism are due to insufficient action of insulin on target tissues resulting due to lack of insulin.

2. METHODOLOGY

Various Methods:

2.1 K-fold cross validation and classification

In this process, Neural Network instead of using difficult techniques like linear aggregation, it includes non linearity in their aggregation.

The non-linearity reduces profound architecture of these networks. It also helps in generation of high order terms.

Generally, in Neural Network we are using summation at the node. But, in this process we are using multiplication at the nodes.

Here we apply Error-Back Propagation Based Learning using Norm-Square Based Error.

These type of functions are basically used in order to solve complete problems and it needs low number of parameters in comparison with known typical models.

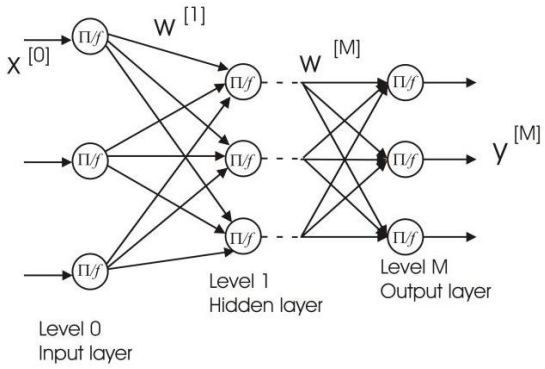

Fig -1: Higher Order Neural Network

The back propagation learning is generally divided into two phases: propagation and weight update. The weights in this process are updated after the entire training sequence applied to the network once. Thus, in the process of finding and refining the foreboding accuracy we used the method of Data Preprocessing. Before the using of this method, the process of training in neural network was very slow. Here we see the simulation of this process. It uses the concept of K-fold cross validation.

k-fold cross validation of data set :

Firstly, the data is partitioned into K equal parts.

Testing	Training	Training
Training	Testing	Training
Training	Training	------- Testing
Training	Training	Training
Training	Training	Training

In this way the partition of data set is created. For each 'k' processes,

'k-1' folds are used for training and the remaining one for testing.

And the average error is calculated.

During the execution of this process various attributes for the detection of diabetes have been taken into consideration in our data set. Some of the attributes used here are: age, class, plasma glucose, diastolic blood pressure, etc.

Thus in this way we see the approach for the detection of diabetes using Higher Order Neural Network.[2]

2.2 Classwise K Nearest Neighbor Method

In this process, a modern class-wise K-Nearest Neighbor (CKNN) classification algorithm for classification of diabetes data. They have used diabetes data set for testing the CKNN algorithm and compared the various performance like accuracy, sensitivity and specificity with simple KNN. The proposed CKNN model gives better classification accuracy as 78.16% compared to simple KNN which is 71.84. Other performance measures are also better than simple KNN. Overall, it is more convenient than simple KNN [3].

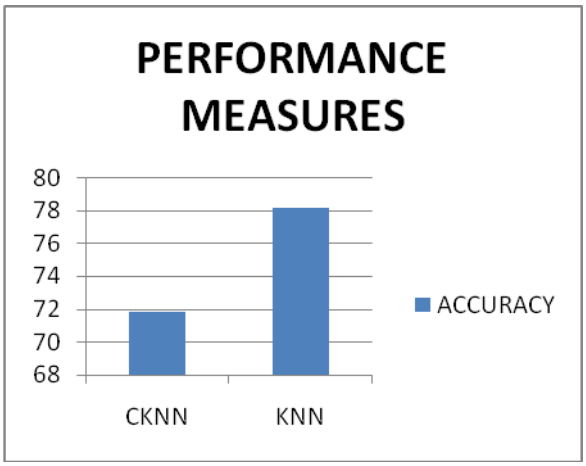

Chart -1: Comparison of Performance

2.3 Using Support Vector Machine

Support Vector Machine (SVM) with Radial basis kernel function for classification of diabetes data. They have used Pima Indian diabetes data set which is collected from UCI repository and trained and tested on SVM as classifier. The proposed model achieved 78% accuracy which can be successfully used for diagnosing diabetes disease. It is a collection of medical diagnostic reports from 768 records of female patients of at least 21 years old. It indicates two classes 0 or 1, 1 stands for positive result and 0 stands for negative result. As, they collected total of 768 samples with which 500 cases are of class '0' i.e. negative result and 268 cases are of class '1' i.e. positive result.

Although, out of these 768 cases, there are many of the cases which has missing values, so after the deletion of all these missing values cases there were 460 cases left with genuine values. The final training, testing and the process of finding accuracy was performed in this cases [4].

2.4 Using LDA-Support Vector Machine and Feed Forward Neural Network

Linear Discriminant Analysis and Support Vector Machine for the diagnosis of Pima Indians Diabetes dataset, where LDA reduces feature subsets and SVM is responsible to classify the data. They have also compared SVM with feed

forward neural network (FFNN) but our proposed SVM+LDA gives better classification accuracy as 77.60% with 2 features [5].

