Severity Estimation & Food Recommendation for Diabetic Patient using Data Mining

Sandeep Gupta[1] Ashish Gupta[2] Nilima Bhoir[3]
[1,2,3]Department of Computer Engineering
[1,2,3]Ideal Institute of Technology, Maharashtra-421303, India

Abstract— This paper analyzes the Diabetes prediction using various DM techniques. Some of the most important and popular data mining techniques are classification, clustering, prediction, Naive Bayes, Decision Tree are analyzed to predict the diabetes disease. Data mining plays an efficient role in prediction of diseases in health care industry. Diabetes is a metabolic disease where the improper management of blood glucose levels led to risk of generating abnormalities in functioning of critical organs like heart, kidney, eye etc. So the patients need to visit their specialist for observing sugar level related risks and sit tight for a day or more to get their conclusion report, they need to squander their cash futile as well. The proposed system calculates severity by using ID3 algorithm, Machine Learning approach based on the input given and at the same time it will recommend what food should be consumed by the patient to avoid health hazards. Data mining techniques can be used for early prediction of the disease with greater quality in order to save the human life and it will also reduce the treatment cost. According to International Diabetes Federation stated that 382 million people are affected with diabetes worldwide.

Key words: Data Mining, ID3 Algorithm, Asp.Net, Classification, Prediction

I. INTRODUCTION

Data mining is described as the process of discovering correlations, patterns and trends to search through a large amount of data stored in repositories, databases, and data warehouses. Humans in that sensitivity are limited by data overload so there are new tools and techniques are being progress to solve this problem through automation. Data mining adopts a sequence of pattern recognition technologies and statistical and mathematical techniques to identify the possible rules or relationships that govern the data in the databases. Diabetes mellitus is a chronic disease and a major public health challenge worldwide. It occurs when a body is not able to react or outgrowth properly to insulin, which is needed to maintain the rate of glucose. Diabetes can be controlled with the help of insulin injections, a healthy diet and regular exercise but there is no whole cure is available. Diabetes leads to many other diseases such as blindness, blood pressure, heart disease, and kidney disease and nerve damage. Severe diabetes may leads to develop high risk cause of death. There are three primary types of diabetes mellitus are as follows:

1) Type 1: Diabetes Mellitus results from the body fails to produce insulin. This form was previously specified to as insulin dependent diabetes mellitus.
2) Type 2: Diabetes Mellitus conclusion from insulin resistance which is a condition in which cells fail to use insulin properly, although for sometimes also with an absolute insulin deficiency. This type was previously specified to as noninsulin-dependent diabetes mellitus.
3) Type 3: Gestational diabetes occurs when a pregnant woman previously seem diagnosis of diabetes develops a high blood glucose level.

II. LITERATURE SURVEY

Gyorgy J. Simon, Pedro J.Caraballo [1] proposed the method of distributional association rule mining to identify sets of risk factors and the corresponding patient subpopulations that are significantly increased a risk of progressing to diabetes. And to discover sets of a risk factor here uses bottom up summarization algorithm which produces the most suitable summary that describes International Conference on Innovative Mechanisms for Industry Applications subpopulations at high risk of diabetes. The Subpopulation identified by this summary covered most high risk of patients, had low overlap and were at very high risk. This method is used for when the patient having high risk.

Dr. Zuber Khan, Shaifali Singh [2] worked on the concept of Diabetes Mellitus using k-Nearest Neighbor algorithm which is a most Important technique of Artificial Intelligence. The accuracy rate is showing that how many outputs of the data of the test dataset are same as the output of the data of different features of the trained dataset. The error rate is sighting that how many outputs of the data of the test dataset are not same as the output of the data of different features of the training dataset. The result they showed that as the value of k increases, accuracy rate and error rate will increase.

Mukesh Kumari and Dr.Rajan Vohra [3] worked on the concept of data mining is to extract knowledge from information stored in a dataset and generate the clear and understandable description of patterns. The techniques are attributes selection, data normalization and then classifier is applied on data set to construct the Bayesian model. Bayesian network classifier was proposed for the prediction of person weather diabetic or not. By using Bayesian classifier patient is undergoing classified in classes of Pre-diabetic, Non-diabetic, Diabetic according to the attributes selected. The techniques they applied as preprocessing attribute identification and selection, data normalization. And then classifier is applied to the modified data set to construct the Bayesian model. The Bayesian network has a benefit of it encodes all variables, missing data entries can be handled successfully.

