



Review of Cardiovascular Disease in Diabetic Patients using Data Mining Techniques

Charu .V. Verma¹, Dr. S. M. Ghosh²
Research Scholar¹, Professor²
Department of CSE

Dr. C.V. Raman University, Bilaspur, C.G, India

Abstract:

Data mining is the process of finding useful and relevant information from the databases. There are several types of mining techniques available and applicable magnificently on various sectors. Data mining process may take important role in Health care sector. The healthcare sector assembles enormous quantity of medical data which cannot extract to uncover hidden information for making effective decision. Diabetes mellitus is a chronic condition that occurs when the body cannot produce enough or effectively use of insulin. Patients with diabetes mellitus have a considerably higher risk of cardiovascular disease. Most of this excess risk is associated with an augmented prevalence of well-known risk factors such as hypertension, hyperglycemi and obesity in these patients. Individuals with insulin resistance or diabetes in combination with one or more of these risk factors are at even greater risk of heart disease. Prediction of heart disease in diabetic patient is a challenging task. In this paper, we present a review of the research work being carried out using different data mining techniques to enhance diagnosis and prediction of heart disease in diabetic patient.

Keywords: Data mining techniques, Diabetes, Cardiovascular disease.

I. INTRODUCTION

The main objective of our paper is to review the different techniques of data mining used for the prediction of cardiovascular diseases in diabetic patient by using different mining tools. Diabetes Mellitus is commonly referred as Diabetes. Diabetes is one of the major health challenges all over the world. The prevalence of diabetes is increasing at a fast pace. The World Health Organization (WHO) estimated there were 30million people who had diabetes worldwide in 1985. This estimation was increased to 135 million by 1995 and reached 217 million in 2005. By the year 2030 WHO predicts that the number will be increase at least 366 million [9].Diabetes is one of the biggest concerns in India, as the figures are quite astounding. Prevention and prediction of diabetes mellitus is increasingly gaining interest in healthcare community. Most of the food we eat is turned into glucose or sugar for energy. The pancreas, an organ makes a hormone called insulin to help glucose get into the cells of our bodies. When a body is affected with diabetes, it couldn't make enough insulin or couldn't use its own insulin. This causes sugar to build up into blood. Diabetes is a life threatening disease in rural and urban, then developed and under developed countries. The common symptoms for the diabetic patients are frequent urination, increased thirst, weight loss, slow-healing in wound, giddiness, increased hunger etc. In order to reduce the population with diabetes mellitus it should be detected at an earlier stage, hence a quick and efficient detection mechanism has to be discovered. Diabetes is caused due to Genetic susceptibility, Lifestyle, Eating habits, Profession and Environment etc.

Types of diabetes [11]

- Type 1Diabetes is called insulin-dependent diabetes mellitus (IDDM) or juvenile-onset diabetes. It often begins in childhood. Type 1diabetes is an autoimmune condition where

the immune system wrongly identifies and subsequently attacks the pancreatic cells that produce insulin, leading to little or no insulin production.

- Type 2 Diabetes is called non-insulin-dependent diabetes mellitus (NIDDM) or adult-onset diabetes. In the type 2 diabetes, the body usually still produces some insulin, but this is not enough to meet demand and the body cells do not properly respond to the insulin. The latter effect is called insulin resistance, where persistently elevated blood glucose has caused cells to be overexposed to insulin, making them less responsive or unresponsive to the hormonal messenger.

- Gestational Diabetes occurs in pregnancy and typically resolves after childbirth. People who have experienced gestational diabetes do, however, have an increased risk of developing type 2diabetes after pregnancy. Various medical investigations report that if blood sugar is consistently high, over time it can affect the heart, eyes, kidneys, nerves, and other parts of the human body. These problems are called complications. Sometimes people with diabetes don't realize that they have the disease until they begin to have other health problems.

