

MACHINE PREDICTS THE DIAGNOSIS A BRIEF REVIEW OF MEDICAL DIAGNOSIS BY MACHINE LEARNING TECHNIQUES

Tushar Deshmukh

School of Computational Science, S. R. T. M. University Nanded
Nanded, M.S. 431605, India
gemini.tushar@gmail.com

Dr. H. S. Fadewar

School of Computational Science, S. R. T. M. University Nanded
Nanded, M.S. 431605, India
gemini.tushar@gmail.com

Abstract Big data analytics and intelligent decision making systems are gaining importance in today's era. Applications of such systems are wide spread and have changed major areas of human living. Even medical domain is being revamped by these systems. Now intelligent systems predict or suggest the diagnosis based on history of the patient. Early predictions for critical disease like diabetes are very much possible. This review gives a brief insight to such systems with focus on diabetic patients. Survey shows it can also help out to point out the various risk factors that can alarm the scenario well in time. The intension of the paper is to analyze how data mining can be helpful in such early diagnosis of diabetes as well as to see how different researchers has used this techniques for better predictions.

Keywords: Diabetes, prediction of diabetes, data mining technique etc.

1. Introduction

The global rate of diabetes has increased from 4.7% to 8.5% within almost 25 years i.e. 1980 to 2014 [1]. The predictions say that the almost 109 million Indians might be suffering from diabetes [2].

The defects in the secretion of insulin or the use of insulin may cause the metabolism diseases, which is termed as diabetes. [3] Diabetes can be categories into different category.

Type1 Diabetes:

This type of diabetes is also called as juvenile-onset diabetes or insulin dependent diabetes. The β -cells of the pancreas are destructed which are responsible for the production of insulin. The rate of destruction is very fast in younger age. Environmental factors and genetic factors may the suspect for this type of diabetes.

Type 2 Diabetes:

Diabetes mellitus also called Type 2 Diabetes causes sugar to collect in blood stream. It becomes very difficult for the body cell to absorb the insulin or to use the insulin. Type 2 diabetes is more common than Type 1 diabetes. This type of diabetes is sometimes called as adult-onset diabetes. There are few symptoms or risk factors like obesity, age, family history, race etc. but still there are some examples where none of these symptoms may be observed.

Gestational Diabetes:

It is the third type of diabetes. It is defined as any carbohydrates intolerance recognized first time during pregnancy. The rate of such diabetes cases are increasing slowly. The patients have different characteristics of pregnancy than the normal pregnant women. There can be complications in the delivery which may results in fetal death, fetal macrosomia and growth disorder, neonatal hypoglycemia etc. There are multiple risk factors related to it like over obesity, family history, life style of patient, maternal age, miscarriage history which one has to consider.

Regular exercise, good healthy diet and insulin dose may control the diabetes but total cure is as good as impossible. Otherwise normal diabetes can lead to various problems such as [4]:

- a) Kidney failure(nephropathy) here the filtering system is damaged and ends in kidney diseases,
- b) Cardiovascular diseases where the arteries gets narrowed or chest pain with heart attack,
- c) Nerve damage (neuropathy) in which the blood capillary gets damaged which may cause tingling, numbness etc. It may even affect the digestion system.

- d) Eye damage (retinopathy) can lead to cataract or glaucoma, in some cases even leads to total blindness.
- e) Foot damage: severe infections happen if unattended can lead to foot or leg amputation.
- f) Alzheimer's disease: diabetes increases the risk of diabetes.

So it is very much important to detect the diagnosis in very early stages to avoid further complications. Data mining is such instrument which can help us diagnosing the diabetes in its earliest stage.

Data Mining is technique of finding unknown patterns and trends from large datasets. Data mining is the process of finding and using information to build a model using vast data models to uncover previously unknown patterns [5]. Thus data mining can be used to find out the different patterns and trends from the dataset so that it could predict out the diabetes accurately.

2. Related Work

Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni [6] performed a series of experiments by keeping the dataset same to perform predictive data mining. Medical data mining had great potential to churn various unknown pattern from the data, but there are some basic problems need to be addressed like the data is found to be heterogeneous, unorganized and voluminous. The authors also point out importance of the automation of such predictive tasks which may save high costs, bias towards patients and manual errors. The research shows that accuracy of algorithms like Decision tree, Bayesian classification can be increased a lot when they are combined with genetic algorithms.

Riccardo Bellazzi, Paolo Magni, and Giuseppe De Nicolao[7] worked on live data of 14 years old male patient, who was on insulin therapy i.e. 3 doses of insulin thrice a day. His blood glucose level was taken before all his three lunch. This data was applied to Naïve Bayesian algorithm. This method can be easily used as off line tool for blood glucose monitoring of patient or for telemedicine approach. It was novel approach to extract the structural component like daily trend and pattern from time series data using Naïve Bayesian Approach.

The fractal dimension of diabetic patient's retinal vascular distribution is more than normal human's. Shu-Chen Cheng and Yueh-Min Huang [8] has used this theme for their research. Its image processing technique to show fractal dimension and measure of lacunarity has a role to play in classification. There were 4 different approaches used in this study, back propagation algorithm, genetic algorithm, and radial basis function and voting scheme.

