

Survey on the Various Techniques Used For the Diagnosis of Diabetes-Mellitus

Swapnali N. Padawale¹, B. D. Jadhav²

¹(PG Student, E&TC Dept, Rajarshi Shahu College of Engineering/ Savitribai Phule Pune University, India)

²(Faculty, E&TC Dept, Rajarshi Shahu College of Engineering Pune/ Savitribai Phule Pune University, India)

ABSTRACT :Diabetes mellitus is generally known as diabetes, in which a person is diagnosed with high blood glucose level .Diabetes is a metabolic disorder caused when body is failed to produce insulin or the cells do not respond to the insulin that is produced .The presence of high blood glucose leads to damage, malfunction and failure of different organs such as eyes, kidneys, nerves, heart and blood vessels. The diagnosis of diabetes is very important nowadays using various types of techniques. Blood glucose test is the vital method for diagnosing diabetes which is an invasive method. Non Invasive methods are also explained in detail in this paper. 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 neighbour method, Data Mining Algorithm, etc.

Keywords - Diabetes Mellitus, Data Mining, Neural Network, Support Vector Machine

I. INTRODUCTION

Diabetes is one of the most chronic diseases in the world today. Diabetes is a leading cause of death in most developed countries, and is spreading rapidly in many developing nations. According to IDF (International Diabetes Federation) diabetes affects more than 140 million people in the world and 50 million people are suffering from this disease in India. The major symptoms and effects of diabetes mellitus include long-term injury, dysfunction, and functional abnormalities in eyes, nerves, kidneys, blood vessels and heart. Recognizing a type of diabetes to an individual is mostly based on the situation at the time of diagnosis.

There are three major categories of diabetes mellitus, which are-

- 1) Type 1 Diabetes: This type of abnormality occurs because of the malfunction of the body to generate insulin.
- 2) Type 2 diabetes: This type of class results from insulin resistance, where the cells cannot utilize insulin in proper proportion.
- 3) Gestational diabetes: This occurs when pregnant women, without diabetes but having high blood glucose level during pregnancy. It may possibly lead to the development of type 2 DM.

In recent years, the number of diabetic patients has increased drastically mainly due to the aging population and irregular western food habits. Genetic inheritance is the main reason for the cause type 1 and 2 diabetes classes. The main aim of treating diabetes is to control the sensitive complications of diabetes, and to eliminate the chronic complications of diabetes. For effective diagnosis of diabetes, the main factor that has to be considered is the risk of diabetic complications, early and accurately.

Numerous techniques have been developed for the diagnosis of diabetic mellitus. Most of the techniques used clustering and classification for the effective diagnosis of the diabetic mellitus disease. But, there is always a scope for improvement and still several techniques are being developed to overcome the limitations of the existing techniques. This paper presents an analytical study on the existing techniques available for diabetes mellitus. The characteristic features of the approaches are investigated to develop a better approach for the early and efficient diagnosis of the disease. Blood glucose monitoring is known as the gold standard for diabetes diagnostics and self-monitoring. But, this process is invasive and uncomfortable for patients. Furthermore, the process must be completed several times a day to successfully manage the disease, which greatly contributes to the immense need for non-invasive monitoring options. We have also summarized non invasive methods using image processing.

II. METHODOLOGIES

This section discusses about the existing techniques and algorithms used for the diagnosis of diabetes mellitus. Each and every algorithm used for the diagnosis of diabetes mellitus has their own limitations and advantages. This section presents an analytical study on the features of the existing techniques.

2.1 Using Intelligent Support Vector Machines

A hybrid model has proposed to diagnose diabetes mellitus by integrating three different data mining techniques. It uses supervised learning algorithm for model building and unsupervised learning algorithm for sampling. The inputs were processed with the help of Support Vector Machine (SVM) and approach was used for extracting rules. Input for this system is real time data set from that rules are formulated to describe relationships between input features and output class labels. Results show that intelligible SVMs give a capable tool for the diagnosis of diabetes. An intelligible rule set has been generated, with prediction accuracy of 94%, sensitivity of 93%, and specificity of 94%. Two rules are generated by the proposed approach from those one was found to be inconsistent with generally acceptable medical knowledge in diagnosis of diabetes [4].

2.2 Data Mining Algorithms

Farahmandian M., Lotfi Y., Maleki I. [11] designed a data mining algorithm, which 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 which is feed to it, searching for specific types of trends or patterns. 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 for extracting patterns that can perform action and brief statistics Diabetes data set on various classification algorithms like SVM, KNN, 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.