II. CORRELATION BETWEEN DIABETES AND CARDIAC DISEASES

Heart is an essential part of our body. Various factors increase the risk of Heart disease. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases. In 2008, 17.3 million people died due to Heart Disease. Over 80% of deaths in world are because of Heart disease. WHO estimated by 2030, almost 23.6 million people will die due to Heart disease [10]. The number of people with diabetes around the World is ever increasing, so the outlook for heart disease is becoming even more alarming. The following statistics of The American

Heart Association speaks loud and clear that there is a strong correlation between Heart disease and diabetes. They consider diabetes to be one of the seven major controllable risk factors for cardiovascular disease [12]. Heart disease is the main cause of death and disability among people with type 2 diabetes. At least 68 percent of people age 65 or older with diabetes die from some form of heart disease. Adults with diabetes are two to four times more likely to have heart disease than adults without diabetes. Most of the diabetic patients are affected by metabolic and hormonal disorders. Some of the risk factors of heart diseases among diabetic patients are lack of physical inactivity, alcohol consumption, tobacco chewing, smoking, obesity etc. The biological link mechanisms behind heart diseases and diabetes are hypertension, hyperglycemia and inflammation in the blood. Very high blood sugar damages blood vessels can lead to blockage. People with diabetes have two to four times risk of developing heart disease. People with heart failure can live

longer and more active lives if the condition is diagnosed earlier and if they follow their treatment plans regularly. Several new techniques are used for diagnosing both heart diseases and diabetes. The machine learning techniques are one of the existing techniques which have a transparent diagnostic knowledge to diagnose diseases.

II. LITERATURE REVIEW

Large number of work has been done to find out efficient methods of medical diagnosis for various diseases. This paper aims to survey and analyze the various data mining techniques introduced in recent years for the predictions of cardiovascular disease in diabetic patients. Author of each paper used either single or multiple data mining techniques and find the accuracy for the same. Tabular representation of review is presented here in Table I.

Table.I. table shows different data mining techniques used for the predictions of Cardio vascular diseases in diabetic patients

S.No	Author	Year	Techniques used	Objective
[1]	Parthiban, G. et. al	2012	Naïve bayes, SVM	Applying machine learning methods in diagnosing heart diseases for diabetic patients
[2]	Radha, P. et. al	2014	IFCM-SVM	Diagnosing heart diseases for Type 2 diabetic patient by cascading the data mining techniques
[3]	Kalaiselvi, C. et. al	2015	AGKNN, ANFIS	Prediction of heart diseases and cancer in Diabetic patients using data mining techniques
[4]	Pawar, P. R. et. al	2015	GNN	Prediction of heart disease for diabetic patients using GNN
[5]	Rajasekaran,T.et. al	2016	J48, Naïve Bayes	Heart attack Prediction for diabetic patient using data analysis
[6]	Phul, J. S. et. al	2016	J48, Naïve Bayes	Diagnose and predict Diabetes heart diseases using Data mining classification techniques
[7]	Gokilam, G.G. et. al	2016	J48, Naïve Bayes, Random Tree, Random Forest, REP	Performance analysis of various data mining classification algorithms on diabetes heart dataset
[8]	Dhivya, S. et. al	2016	SAM Algorithm	Heart disease risk prediction strategy using combinatorial and genetic algorithm

III. TECHNIQUES AND TOOLS USED

Data mining application and its types

Data mining is the computer based process of extracting useful information from enormous sets of databases. Data mining is

most useful in an exploratory analysis because of nontrivial information from large volumes of data. The usage of data mining techniques in disease prediction is to reduce the test and increase the accuracy of rate of detection. The data mining tasks can be classified generally into two types based on what a

specific task tries to achieve. Those two categories are descriptive tasks and predictive tasks as shown in figure 1.

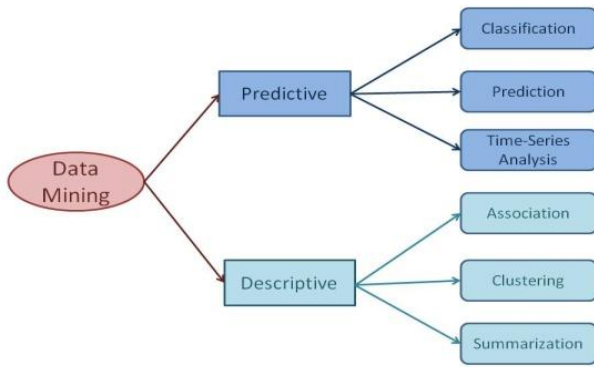


Figure.1. Data Mining Techniques

1. Predictive

Predictive data mining tasks come up with a model from the available data set that is helpful in predicting unknown or future values of another data set of interest. A medical practitioner trying to diagnose a disease based on the medical test results of a patient can be considered as a predictive data mining task.

