ABSTRACT

Data mining techniques are used to find interesting patterns for medical diagnosis and treatment. Diabetes is a group of metabolic disease in which there are high blood sugar levels over a prolonged period. This paper concentrates on the overall literature survey related to various data mining techniques for predicting diabetes. This would help the researchers to know various data mining algorithm and method for the prediction of diabetes mellitus.

KEYWORDS: Diabetes Mellitus, Data mining, Prediction, Decision Tree, Classification

INTRODUCTION

Diabetes Mellitus is a chronic disease for which there is no known cure except in very specific situations management concentrates on keeping blood sugar levels as close to normal as possible without causing hypoglycemia. This can be controlled with diet, exercise and use of appropriate medications.

Diabetes Mellitus occurs throughout the world and it is more in developed countries. The increase in rates in developing countries follows the trend of urbanization and life style changes, including a “western-style” diet. This is because of less awareness.

The purpose of data mining is to extract useful information from large databases or data warehouses. Data mining applications are used for commercial and scientific sides [1].

Data mining is process of selecting, exploring and modeling large amounts of data in order to discover unknown patterns or relationships which provide a clear and useful result to the data analyst [2].

KDD process may consists several steps: like data selection, data cleaning, data transformation, pattern searching i.e. data mining, finding presentation, finding interpretation and finding evaluation [3].
**DIABETES OVERVIEW**

Diabetes Mellitus (DM) is a set of related diseases in which the body cannot regulate the amount of sugar in the blood. In a healthy person, the blood glucose level is regulated by several hormones, including insulin. Insulin is produced by the pancreas, a small organ between the stomach and liver. The pancreas secretes other important enzymes that help to digest food. Insulin allows glucose to move from the blood into liver, muscle, and fat cells, where it is used for fuel.

**CAUSES OF DIABETES**

Hereditary and genetics factors, Infections caused by viruses, Stress, Obesity, Increased cholesterol level, High carbohydrate diet, Nutritional deficiency, Excess intake of oil and sugar, No physical exercise, Overeating, Tension and worries, High blood pressure, Insulin deficiency, Insulin resistance.

**TYPES OF DIABETES**

**Type 1 Diabetes**

It usually starts in childhood or young adulthood. The body's immune system destroys the cells that release insulin, eventually eliminating insulin production from the body. Without insulin, cells cannot absorb sugar (glucose), which they need to produce energy.

**Type 2 Diabetes**

It can develop at any age and usually discovered during adulthood. Now it is found that increasing number of children are being diagnosed. This can be prevented or delayed with a healthy lifestyle, including maintaining a healthy weight with regular exercise.

**Gestational Diabetes**

Diabetes that is triggered by pregnancy is called gestational diabetes. It is often diagnosed in middle or late pregnancy period. High blood sugar levels in a mother are circulated through the placenta to the baby and it must be controlled to protect the baby's growth and development. It creates greater risk to mother and even to the unborn baby.

**METHODOLOGY**

Publications and journals has been analysed and data mining techniques which is given below have been applied for predicting diabetes.

**Decision Tree:** Decision tree is one of the popular and important classifier which is easy and simple to implement. It doesn’t have domain knowledge or parameter setting. It handle huge amount of dimensional data. It is more suitable for exploratory knowledge discovery. The results attained from Decision Tree are easier to interpret and read [4].

**Naive Bayes:** Naive In simple terms, a naive Bayes classifier assumes that the value of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 3” in diameter. A Naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features [4].

**K-nearest neighbor’s algorithm (k-NN):** is the one of the important method for classifying objects based on closest training data in the feature space. It is simplest among all machines learning algorithm but, the accuracy of k-NN algorithm can be degraded by presence of noisy features [5].

**Classification via Clustering:** Clustering is the process of grouping same elements. This technique may be used as a preprocessing step before feeding the data to the classifying model. The attribute values need to be
normalized before clustering to avoid high value attributes dominating the low value attributes [6].

A clinical Decision Support System based on OLAP with data mining to diagnose whether a patient can be diagnosed with diabetes with probability high, low or medium. The system is powerful because it discovers hidden patterns in the data and can, it enhances real-time indicators and discovers bottlenecks and it improves information visualization [7].

**Neural Network:** An artificial neural network (ANN), often just called a "Neural network" (NN), is a mathematical model or computational model based on biological neural network. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem [9].

In medicine, ANNs have been used to analyze blood and urine samples, track glucose levels in diabetics, determine ion levels in body fluids and detect pathological conditions [10].

Artificial Neural networks are well suited to tackle problems that people are good at solving, like prediction and pattern recognition. Neural networks have been applied within the medical domain for clinical diagnosis, image analysis and interpretation [10], signal analysis and interpretation and drug development [11].

**CONCLUSION**

Different approaches for the prediction of Diabetes Mellitus and its types are concentrated in this study. Data mining is a technique used to extract useful information from existing large volume of data which enable us to gain more knowledge. In this way data mining techniques are applied in health care sector in order to predict various diseases and to find out efficient ways to treat them as well.

**REFERENCES**