International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

A Survey on Prediction of Diabetes using Data Mining

Shravani S. Shinde, Rohini M. Rajmane, Shradhda S. Chindage, Shweta S. Gundale, Uday B. Mane (Asst. Prof.)
Computer Science & Engineering, Shivaji University/Sanjay Ghodawat Institute, Atigre/ Kolhapur, Maharashtra, India

ABSTRACT

With recent advances in computer technology, large amounts of data will be collect and store, but all this data becomes more useful when it is analyze and some dependencies and correlations has found. This should be an accomplished with machine learning algorithms. The automatic diagnosis of diabetes is an important real-world medical problem. Detection of diabetes in its early stages is the key for treatment.

A study of existing systems reveals that it is important to discover the hidden knowledge from a particular dataset to improve the quality of health care for diabetic patients. Techniques used in existing system's are namely: EM algorithm, H-means+ clustering and Genetic Algorithm, for the classification of the diabetic patients. In addition, Ant Colony Optimization (ACO) has used in data mining field to extract rule based on classification systems. Survey work shows that there are data mining algorithms which can be used in the analysis to develop a more efficient prediction technique.

Keywords: Classification, Data Mining, Decision Tree.

I. INTRODUCTION

Effects of diabetes have been reported to a more fatal and worsening impact on women than on men because of their lower survival rate and poorer quality of life. World Health Organization (WHO) reports state that almost one –third of the women who suffer from diabetes have no knowledge about it. The effect of diabetes is unique in case of mothers because the disease is transmitting to their unborn children. Strokes, miscarriages, blindness, kidney failure and amputations are just some of the complications that arise from this disease. The analyses of diabetes cases have been restricted to pregnant women.

Nowadays, large amount of information is collect in the form of patient records by the hospitals. Data mining is analysis technique that helps in proposing inferences. This method helps in decision-making through algorithms from large amounts of data generated by these medical centers. Considering the importance of early medical diagnosis of this disease, data mining techniques can applied to help the women in detection of diabetes at an early stage, which may help in avoiding complications.

Types of Diabetes

The two main types of diabetes are as follow:

Type 1: Though there are only about 10% of diabetes patients have this form of diabetes, recently, there has been a rise in the number of cases of this type in the United States. The disease manifest as an autoimmune disease occurring at a very young age of below 20 years hence also called juvenile-onset diabetes. In this type of diabetes, the pancreatic cells that produce insulin have destroyed by the defense system of the body.

Type 2: This type accounts for almost 90% of the diabetes cases and commonly called the adult-onset

diabetes or the non-insulin dependent diabetes. In this case, the various organs of the body become insulin resistant, and this increases the demand for insulin.

II. LITERATURE SERVEY

[1] Diabetes is most dangerous disease in the 21st century in the world. Diabetes is affects the human lifetime. Diabetes is unique in case of mothers because the disease is transmitting to their unborn children. Diabetes may develop serious complications such as heart diseases, Strokes, blindness, Premature Death, and kidney failure. Now a day, large amount of information is collect from the rural and urban peoples in the form of patient records. Data mining is the term used to describe the process of extracting value from a database.

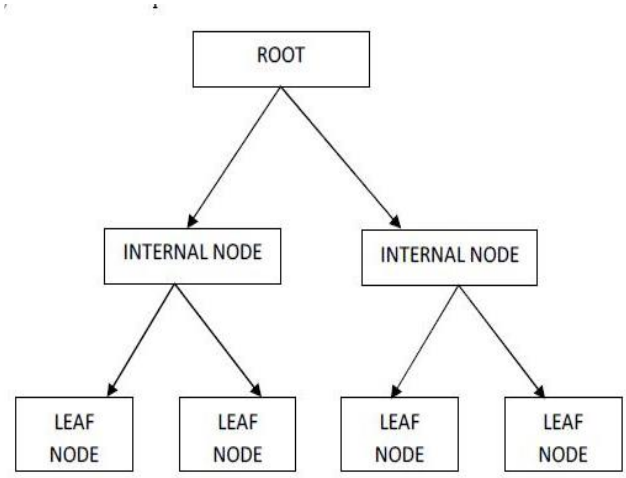

Figure 1. Shows a sample decision tree structure

Knowledge discovery for predictive purposes have done through data mining, which is an analysis technique that helps in proposing inferences. This method helps in decision-making through algorithms from large amounts of data.

Figure 1 depicts the decision tree is a tree structure, which is in the hierarchical form of a flowchart. It has used as a method for classification and prediction with representation using nodes. The root and internal nodes are the test cases that have used to separate the instances with different features. Internal nodes themselves are the result of attribute test cases. Leaf nodes denote the class variable.

[2] Ant colony optimization (ACO) has used successfully in data mining field to extract rule based classification systems. Diabetes is one of the most dangerous diseases, named Silent killer. This disease is a major health problem in both industrial and developing countries, and its incidence is rising. It is a disease in which, either the body does not produce enough insulin or the cells ignore the insulin. Insulin is necessary for the body to be able to use glucose for energy. Diabetes increases the risk of blindness, blood pressure, heart disease, and kidney disease and nerve damage. This disease has two main type's type1 and type 2. The most usual form of diabetes is diabetes type 2 or Diabetes mellitus type 2. Millions of people have diagnosed with diabetes type 2,and unfortunately, many more are unaware that they are at high risk.