Predictive Data Mining technique is categorized into three types:

- **Classification:** Classification derives a model to determine the class of an object based on its attributes. A collection of records will be available, each record with a set of attributes. Mainly classification is used to classify every item in a set of data into one of predefined set of classes or groups. Classification technique makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics.
- **Prediction:** Prediction task predicts the possible values of missing or future data. Prediction involves developing a model based on the available data and this model is used in predicting future values of a new data set of interest.
- **Time - Series Analysis:** Time series is a sequence of events where the next event is determined by one or more of the preceding events. Time series reflects the process being measured and there are certain components that affect the behavior of a process. Time series analysis includes methods to analyze time-series data in order to extract useful patterns, trends, rules and statistics

2. Descriptive

Descriptive data mining tasks usually finds data describing patterns and comes up with new, significant information from the available data set. A retailer trying to identify products that are purchased together can be considered as a descriptive data mining task.

Descriptive Data Mining technique is categorized into three types:

- **Association:** Association discovers the association or connection among a set of items. Association identifies the relationships between objects. Association analysis is used for commodity management, advertising, catalog design, direct marketing etc.

- **Clustering:** Clustering is used to identify data objects that are similar to one another. The similarity can be decided based on a number of factors like purchase behavior, responsiveness to certain actions, geographical locations and so on.

- **Summarization:** Summarization is the generalization of data. A set of relevant data is summarized which result in a smaller set that gives aggregated information of the data. Data can be summarized in different abstraction levels and from different angles.

Data mining tools

Since widespread use and complexity involved in building data mining applications, a large number of Data mining tools have been developed over decades. The development and application of data mining algorithms requires use of very powerful software tools. As the number of available tools continues to grow the choice of most suitable tool becomes increasingly difficult. Data mining tools predict future trends, behaviors, allowing business to make proactive, knowledge driven decisions. In this paper different data mining tools along with resultant accuracy of various above reviewed papers are discuss in table II.

- **WEKA:** Waikato Environment for Knowledge Analysis. Weka is a collection of machine learning algorithms for data mining tasks. These algorithms can either be applied directly to a data set or can be called from your own Java code. The Weka (pronounced Weh-Kuh) workbench contains a collection of several tools for visualization and algorithms for analytics of data and predictive modeling, together with graphical user interfaces for easy access to this functionality.

- **MATLAB:** MATLAB is a high-level scientific and engineering programming environment which provides many useful capabilities for plotting and visualizing data and has an extensive library of built-in functions for data manipulation. It is a programming language developed by Math Works. It provides functions for integrating MATLAB based algorithms with external applications and languages such as C, Java, .NET and Microsoft Excel.

- **R:** Revolution is a free software programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed.

- **RAPIDMINER:** Rapid miner is a software platform developed by the company of the same name that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics. Rapid Miner uses a client/server model with the server offered as Software as a Service or on cloud infrastructures.

- **Orange:** Orange is a component-based data mining and machine learning software suite, featuring a visual programming front-end for explorative data analysis and visualization, and Python bindings and libraries for scripting. It includes a set of components for data preprocessing, feature scoring and filtering, modeling, model evaluation, and exploration

techniques. It is implemented in C++ and Python. Its graphical user interface builds upon the cross-platform framework.