A. Problem Definition

To compile the aggregated real time data set from reputed Hospitals International Journal of Pure and Applied Mathematics Special Issue 220.To Pre-process the clustered real time patient data for foot ulcer-diabetic prediction consuming both manual techniques and Preprocessing tools like Weka. To promote sustainable development, the smart city implies a global vision that merges artificial intelligence, big data, decision making, information and communication

technology (ICT), and the internet-of-things (IoT). The ageing issue is an aspect that researchers, companies and government should devote efforts in developing smart healthcare innovative technology and applications. In this paper, the topic of disease diagnosis in smart healthcare is reviewed. Typical emerging optimization algorithms and machine learning algorithms are summarized. Evolutionary optimization, stochastic optimization and combinatorial optimization are covered. Owning to the fact that there are plenty of applications in healthcare, four applications in the field of diseases diagnosis (which also list in the top 10 causes of global death in 2015), namely cardiovascular diseases, diabetes mellitus, Alzheimer' s disease and other forms of dementia, and tuberculosis, are considered. In addition, challenges in the deployment of disease diagnosis in healthcare have been discussed.

B. Proposed System

The following figure 1, shows a flow of Proposed work including Data set then identification of parameters are done like Age, BMI, HbA1C, FBG, PMBG. Then the proposed approach includes estimation of ranges for each test. After ranges are estimated by getting patient history records the Eclat Algorithm is applied which calculates control and Un-control condition of diabetes. After applying Eclat frequent patterns are obtained. By considering frequent pattern range the comparison is done with standard dataset forming Prediction of severity estimation on organs. The severity estimation of organs like heart, kidney, eye etc. The proposed approach gives the quantification of severity on organs of a body. Below shows a flow of proposed system.

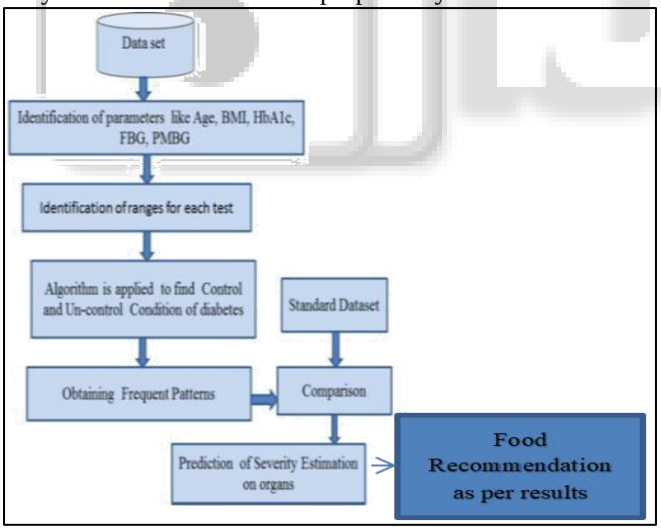

Fig. 1: Shows a Flow of Proposed Work

C. Future Scope

In Future work, more test will be performed for prediction and severity estimation. Some other different parameters are considered. Their might be other risk factors that did not consider, Factors include family history, smoking, metabolic syndrome, inactive lifestyle. By considering all other attributes more accuracy of prediction and quantification of severity estimation may be found out.

III. METHODOLOGY FOR IMPLEMENTATION

A. Data Set Collection

The Data set is collected from Pima Indian diabetes dataset containing various attributes like Age, Sex, BMI, Test Results of diabetes. Dataset is also made from test results of diabetic and nondiabetic patients. TheHbA1c, FBG, PMBG test results for diabetic patient forms monthly records to form the dataset. The Identification of Attributes can be done according to recent test reports of diabetic patient and Parameters are selected like Age, Gender, BMI, HbA1c, FBG, PMBG (test results).

B. Identification of Ranges

The ranges are identified according to Standard ranges for diabetes:

1) For HbA1c

The Normal range is below 5.7%, For Pre-diabetes the range lies between 5.7 to 6.4%, and for Diabetes it is greater than or equal to 6.5%. For FPG: The Normal range is greater than or equal to 100 mg/dl, For Pre-diabetes the range lies between 100 to 125 mg/dl, and for Diabetes it is greater than or equal to 126 mg/dl. For PMBG: The Normal range is greater than or equal to 140 mg/dl, For Pre-diabetes the range lies between 140 to 199 mg/dl, and for Diabetes it is greater than or equal to 200 mg/dlInput

C. ID3 Algorithm

The ID3 algorithm begins with the original set as the root node. On each iteration of the algorithm, it iterates through every unused attribute of the set and calculates the entropy (or information gain) of that attribute. It then selects the attribute which has the smallest entropy (or largest information gain value. The set is then split or partitioned by the selected attribute to produce subsets of the data. (For example, a node can be split into child nodes based upon the subsets of the population whose ages are less than 50, between 50 and 100, and greater than 100.) The algorithm continues to recur on each subset, considering only attributes never selected before.

