A SURVEY ON DIABETES DISEASE DIAGNOSIS USING VARIOUS DATA MINING TECHNIQUES

[1]Sathish G, [2]Manjula K
[1]Assistant Professor , [2]II MCA Student

[1] [2]Department of Computer Application,

[1] [2]Priyadarshini Engineering College, Vaniyambadi, Vellore, Tamilnadu, India

_____
_____

Abstract: Data mining looks through a large amount of data to extract useful information. Diabetes has affected over 425 million people World Wide with a majority of them being women. It is one of the World's major diseases. According to the WHO report by 2025 this number is expected or rise to over 380 million. This paper features various Data Mining techniques such as Classification, Clustering, and Association, pattern Recognitions, Diagnosis of Diabetes and also highlights related work to analyses and predict Diabetes disease diagnosis

*IndexTerms - **Data Mining techniques, Classification, Clustering, Association, Pattern Recognitions***.

_____

Introduction

Diabetes is a disease that occurs when your blood glucose also called blood sugar is too high insulin a hormone made by the pancreas helps glucose from food into your cells to be used for energy .A group of disease that result in too much sugar in the blood.

TYPES OF DIABETES DISEASE

Type1: More than 1 million cases per year in India. It caused by the immune system destroying the cells in the pancreas that make insulin. The symptoms are fast heart rate, headache, sleepiness or weight loss. Treatment can be help, but this condition can't be cured.

Type2: More than 10 million cases per year in India. A chronic condition that affect the way the body process blood sugar(glucose).The symptoms that are include increase thirst, frequent urination, hunger ,fatigue, and blurred vision.

Pre diabetes: More than 10 million cases per year in India. This causes it occurs when your body is not able to keep your sugar (glucose) at a normal level in your blood. It haves a no symptoms.

Gestational Diabetes: More than 1 million cases per year in India. A form of high blood sugar affecting pregnant women. A blood sugar test during pregnancy is used for diagnosis.
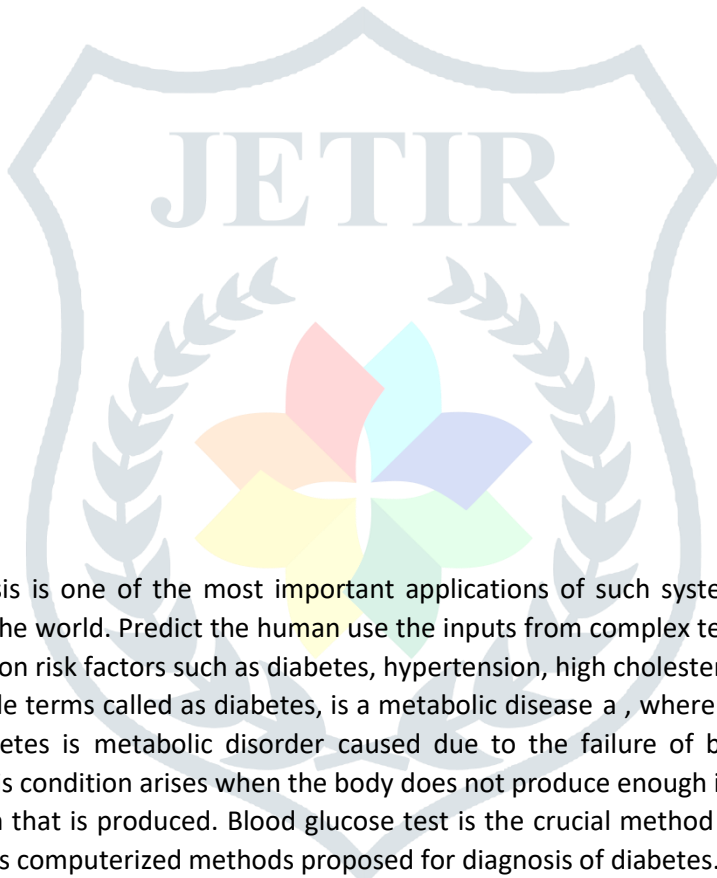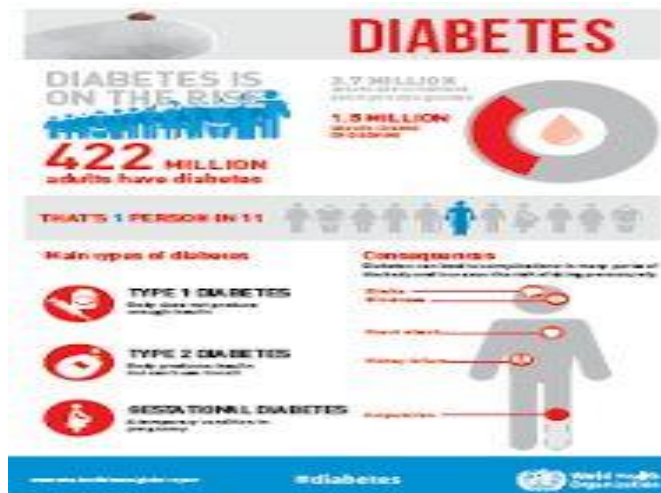
Literature Survey