[3] Medical information systems in modem hospitals and medical institutions become larger and larger; it causes great difficulties in extracting useful information for decision support. Traditional manual data analysis has become inefficient and methods for efficient computer based analysis are essential. It has proven that the benefits of introducing machine learning into medical analysis are to increase diagnostic accuracy, to reduce costs and to reduce human resources. Artificial Neural Networks (ANN) is currently the next promising area of interest. Already it could successfully apply to various areas of medicine such as diagnostic systems, bio chemical analysis, and image analysis and drug development. The benefit of using ANN is that they are not affect by factors such as fatigue, working conditions and emotional state.

[4] Data mining and machine learning algorithms in the medical field extracts distinctive concealed patterns from the medical data. They can utilized for the examination of vital clinical parameters,

expectation of different diseases, estimating assignments in pharmaceutical, extraction of medical knowledge, treatment planning support and patient administration. Data mining gives a diversity of methods to investigate large data keeping in mind the end goal to find hidden knowledge. This study is an effort to plan and execute a descriptive data mining approach and to devise association standards to envisage diabetes behavior in arrangement with particular life style parameters, including physical activity and emotional states, especially in elderly diabetics.

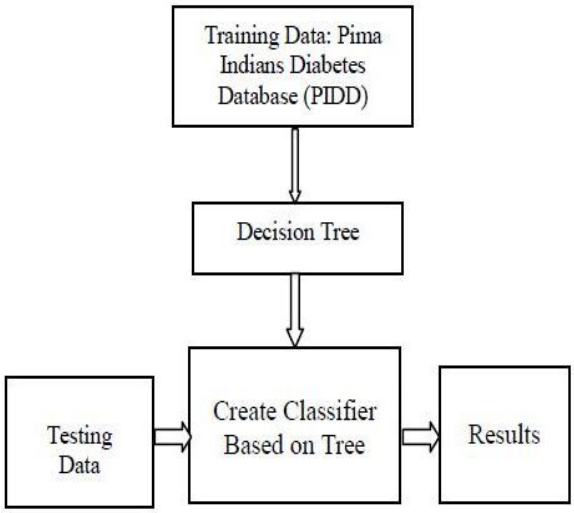

Figure 2. Algorithmic Sequence of Random Forest Classifier

A random forest classifier is the assembly of tree-structured classifiers (Figure 2). This algorithm supplements the objects from array of input to every tree of forest. The elements of the unit vector have individually voted for classification by every single tree. The forest filters the most voted classifications out of the forest.

Random Forest Classifier based approach outperforms better with the accuracy of 99.7%. Future work will direct on the use of hybrid classification algorithms to enhance the overall performance.

[5] Decision trees are few of the most extensively researched domains in Knowledge Discovery. Irrespective of such advantages as the ability to explain the choice procedure and low computational costs, decision trees also usually produce relatively great outcomes in assessment with other machine finding out formulas. A new decision tree algorithm based on J48 and reduced error pruning. Tree obtained is fast decision tree learning and based on the information gain or reducing the variance.

III. CONCLUSION

In the survey work methodologies are used for prediction of diabetes. A study conducted in order to discover the methodologies and techniques used in existing systems. We found that it is necessary to discover the hidden knowledge from a particular dataset to improve the quality of health care for diabetic patients. Different algorithms and technique used for prediction and analysis of diabetes. These algorithms are the data mining algorithms differentiated based on working, complexity and accuracy. The survey work clearly shows that there are some new algorithms, which are not yet applied in this domain. Based on the study of data mining algorithms there is scope for improving upon the diabetes prediction techniques.

The survey work will be extended to implement the new algorithms for the development of efficient prediction techniques.

IV. REFERENCES

[1]. Sankaranarayanan. S. and Dr. Pramananda Perumal. T., "Predictive Approach for Diabetes Mellitus Disease through Data Mining Technologies", World Congresson Computing and Communication Technologies, 2014, pp. 231-233

[2]. Mostafa Fathi Ganji and Mohammad Saniee Abadeh, "Using fuzzy Ant Colony Optimization for Diagnosis of Diabetes Disease", Proceedings of ICEE 2010, May 11-13, 2010

[3]. T. Jayalakshmi and Dr. A. Santhakumaran, "A Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural

Networks",International Conference on Data Storage and Data Engineering, 2010, pp. 159-163

[4]. Sonu Kumari and Archana Singh, "A Data Mining Approach for the Diagnosis of Diabetes Mellitus", Proceedings of 71hlnternational Conference on Intelligent Systems and Control (ISCO 2013)

[5]. Neeraj Bhargava, Girja Sharma, Ritu Bhargavaand Manish Mathuria, Decision Tree Analysis on J48Algorithm for Data Mining. Proceedings of International Journal of Research in Computer.