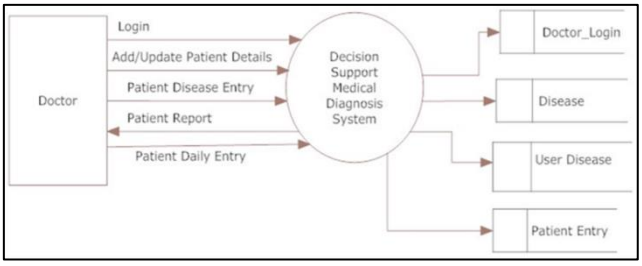

Fig. 2: The Algorithm use Above Procedures

Throughout the algorithm, the decision tree is constructed with each non-terminal node (internal node) representing the selected attribute on which the data was split, and terminal nodes (leaf nodes) representing the class label of the final subset of this branch.

IV. CONCLUSION

The proposed system used to find out prediction and severity estimation of diabetes. The Prediction can be done by whether the patient is Normal, Prediabetic, Diabetic. If the Patient lies in diabetic condition and had diabetes from

previous many years then the Eclat algorithm calculates Control and Un-Control Condition of diabetes, then quantification of Severity on an organ is calculated if the patient has Un-Control Condition of diabetes. Early Estimation of Severity may prevent developing high risk and may use to prevent death from severe diabetes.

REFERENCES

[1] Type-2 diabetes in India: Challenges and possible. solutions. Available from:http://www.apiindia.org/medicine_update_2013/chap40.pdf
[2] JaliMV, HiremathMB. Diabetes. Indian Journal of Science and Technology.
[3] 2010 Oct; 3(10). International Journal of Scientific and Research Publications, Volume 7, Issue 6, June 2017 709 ISSN 2250-3153 www.ijsrp.org
[4] LanordM, Stanley J, Elantamilan D, Kumaravel TS. Prevalence of Prehypertension and its Correlation with Indian Diabetic Risk Score in Rural Population. Indian Journal of Science and Technology. 2014 Oct; 7(10):1498–503.
[5] Han Kamber M. Data mining concepts and techniques.. 2nd ed. Amsterdam, Netherlands: Elsevier Publisher; 2006. p. 383–5.
[6] Han, Kamber M. Data mining concepts and techniques. 2nd ed. Burlington, Massachusetts: Morgan Kaufmann; 2006. p. 285–8.
[7] Gao, Denzinger J, James RC. CoLe: A cooperative data mining approach and its application to early diabetes detection. Proceedings of the 5th International Conference on Data Mining (ICDM'05); 2005
[8] Rajesh K, Sangeetha V. Application of data mining methods and techniques for diabetes diagnosis. International Journal of Engineering and Innovative Technology (IJEIT). 2012; 2(3):224
[9] AfrandP,Yazdani NM, Moetamedzadeh H, NaderiF,Panahi MS. Design and implementation of an expert clinical system for diabetes diagnosis. Global Journal of Science, Engineering and Technology; 2012. p. 23–31. ISSN:2322-2441.
[10] Adidela DR, Lavanya DG, Jaya SG, Allam AR. Application of fuzzy ID3 to predict diabetes. In AdvComput Math Sci. 2012; 3(4):541–5.
[11] PatilBM, Joshi RC, Toshniwal D. Association rule for classification of type2 diabetic patients. 2nd International Conference of IEEE on Machine Learning and Computing; 2010. p. 67. DOI 10.1109/ICMLC.
[12] AljarullahAA. Decision tree discovery for the diagnosis of type II diabetes. International Conference on Innovative in Information Technology; 2011. p. 303–7.
[13] Jaya Rama Krishnaiah VV, Chandra Shekar DV, Satya Prasad R, Rao KRH. An empirical study about type-2 diabetes suing duo mining approach. International Journal of Computational Engineering Research. 2012; 2(6):33–42.