Diabetes diagnosis is one of the most important applications of such system as it is one of the leading causes of deaths all over the world. Predict the human use the inputs from complex tests conducted in labs and also predict the disease based on risk factors such as diabetes, hypertension, high cholesterol, physical inactivity, obesity. Diabetes mellitus, in simple terms called as diabetes, is a metabolic disease a , where a person is affected with high blood glucose level. Diabetes is metabolic disorder caused due to the failure of body to produce insulin or to properly utilize insulin. This condition arises when the body does not produce enough insulin, or because the cells do not respond to the insulin that is produced. Blood glucose test is the crucial method for diagnosing diabetes. Also, there have been numerous computerized methods proposed for diagnosis of diabetes.

There are different kinds of data mining techniques are available. Classification, Clustering, Association and Pattern Recognitions.

Techniques

Diabetes diseases has become a common health problem nowadays, which would affect people and lead to various complications .

Classification

Classification is categorical and data is classified based on the training set and values, for prediction of Diabetes.

Multilayer Perceptron (MLP)

This is a commonly used neural network classification algorithm.

Architectures that use MLP during simulations on PIDD dataset consist of three layer feed-forward

neural network namely input, hidden, and output layer.

BayesNet

Bayesian networks are used with the presumptions that attributes are nominal with no missing values.

JRip

JRip to RIPPER is a basic and important algorithm in data mining classification.

C4.5

Is a Decision tree classifier to classify a new item and needs to create a decision tree using the

training data.

Fuzzy Lattice Reasoning (FLR)

This Classifier is used descriptive and decision-making.

Clustering

Clustering is different from classification; it does not have predefined classes. A large database is divided into number of small subgroups called clusters. Clustering algorithms discovers collections of the data such that objects in the same cluster are more identical to each other than other groups examined the gene expression data with support of hierarchical clustering approach by using genetic algorithm.

Association

Association also has great impact in the diabetes disease to discover the relationships between diseases, state of human health and the symptoms of disease. It used association in order to have used this integrated approach of association and classification for studying diabetes data. This integrated approach is useful for determining rules in the database and then by using these rules, an effective classifier is raised. The study made experiment on the data of diabetes.

Patients and generate rules by weighted associative classifier. Thus, Association also has an example influence in the healthcare field to identify the relationships among diabetic diseases, state of human health and the symptoms of disease.

Pattern Recognitions: Pattern recognition refers to the study of methods and algorithm for putting data object into categories. pattern recognition methods are used when designing practical systems.

DIABETES DISEASE DIAGNOSIS

This is used to detect diabetes or prediabetes. An cral glucose tolerant test measure your blood sugar you have gone at least eight your drink a glucose containing beverad. This test can be used to diagnosis diabetes or prediabetes. This helps lower blood sugar levels and improves insulin sensitively. Regular exercises along with a good diet can help to reduce cholesterol levels and high blood pressure. Many people can keep their blood glucose in a healthy range without medications if they lose weight and keep their weight down, are regularly physically active and follow a meal plan that helps them keep potion sizes under control.The way to lower blood sugar levels naturally exercise regularly,control your carb intake,increase fiber intakes, drink water and stay Hydrated,implement portion control.