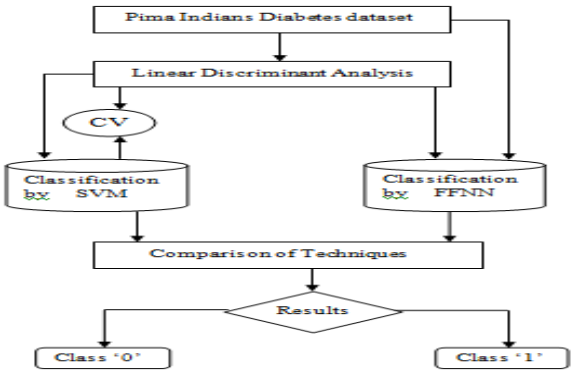

Fig -2: Proposed Architecture

2.5 Using Artificial Neural Network

A fuzzy approach and the process of normalization characterized by Artificial neural networks as 'Relate to Type 2 diabetic' or not. This outcome indicates the successfulness of proposed algorithm to optimal model, the diagnostic method for small or large datasets; the main cause behind this is its enumerated simplicity. Here we combined Fuzzy modeling and artificial neural network architecture [6].

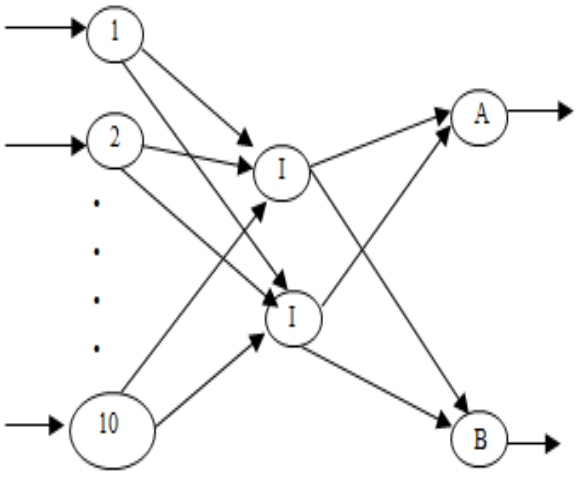

Fig -3: Back Propagation Neural Network

2.6 Statistical Normalization and Back Propagation for Classification

The artificial neural network has currently been used in some of the areas including diagnosis of diseases like diabetes, thyroid, etc. Even though its successful application for many of the diseases but there are some major issues that can be taken into account before using the neural network models, such as the network topology, learning parameter, and normalization process for the input vectors. Taking into consideration all the factors, it is used in the application of diagnosis of diseases.

Various normalization methods applied in back propagation neural networks to increase the effectivity of the trained network. The experimental results explains that the performance of the diabetes data classification model with the use of neural networks was based on the normalization methods [7].

Normalization Method	Accuracy
Z-Score	69.45
Min-Max	69.67
Median	64.31
Sigmoid	69.28
Mean and Standard Deviation	64.44
Statistical Column	72.55

Fig -4: Normalisation Accuracy

2.7 Data Mining Methods and Techniques for Diabetes Diagnosis

Data mining is used in the field of medical science for the diagnosis of diabetes. In Data Mining, Classification Algorithms usually needs that the classes should be defined on the basis of data attribute values. They generally explain these classes by looking at the classification of data already known to class. Pattern Recognition is a type of classification where an input pattern is classified into one of the several classes based on its similarity to these predefined classes. Knowledge Discovery in Databases (KDD) is the process of finding useful information and patterns in data which involves Selection, Pre-processing, Transformation, Data Mining and Evaluation [8].

