ISSN 2348 – 9928
IJAICT Volume 3, Issue 9, September 2016 Doi:01.0401/ijaict.2014.07.05 Published on 05 (9) 2016

EVALUATION ON RISK FACTORS AND COMORBIDITY CONDITIONS OF DIABETESUSING MINING ALGORITHMS AND SEARCHING METHODS

Ms. N. Kanagavalli
Assistant Professor,
Information Technology,
Saveetha Engineering College,
Chennai, Tamilnadu, India

Mr. A .Vijayaraj
Associate Professor,
Information Technology,
Saveetha Engineering College,
Chennai, Tamilnadu, India

Mr. B. Sakthi Saravanan
Assistant Professor,
Information Technology,
Saveetha Engineering College,
Chennai, Tamilnadu, India

Abstract— The World Health Organization estimates that by 2030 there will be approximately 450 million people with diabetes, associated with renal complications, heart disease, stroke and peripheral vascular disease. Our aim is to analyses the risk factors and co morbidity conditions to detect diabetes early. Using PubMed and EMBASE databases to identify and extract key information that describes aspects of developing a prediction model, sample size and number of events, risk predictor selection. Using these two methods, we identify the various attributes and assign a value to each and every attribute or parameter. Based on the parameters, the analyses of high risk factors of developing diabetes are identified using association rule mining.

Keywords—Risk Factors, Comorbidity, Searching Methods, Diabetes, Insert.

I. INTRODUCTION

Diabetes is a growing epidemic that affects millions of people in the world and they do not know that they have the disease. Diabetes leads to significant medical complications including ischemic heart disease, stroke, nephropathy, retinopathy, neuropathy and peripheral vascular disease. Multiple risk factors have been identified affecting a large proportion of the population [1]. For example, prediabetes (blood sugar levels above normal range but below the level of criteria for diabetes) is present in approximately 35% of the adult population and increases the absolute risk factors of diabetes 3 to 10 fold depending on the presence of additional associated risk factors, such as obesity, hypertension, hyperlipidemia, etc. Comprehensive medical management of this large portion of the population to prevent diabetes represents an unbearable burden to the healthcare system[2].

Association rules are implications that associate a set of potentially interacting conditions (e.g. high BMI and the presence of hypertension diagnosis) with elevated risk. The use of association rule is particularly beneficial, because in addition to quantifying the diabetes risk, they also readily provide the physician with a "justification", namely the associated set of conditions. This set of conditions can be used to guide treatment towards a more personalized and targeted preventive care or diabetes management.

Association rule mining (ARM) describes how two items are related using a special method of exploring patterns different from other analysis techniques. The association rule generated from ARM can formulate the relation between X and Y in the form of "X → Y" or "If X…, then Y…," and analyse it as "If item X exists, item Y coexists". A rule does not necessarily imply cause and effect. Instead, it identifies simultaneous occurrence between items in antecedent X and consequent Y.

ARM makes it possible to analyse the association between not only two diseases, but also among three or more comorbidities that can be calculated from existing statistics. One study revealed the accompanying diseases of attention deficit/hyperactivity[3] disorder by applying ARM to diagnostic data from the National Health Insurance Database of Taiwan. Another study analysed stroke and its comorbidity diseases by ARM. Therefore, the current study was conducted to determine the relations among complications, the various diseases that accompany diabetes and three or more comorbidities, using ARM based on large amounts of clinical data.

II. EXISTING SYSTEM

In existing system the information of diabetes patients and the drugs used by them are collected. This information is used for the analysis of risk factors [4]. Using association rule mining the risk factors are identified from the information of diabetes and the drugs used. Based on the association rule mining, the risk scores are generated. The risk score only provide a

quantification of the risk, not the impact of the risk. That is, it does not provide the information that what is the cause for the elevation of risk.

III. PROPOSED SYSTEM

In proposed system general information of patients are collected from different databases. Using search methods (PubMed and EMBASE), the information of diabetes patients are extracted from the database and the attributes of diabetes are listed. Then assigning values to each attribute to ensure whether the value lies within the range specified or outside the range. If the value lies within the range then the risk factors are identified and calculated. If the value lies outside the range, then there are no risk factors [5].

IV. METHODS

4.1 Embase

Embase (often styled EMBASE for Excerpta Medica dataBASE) is a biomedical and pharmacological database. Embase enables tracking, extracting and retrieval of information from the database. Embase conduct quick searches with keyword, advanced searches using Emtree and subheadings, use filters to limit results. The other features of Embase are the following:

- Includes graphic – based filters.
- Register with Embase to save searches and set up email alerts.
- Export citations to Refworks automatically.

Embase is also used for natural language searches, in – depth drug searches, for drug/pharmacy topics, to find conference abstracts, for extensive search limiting options etc.

4.2 PubMed

PubMed is a method used to compare the values of attributes. PubMed contains in-process citations, some older citations, etc. It is used for easy keyword searching, clinical queries, genetic topics. PubMed conduct quick searches with keywords. Keywords are automatically mapping to medical subject headings. It also uses filters to limit results. The other features of PubMed are:

- Includes "Related Citations", highlighting review articles.
- Export citations of Ref works.

PubMed is also used to map keywords to subjects easily, to explore the medical subject headings database as a standalone, and for clinical queries and genetic topics.

V. MINING ALGORITHMS

5.1 Association Rule Mining

Association rule mining is an important component of data mining. Association rules are an important class of methods of finding regularities/pattern in data. It perhaps the most important model invented and has been used in many applications. Such knowledge can assist in making decisions of particular interest. Basic objective of finding association rules is to find all co-occurrence relationship called associations.