2.3 Classwise K Nearest Neighbor (Cknn) Method

Ajita Satheesh, Ravindra Patel [7] designed a dynamic nearest neighbor classifier for data integrated via object oriented concept and generalization. They integrated the data collected from multiple service providers into a single consolidated unit (training instance). The traditional KNN is enhanced using normalization and majority voting. As compared to the traditional KNN, the results show that the dynamic nearest neighbor is more efficient. This paper address the ways to improve the classification of the KNN algorithm for classification of Diabetes data with imputed missing values a new class wise nearest neighbor algorithm(CKNN) for improving the performance of the standard KNN. In this process, a modern class-wise K-Nearest Neighbor (CKNN) classification algorithm is used for classification of diabetes data. They have used diabetes data set for testing the CKNN algorithm and compared the various performances like accuracy, specificity and sensitivity 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.

2.4 Using B-Colouring Technique In Clustering Analysis

Lakshmi V. & Thilagavathi K [1] proposed a clustering algorithm based on a graph b-colouring technique, which was used to cluster Pima Indian diabetic dataset. They have implemented, performed experiments, and compared their approach with KNN Classification and K-means clustering. The results show that the clustering based on graph colouring out performance than the other clustering approach in terms of purity and accuracy. The proposed technique offers a real representation of clusters by dominant objects that guarantees the inter cluster disparity in a partitioning and used to calculate the quality of cluster. Accuracy of b-colouring approach samples is 93.676%.

2.5 Lda-Support Vector Machine And Feed Forward Neural Network

Parashar A., Burse K., Rawat K. [10] proposed Linear Discriminant Analysis and Support Vector Machine for the diagnosis of Pima Indians Diabetes dataset. In this method LDA reduces feature subsets and SVM is used to classify the data. For making diabetes diagnosis easier for Physicians, there have been several methods employed and for attaining greater performance they have reduced attributes of diabetes dataset using LDA. Then they have implemented and compared three different techniques Support Vector Machine, Feed Forward Neural Network and LDA with Feed forward networks. On comparison of these methods, then they have concluded that LDA-SVM has proved to be much better than the other two methods and is more efficient to be handled due to its reduced complexity. They have also compared SVM with feed forward neural network (FFNN). Proposed SVM+LDA gives better classification accuracy as 77.60% with 2 features.

2.6 Artificial Neural Networks Technique

Rajeeb Dey [8] has proposed a method to diagnose diabetes mellitus using back propagation algorithm of Artificial Neural Network (ANN). Diagnosis of diabetes was done using a binary classification, those predicted to be diabetic falls under category 1 and others under category 0. The present system is a classification problem applied for diagnosis of diabetes mellitus using back propagation algorithm of ANN [18]. ANN architecture is used for classification. In this system parameters considered to diagnose diabetes are Random Blood Sugar test result, Post Plasma Blood Sugar test, Fasting Blood Sugar test result, Sex, age and their occupation. This system achieves classification performance of 92.5%.

Jaffar, Ali [5] proposed a method to diagnosis diabetes. The diagnosis is done using back propagation neural network algorithm. 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 are the inputs to the system. The missing values in the data set were biggest challenge in this system. This system has been modified and obtained [12]. An idea has proposed to overcome the missing values those were not addressed in previous system, by constructing the data sets with reconstructed missing values. This has helped to increase the classification accuracy[19]. They have also proposed a method helps to conquer missing value by doing data pre-processing, this helps to increase the speed of training process by reducing the actual learning time. In this paper, they analyze a variety of missing value techniques and pre-processing methods. This has improved the results and achieved classification accuracy of 99%.

2.7 Using K-Fold Cross Validation and Classification Using HONN and PCA

Anand R. et al. [9] have suggested Principle Component Analysis and Higher Order Neural Network for classification of Pima Indian diabetes data set. They have applied Error-Back Propagation Based Learning using Norm-Square Based Error. These types of functions are basically used in order to solve complete problems. It needs low number of parameters in comparison with known typical models. The proposed model gives faster convergence with PCA pre processing and lower mean square error [20].