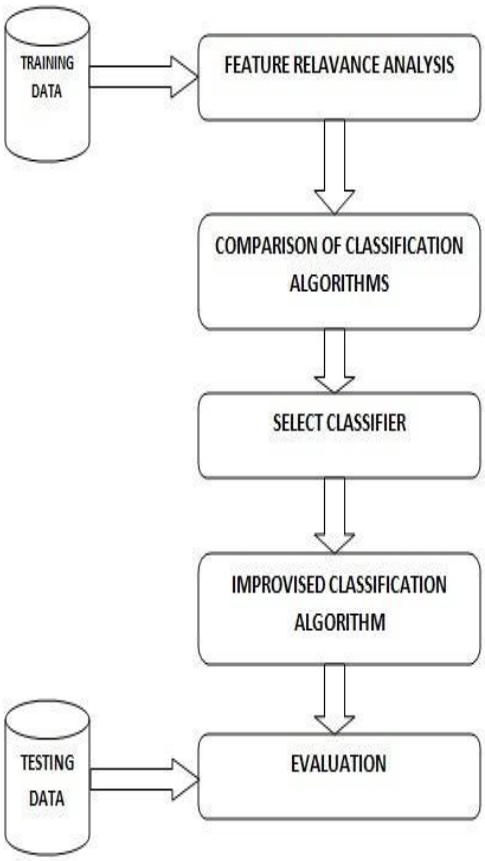

Fig -5: Proposed Architecture

2.8 Data Mining Algorithms Application in Diabetes Diseases Diagnosis

A data mining algorithm consists of a set of heuristics and calculations that extracts a data mining model from data. To generate model in data mining, the algorithm first checks the data you feed to it, searching for specific types of patterns or trends. Then the algorithm uses the outcome of this analysis to generate the optimal parameters for generation of the mining model. These parameters are then passed to the entire data set to extract patterns that can perform action and breif statistics.

Diabetes data set on various classification algorithms like SVM, KNN, Naïve bayes,ID3, CART and C5.0 to classify the diabetes data. They have compared the classification accuracy of these models. SVM gives best classification accuracy as 81.77% compare to others [9].

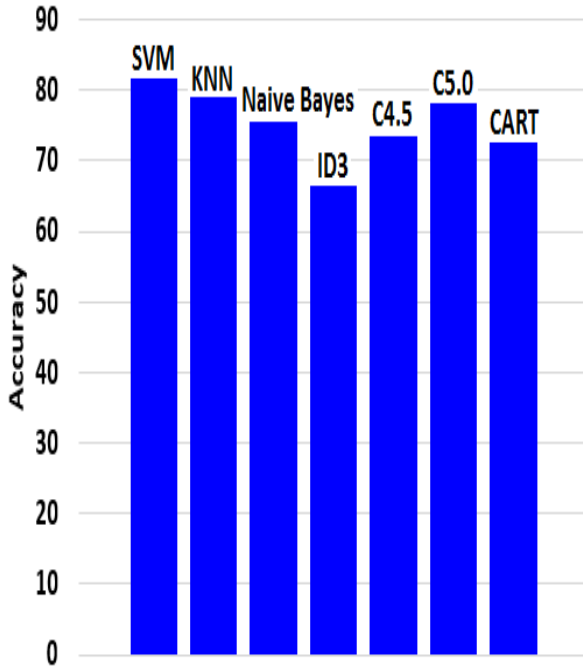

Chart -2: Comparision of Accuracy

3. CONCLUSIONS

We have studied various methods for the diagnosis of Diabetes Mellitus. A brief survey of all the methods is done in order to find the details and process of all the techniques used for the diagnosis of diseases. At last, we can say that Data Mining Algorithms, Machine Learning Techniques and various other techniques have done a remarkable job in the field of medical science and for the diagnosis of diseases.

Thus, we see all the processes for the diagnosis of Diabetes Mellitus.

REFERENCES

[1] "Diagnosis & Classification of Diabetes Mellitus ", Diabetes Care, Volume 37, Supplement 1, 2014.

[2] Raj Anand, Vishnu Pratap Singh Kirar, Kavita Burse, " K-Fold Cross Validation and Classification Accuracy of PIMA Indian Diabetes Data Set Using Higher Order Neural Network and PCA ", IJSCE, Volume-2, Issue-6, January 2013.

[3] Y. Angeline Christobel, P.Sivaprakasam, " A New Classwise k Nearest Neighbor (CKNN) Method for the Classification of Diabetes Dataset", IJEAT, Volume-2, Issue-3, February 2013.

[4] Kumari V. Anuja, Chitra R. (2013). Classification of Diabetes Disease Using Support Vector Machine.

International Journal of Engineering Research and Applications. Vol. 3, pp. 1797-1801, ISSN: 2248-9622.

[5] Parashar A., Burse K., Rawat K. (2014). A Comparative Approach for Pima Indians Diabetes Diagnosis using LDA-Support Vector Machine and Feed Forward Neural Network. International Journal of Advanced Research in Computer Science and Software Engineering. Vol. 4, pp. 378-383, ISSN: 2277 128X.

[6] Rajeswari K., Vaithiyanathan V. , (2011) Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network, IJCSNS International Journal of Computer Science and Network Security, VOL.11, pp. 126-130.

[7] Jayalakshmi T., Santhakumaran A., (2011) Statistical Normalization and Back Propagation for Classification, International Journal of Computer Theory and Engineering, Vol.3, No.1, February, 2011, 1793-8201, pp. 89-93.

[8] Rajesh K., Sangeetha V.,(2012) Application of Data Mining Methods and Techniques for Diabetes Diagnosis, International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 3, September 2012, ISSN: 2277-3754, pp. 224-229.

[9] Farahmandian M., Lotfi Y., Maleki I. (2015). Data Mining Algorithms Application in Diabetes Diseases Diagnosis: A Case Study. MAGNT Research Report. Vol. 3, PP. 989-997, ISSN. 1444-8939.