Table .2. Table Shows Different Data Mining Tools Used For Prediction Of Heart Diseases In Diabetic Patients With Accuracy

S.No	Author	Technique used	Tools used	Attributes	Accuracy
[1]	Parthiban, G. et. al	Naïve bayes, SVM	WEKA	9	94.60%
[2]	Radha, P. et. al	IFCM-SVM		9	93.8%
[3]	Kalaiselvi, C. et. al	AGKNN, ANFIS	MATLAB	13	98%
[4]	Pawar, P. R. et. al	GNN	MATLAB	12	
[5]	Rajasekaran,T.et. al	J48, Naïve Bayes	WEKA	4	85.7%
[6]	Phul, J. S. et. al	J48, Naïve Bayes	WEKA	12	95.53%
[7]	Gokilam, G.G. et. al	J48, Naïve Bayes, Random Tree, Random Forest, REP	WEKA	7	95.45%
[8]	Dhivya, S. et. al	SAM Algorithm	JAVA	11	96.8%

IV. CONCLUSION AND FUTURE SCOPE

In this review paper, the different data mining techniques are discussed for the prediction of Cardiovascular Disease in diabetic patients. Different techniques were implemented on various available tools for data mining. From this review we can find that every paper used either single or multiple data mining techniques with different number of attributes to find accuracy. Mainly different classification and association techniques namely Naïve Bayes, Decision Tree, Neural Networks, SVM and SAM are focused in this review. This paper provides various results in the form of accuracy of different techniques. From this Research work, it shows that Neural Network provides more accuracy compared to other techniques used. Some Research paper provides additional points that in case of changing the number of attributes the performance may vary. Our future work will be on a common dataset, in which we will focus on improving the predication of cardio vascular diseases in diabetic patient by implementing and utilizing different data mining techniques and features. By this we will identify the accuracy of the implemented techniques and then check the efficient algorithm among them.

V. REFERENCES

[1]. G. Parthiban, S.K.Srivatsa, (2012), “Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients”, International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 3– No.7, and August 2012.

[2]. P. Radha, Dr. B. Srinivasan, (2014), “Diagnosing Heart Diseases for Type 2 Diabetic Patients by Cascading the Data Mining Techniques”, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 8 2503 – 2509.

[3]. C. Kalaiselvi, G. M. Nasira, (2015), “Prediction of Heart Diseases and Cancer in Diabetic Patients Using Data Mining Techniques”, Indian Journal of Science and Technology, Vol

8(14), DOI: 10.17485/ijst/2015/v8i14/72688, July 2015 ISSN, (Print): 0974-6846 ISSN (Online): 0974-564.

[4]. P.R.Pawar, D. Vora,(2015), “Prediction of Heart Disease for Diabetic Patients using Genetic Neural Network”, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169, Volume: 3 Issue: 7 5028 – 5030.

[5]. T.Rajasekaran, R.Tharani Priya, V.Pavithra, P.Sudharshanam, (2015), “Heart Attack Prediction for Diabetic’s Patient using Data Analytics”, South Asian Journal of Engineering and Technology, Vol.2, No.17 (2016) 103–110, ISSN No: 2454-9614.

[6]. J.S.Phul, S.Vatta, (2016), “Diagnose and Predict Diabetic Heart Diseases using Data Mining Classification Techniques”, International Journal of Advance Research in Science and Engineering, Vol. no. 5, Issue no. 11, November 2016, ISSN (O) 2319-8354 , ISSN(P)2319-8346.

[7].G.G.Gokilam, Dr K.Shanthi, (2016), “Performance Analysis of Various Data mining Classification Algorithms on Diabetes Heart dataset”, COMPUSOFT, An international journal of advanced computer technology, 5 (3), March - 2016 (Volume-V, Issue-III), ISSN 2320-0790.

[8]. Dhivya.S, E.M.Mercy, (2016), “HEART DISEASE RISK PREDICTION STRATEGY USING COMBINATORIAL AND GENETIC ALGORITHM”, International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-5, May 2016 ISSN: 2395-3470.

[9].S. Wild, G. Roglic, A. Green., “Global prevalence of diabetes: estimates for the year 2000 and projections for 2030,” Diabetes Care, vol. 27, no. 5, pp. 1047–1053.

[10]. C.S.Dangare, Dr.S. S. Apte, “A data mining approach for prediction of heart disease using neural networks”, International journal of computer engineering and technology, 2012.

[11]. H.Nichols, <http://www.medicalnewstoday.com/articles/7504.php>, Last updated: Fri 24 July 2015.

[12].http://www.heart.org/HEARTORG/Conditions/More/Diabetes/WhyDiabetesMatters/Cardiovascular-Disease-Diabetes_UCM_313865_Article.jsp#.WPbnZvmGPIU42297SHARE54