The classic application of association rule mining is market basket data analysis, which aims to discover how items purchased by customers in a supermarket are associated. Association rule are of form $X \rightarrow Y$, where X and Y are collection of items and intersection of X and Y is null. A rule may contain more than one item in antecedent and consequent of rule. Every rule must satisfy two users specified constraints: one is measure of statistical significance called support and other is measure of goodness called confidence [6].

The key element that makes association rule mining practical is the minsup threshold. It is used to prune the search space and to limit the number of frequent itemsets and rules generated. Traditionally, association analysis has been considered as an unsupervised technique, so it has been applied for knowledge discovery tasks. Recent studies have shown that knowledge discovery algorithms such as association rule mining can be successfully applied for prediction in classification problems.

Association rule mining is an important area of data mining research and a comparatively [7] a younger member of data mining community. In addition to finding co-occurrence relation between items, which is basic objective, the algorithm has been applied for diverse applications.

5.2 FEKM Algorithm

Fast and Exact out-of-core K-Means Clustering (FEKM) requires only one or a small number of passes on the entire dataset, and provably produces the same cluster centres as reported by the original K-means algorithm. It is used to speed up the process of finding the cluster centre.

Clustering has been one of the most widely studied topics in data mining and k-means clustering has been one of the popular clustering algorithms. K-means requires several passes on the entire dataset, which can make it very expensive for large disk-resident datasets.

FEKM algorithm which typically requires only one or a small number of passes on the entire dataset, and provably produces the same cluster centers as reported by the original k-means algorithm. The algorithm uses sampling to create initial cluster

centers, and then takes one or more passes over the entire dataset to adjust these cluster centers. Experimental results from a number of real and synthetic datasets show speedup between a factor of 2 and 4.5, as compared to k-means.

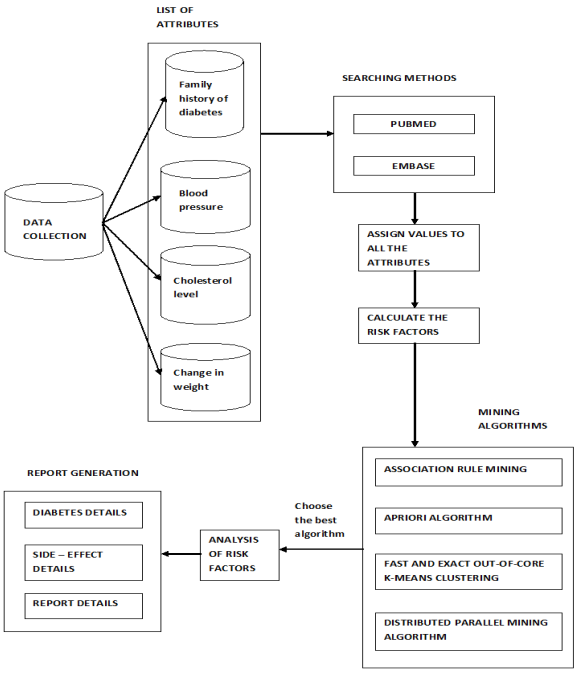

Fig 1: Report generation based on risk factor calculation

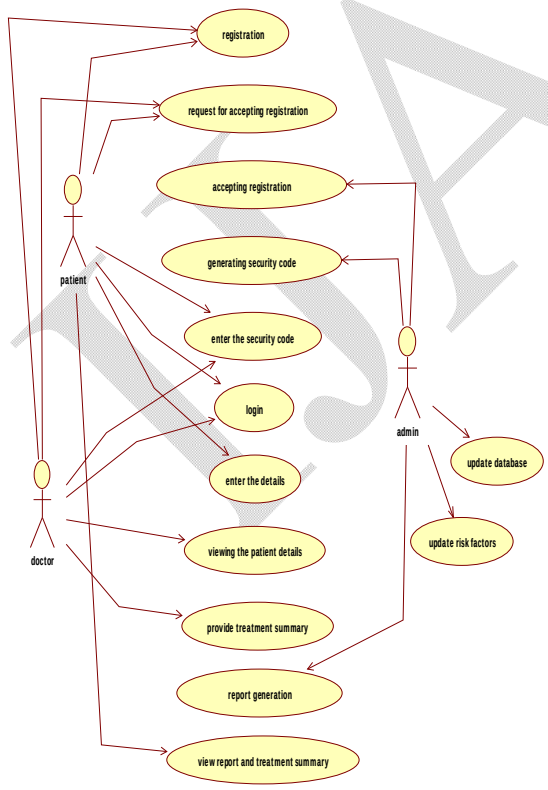

Fig 2: Class Diagram Report generation based on risk factor calculation

VI. CONCULSION

Finally generated report helps the doctor to provide better treatment for the diabetes patients who have high risk factors. The analysis of risk factors helps to identify the level of diabetes and their comorbidity conditions. The doctor can improve the treatment based on the risk level of diabetes.

References

[1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In VLDB Conference, 1994.

[2] Yonatan Aumann and Yehuda Lindell. A statistical theory for quantitative association rules. In Knowledge Discovery and Data Mining, 1999.

[3] Pedro J. Caraballo, M. Regina Castro, Stephen S. Cha, Peter W. Li, and Gyorgy J. Simon. Use of association rule mining to assess diabetes risk in patients with impared fasting glucose. In AMIA Annual Symposium, 2011.

[4] Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the united states, 2011.

[5] U.S. Department of Health and Human Services, Centers for Disease Control and Prevention http://www.cdc.gov/diabetes/ pubs/factsheet11 .htm, 2011.

[6] Bing Liu, Wynne Hsu, and Yiming Ma. Integrating classification and association rule mining. In ACM International Conference on Knowledge Discovery and Data Mining (KDD), 1998.

[7] Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England Journal of Medicine, 346(6), 2002.