Table1: Classification Accuracy of Different Methods

Paper	Method	Classification Accuracy %
An Improved Data Mining Model to Predict the Occurrence of Type-2 Diabetes using Neural Network[16]	C4.5	92.38
Diabetes mellitus forecast using ANN[5]	ANN	80.88
Prediction of Diabetic disease by using b-Colouring Technique in Clustering Analysis[1]	b-Clustering	93.67
Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus[4]	Intelligible SVM	94
A Hybrid Model of Hierarchical Clustering and Decision Tree for Rule-based Classification of Diabetic Patients[17]	Hybrid model of hierarchical clustering	81
Classification of Diabetes Disease Using Support Vector Machine[15]	SVM with RBF	78
Application of artificial neural network technique for diagnosing diabetes mellitus[8]	ANN	92.5
A Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks[12]	ANN	99
A Comparative Approach for Pima Indians Diabetes Diagnosis using LDA Support Vector Machine and Feed Forward Neural Network[10]	LDA-SVM and FFNN	75.65

2.8 Using Image Processing

2.8.1 Noninvasive Method Using Facial Block Color with A Sparse Representation Classifier

Bob Zhang et.al. [2] have proposed new methods to detect DM non-invasively. Such as, facial block color features with the Sparse Representation Classifier (SRC) is described to detect DM. They classified DM vs. Healthy based on a combination of four facial blocks with the SRC. The four facial blocks were located on

the nose, the forehead, and below the right and left eyes, with each block represented by 6 colors. They achieved an accuracy of 97.54% through experiments with 284 DM samples and 142 Healthy samples. However, texture features from those blocks were not extracted to see if they can also be used to detect DM.

2.8.2 Based On The Fractal Characteristics Of Retinal Images

Shu-Chen Cheng and Yueh-Min Huang [14] have proposed a novel diagnostic method to develop quantitative indexes for diabetes mellitus. It was discovered that the fractal dimension of a severe diabetic patient's retinal vascular distribution appears greater as compared to a normal human's; hence the fractal dimension of the vascular distribution was estimated. The issue of how to yield an accurate fractal dimension is to use high quality of images. An appropriate image-processing algorithm adopted to achieve better image processing results. The measure of lacunarity was another important fractal feature introduced in this paper. Measure of lacunarity describes the characteristics of fractals. That have same fractal dimension but the different appearances. Further classification can be made using the degree of lacunarity for those vascular distributions in the same fractal dimension.

The resolution of the original image was also discussed in addition to the image-processing technique. They have presented a system for examine the resolution of the source image is high or not. Using extraordinarily high-resolution images is simply a waste of disk space and computation time and cannot achieve higher accuracy of fractal dimension. Diagnosis results were classified using four different methods the radial basis function network, the back propagation algorithm, the genetic algorithm, and the voting scheme were compared to achieve higher accuracy. In this paper, the fractal dimension and the measure of lacunarity have found to be so successful for the classification of diabetes that they are adequate for use as quantitative indexes. The best classification result was 89.2% correctness.

2.8.3 Using Tongue Color, Texture, and Geometry Features

Bob Zhang et.al. [6] have proposed a non-invasive approach to classify Healthy/DM and NPDR/DM-sans NPDR samples using three groups of features extracted from tongue images. Color, texture, and geometry included in these three groups. A tongue color gamut was first applied such that each tongue image can be represented by 12 colors. Afterward, eight blocks strategically located on the tongue were extracted and its texture value calculated. Finally, 13 geometry features from tongue images based on measurements, their ratios, distances and areas were extracted. Numerical experiments were carried out using 130 Healthy and 296 DM tongue images. By applying each feature individually to separate Healthy/DM, the highest average accuracy achieved (via SVM) was only 66.26%. However, employing SFS with the SVM, nine features (with elements from all the three groups) were shown to produce the optimal result, obtaining an average accuracy of 80.52%. As for NPDR/DMSans NPDR classification, the best result of 80.33% was attained using five features: one from texture, one from geometry and three from color.

2.8.4 Based On Facial Block Texture Features Using The Gabor Filter

Shu Ting; Zhang, B. [3] have proposed a novel method to detect DM non-invasively based on facial block texture features. This method gives an accuracy of 99.64%, a sensitivity of 99.29%, and a specificity of 100%. This method first extracts a texture feature value from each facial block using a Gabor filter bank consisting of 40 filters (five scales and eight orientations). Afterwards, these texture feature values of the facial blocks are then used to classify DM samples vs. Healthy samples on a dataset of 284 and 142 corresponding, using SVM and *k*-NN with 10-fold cross validation. From the experimental results it can be concluded that texture features extracted from facial blocks are very effective and efficient at DM detection.