In another study, Nahla H. Barakat, Andrew P. Bradley, and Mohamed Nabil H. Barakat [9] propose support vector machines (SVM) as the tool for diagnosis of diabetes. The paper also talks about accuracy of SVM for the prediction. The SVM and the rules extracted from it can be proved as the second opinion. The research put forward some of the risk factors like obesity. The data shows that the waist circumference for male and female is deciding splitter for classification. Based on such outcome early predictions can be made easily.

K. R. Lakshmi and S.Prem Kumar [10] in their study, mention in detail the process of data mining. They also mentions the various data mining methods based on their functionality. Using Tanagra as data mining tool, the comparison was made amongst the different data mining algorithm. The algorithms are filtered on various grounds like lowest computing time(<550), positive precision values(>0.1), cross validation error rate(<0.3), bootstrap validation error rate(<0.29). The entire aim was to get best algorithm in terms of highest accuracy and lowest computing rates. The research shows that PLS-DA is the best in comparison. The researchers have considered PIMA Indian dataset for this study.

Shankaracharya, Devang Odedra, Subir Samanta, and Ambarish S. Vidyarthi [11] mention the importance of early detection of diabetes and for that various scores has to be devised. They mentions about supervised learning algorithm called mixture of experts(ME) where each problem is divided into subtasks and solved by simple network expert. When compared this algorithm has given maximum accuracy. The researchers also talked about the practical problems in the data set. As the statistical models are time dependent, even though the model works well for readymade data but the same might not work well with real or live data. It is of very much importance that the medical practitioners try to maintain the clinical data so that the even other practitioners will have a instant predictions as a second opinion.

Rashedur M. Rahman, Farhana Afroz [12] has made research to compare the different data mining techniques for diagnosis of diabetes. The dataset was kept same but they used different tools for data mining. Weka was used as performance measurer for the algorithms like Multilayer Perceptron Neural Network, Bayes Network classifier, J48graft, JRip and Fuzzy Lattice reasoning whereas the algorithm like MLP, Naïve Bayes, C4.5 were conducted on Tanagra and lastly MATLAB was used for experimentation on algorithms like FIS, ANFIS. The data set used was divided into 2 parts with the ratio of 66:34 for training and testing respectively. The analysis shows that the highest accuracy was gained by J48 that is 81.33% in comparison FLR which got a accuracy of only 51.43% was lowest accuracy. Amongst the performance of tools Tanagra was the best. When FIS and ANFIS were compared ANFIS gives more accuracy.

3. Conclusion

In recent past number of machine learning techniques have been used to predict the diagnosis of diabetic patients. Few important parameters are body mass index, waist ratio, weight, family history etc.

In literature, systems with higher accuracy on sample data were having few drawbacks on real world data. Also, some techniques report higher false positive rates. So, accuracy and false positives remain major concern. Latest techniques like fuzzy logic or neural networks have given promising results in many domains. Such techniques need to be experimented in more details for medical domain in near future.

4. References

- [1] Matthias Ploug Larsen and Signe Sørensen Torekov, "Glucagon-Like Peptide 1: A Predictor of Type 2 Diabetes," *Journal of Diabetes Research*, p. 13, 2017.
- [2] A. Ramachandran, "Know the signs and symptoms of diabetes," vol. 140, no. 5, 2014.
- [3] American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus," vol. 37, no. 1, 2014.
- [4] clinical staff mayo. (1998-2017) www.mayoclinic.org. [Online]. HYPERLINK "<http://www.mayoclinic.org/diseases-conditions/diabetic-neuropathy/basics/complications/con-20033336>"
- [5] Hian Chye Koh and Gerald Tan, "Data Mining Applications in healthcare," *Journal of Healthcare Information Management*, vol. 19, no. 2.
- [6] Ujma Ansari, Dipesh Sharma, Sunita Soni Jyoti Soni, "Predictive Data Mining for Medical Diagnosis: An Overview," *International Journal of Computer Applications (0975 – 8887)*, vol. 17, no. 8, March 2011.
- [7] Paolo Magni, and Giuseppe De Nicolao Riccardo Bellazzi, "Bayesian Analysis of Blood Glucose Time Series from Diabetes Home Monitoring," *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. 7, p. 47, July 2000.
- [8] Shu-Chen Cheng and Yueh-Min Huang, "A Novel Approach to Diagnose Diabetes Based on the Fractal Characteristics of Retinal Images," *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, vol. 7, no. 3, September 2003.
- [9] Andrew P. Bradley, and Mohamed Nabil H. Barakat Nahla H. Barakat, "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus," *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, vol. 14, no. 4, July 2010.
- [10] K. R. Lakshmi and S.Prem Kumar, "Utilization of Data Mining Techniques for Prediction of Diabetes Disease Survivability," *International Journal of Scientific & Engineering Research*, vol. 4, no. 6, June 2013.
- [11] Devang Odedra, Subir Samanta, and Ambarish S. Vidyarthi Shankaracharya, "Computational Intelligence in Early Diabetes Diagnosis: A Review," *The review of Diabetic Studies*, vol. 7, no. 4, December 2010.
- [12] Farhana Afroz Rashedur M. Rahman, "Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis," *Journal of Software Engineering and Applications*, vol. 6, pp. 85-97, March 2013.