RELATED WORK

Insulin is most important hormone in the body. It converts the sugar, starch and other food items in to energy needed for daily life. If the body does not produce insulin the redundant amount of sugar will be driven out by urination. This disease is called as Diabetes. Causes of diabetes are always a mystery even though obesity and lack of exercise plays a vital role. In early the ability to diagnose diabetes plays a major role in treatment process. Diabetes is common in both developed and developing countries. There was estimated 175 million people in 2000 and was expected to increase to 354 million by 2030. It is also given that by 2030, 85 percent of world's diabetic patients will be in developing countries. In India the diabetes count is expected to increase from 31.7 million in 2000 to 79.4 million in 2030.Many researchers have been made in the past decade to diagnose the diabetes mellitus. Applying data mining techniques in medical field is a typical task. The data mining begins with a hypothesis and the results are adjusted to fit the hypothesis in medical research. This differs from the standard data mining task in which datasets without apparent hypothesis is commonly used Diabetes mellitus decreases the resting of blood flows through the skin by disturbing regulation of skin microcirculation. Obesity and lack of exercise plays a significant role in cause for diabetes. The availability of health records and monitoring those leads to accumulation of data which are used by practitioners, health care decision makers, physicians. Since diabetes is a lifelong disease an individual patient record may be massive and difficult to handle. Poor generalization ability is a major issue of data mining in healthcare industry because of the lack of input data and processing .This proposed approach has gained 98.24% accuracy when compared to other data mining algorithms. Data Mining techniques can be used to predict the fast glucose level (FGL) thereby predicting the fluctuations in glucose level.

CONCLUSION

Type-1 and type-2 diabetes may lead to heart problems, kidney diseases and eye related problems. It is important to prevent or control gestational diabetes because Gestational Diabetes Mellitus (GDM) may go away after pregnancy, but women who have GDM seven times more are likely to develop type-2 diabetes than women who do not have GDM in pregnancy. The children of the GDM mother have the risk of obesity and type-2 diabetes. All of these difficulties can be handled by controlling the blood sugar levels. From this study, it was found out that data mining techniques can be used for predicting the type and risk levels of diabetes. Through this study it is found that the Data Mining techniques are important and it leads to valid approaches for predicting the risk of gestational diabetes. So it is our recommendation to use new techniques like data mining for decision making in medical fields, which improves the diagnosis of diseases like gestational diabetes. This research helps the doctors and health organizations in using the Data Mining techniques in the medical field which helps in predicting the type of diabetics and risks levels associated with it. Thus the proposed model helps in improving the diagnosis of the diseases which indeed helps in early cure of disease in the patients.

REFERENCES

[1] Mokdad AH, Tuffaha M, Hanlon M, El Bcheraoui C, Daoud F, et al. (2015) Cost of Diabetes in the Kingdom of Saudi Arabia, 2014. J Diabetes Metab 6: 575

[2] Wang, K. J., Adrian, A. M., Chen, K. H., & Wang, K. M. (2015). An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus. Journal of biomedical informatics, 54, 220-229 .

[3] M. Dura raj, G. Kalaiselvi, " Prediction Of Diabetes Using Soft Computing Techniques- A Survey", International Journal of Scientific & Technology Research, Vol. 4, Issue No.3, March 2015

[4] Ramesh Motka, Viral Parmar, Balbindra Kumar, A. R. Verma, " Diabetes Mellitus Forecast Using Different Data Mining Techniques", International conference on computer and Communication Technology .

[5] SonuKumari, Archana Singh, "A data mining approach for the diagnosis of diabetes mellitus" ,IEEE conference on intelligent systems and control ,pp-373-375,2013.

[6] RaksehMotka,ViralParmar, "Diabetes Mellitus Forecast Using Different Data mining Techniques IEEE International Conference on Computer and Communication Technology (ICCCT),2013.

[7] VeenaVijayan V., AswathyRavikumar, "Study of Data Mining algorithms for Prediction and Diagnosis of Diabetes Mellitus", International Journal of Computer Application, Vol 94,pp .12-16,June 2014

[8] GunasekarThangarasu, P.D.D. Dominic, "Prediction of HiddenKnowledge from Clinical Database using Data mining Techniques" International Conference on Computer and Information Sciences (ICCOINS), .1-5,2014.

[9] JefriJunifer Pangaribuan ,Suharjito "Diagnosis of Diabetes Mellitus Using Extreme Learning Machine ".pp-33-38,2014