Table 2: Accuracy of Different approaches using Image Processing

Paper	Approach	Accuracy %
diabetes based on the fractal characteristics of retinal images[14]	The Fractal Characteristics Of Retinal Images	89.2
Diabetes Mellitus and Nonproliferative Diabetic Retinopathy Using Tongue Color, Texture, and Geometry Features[6]	Tongue Color, Texture, And Geometry Features	80.33
Diabetes Mellitus Detection Based on Facial Block Texture Features Using the Gabor Filter[3]	Facial Block Texture Features Using The Gabor Filter	99.64
Non-invasive diabetes mellitus detection using facial block color with SRC[2]	Facial Block Color With A SRC	97.54

III. CONCLUSION

The various methods have been studied 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. Some systems use Pima Indian dataset for diagnosis purpose. However, some systems used the patient's data from hospital. This paper clearly presents an analytical study of numerous algorithms which includes clustering, classification, vector machines and neural networks. At last, we can say that Data Mining Algorithms, Machine Learning Techniques and various other techniques have done a remarkable job. Also various methods using image processing have done an amazing 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] Lakshmi V. & Thilagavathi K. (2012). "An Approach for Prediction of Diabetic disease by using b-Colouring Technique in Clustering Analysis". *International Journal Of Applied Mathematical Research*, 1(4), 520-530.
- [2] B. Zhang, V. Bhagavatula, and D. Zhang, "Non-invasive diabetes mellitus detection using facial block color with sparse representation classifier," *Biomedical Engineering, IEEE Transactions on*, vol. 61, no. 4, pp1027-1033, 2014.
- [3] Shu Ting; Zhang, B., "Diabetes Mellitus Detection Based on Facial Block Texture Features Using the Gabor Filter," in *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on*, vol., no., pp.1-6, 19-21 Dec. 2014
- [4] Barakat N.; Bradley, A.P.; Barakat, M.N.H., "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus," in *Information Technology in Biomedicine, IEEE Transactions on*, vol.14, no.4, pp.1114-1120, July 2010
- [5] Jaafar, S.F.B.; Ali, D.M., "Diabetes mellitus forecast using artificial neural network (ANN)," in *Sensors and the International Conference on new Techniques in Pharmaceutical and Biomedical Research, 2005 Asian Conference on*, vol., no., pp.135-139, 5-7 Sept. 2005
- [6] Zhang, B.; Kumar, B.V.K.V.; Zhang, D., "Detecting Diabetes Mellitus and Nonproliferative Diabetic Retinopathy Using Tongue Color, Texture, and Geometry Features," in *Biomedical Engineering, IEEE Transactions on*, vol.61, no.2, pp.491-501, Feb. 2014
- [7] Ajita Sathesh, Ravindra Patel, "Dynamic Nearest Neighbours Classifier For Integrated Data Using Object Oriented Concept Generalization" *IJSSST, Vol.11, No. 1*
- [8] 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 ICHIS, Kharagpur, INDIA December 8-10*.
- [9] 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, pp. 436-438, ISSN: 2231-2307.
- [10] 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.
- [11] 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.
- [12] Jayalakshmi, T.; Santhakumaran, A., "A Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks," in *Data Storage and Data Engineering (DSDE), 2010 International Conference on*, vol., no., pp.159-163, 9-10 Feb. 2010
- [13] Peter, S. (2014) "An Analytical Study on Early Diagnosis and Classification of Diabetes Mellitus". *Bonfring International Journal of Data Mining*, 4(2), 07-11.
- [14] Shu-Chen Cheng and Yueh-Min Huang, "A novel approach to diagnose diabetes based on the fractal characteristics of retinal images", in *Information Technology in Biomedicine, IEEE Transactions on*, vol.7, Issue-3, pp. 163-170, Sept. 2003.
- [15] 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.
- [16] S.Priya, R.R.Rajalaxmi, "An Improved Data Mining Model to Predict the Occurrence of Type-2 Diabetes using Neural Network", *International Conference on Recent Trends in Computational Methods, Communication and Controls (ICON3C 2012)*.
- [17] Norul Hidayah Ibrahim, Aida Mustapha, Rozilah Rosli, Nurdhiya Hazwani Helmee, "A Hybrid Model of Hierarchical Clustering and Decision Tree for Rule-based Classification of Diabetic Patients", *International Journal of Engineering and Technology (IJET) Vol 5 No 5 Oct-Nov 2013*.
- [18] Paithane, A. N., and D. S. Bormane. "Analysis of nonlinear and non-stationary signal to extract the features using Hilbert Huang transform." *Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on*. IEEE, 2014.
- [19] Paithane A. N. and D. S. Bormane. "Electrocardiogram signal analysis using empirical mode composition and Hilbert spectrum." *Pervasive Computing (ICPC), 2015 International Conference on*. IEEE, 2015.
- [20] Paithane A. N., D. S. Bormane, and Sneha Dinde. "Human Emotion Recognition using Electrocardiogram Signals." *International Journal on Recent and Innovation Trends in Computing and Communication* ISSN: 2321-8169 Volume: 2 Issue: 2.