Influence of Exercise in Diabetes Mellitus Prediction in Big Data Using Hadoop/Map Reduce Frame work

[#]Praveenkumar K S, [*]Dr. R Gunasundari
[#]*PhD Research Scholar, KAHE, Coimbatore (Asst. Professor, Dept. of CS, SNGIST, Kochi),*
[*]*Associate Professor, Dept. of CS, CA & IT, KAHE, Coimbatore*

Abstract: *Now a day's disease prediction is a developing area of research in the healthcare. There are number of diseases are still cannot be predictable. There are four main types of disease: infectious diseases, deficiency diseases, hereditary diseases (including both genetic diseases and non-genetic hereditary diseases), and physiological diseases. In other words any disorder or malfunctioning of the body or mind that destroys good health can be called as a disease. The status of health of the body in a disease is said to be compromised. A disease can be caused due to a variety of reasons. Every disease has characteristic symptoms through which we can identify the types of diseases. Early prediction and proper treatments can possibly stop, or slow the progression of disease.*
In the proposed study we consider the disease Diabetes Mellitus (DM), predictions using Big Data Tools. Although many data mining techniques have been applied to assess the main causes of diabetes, but only few sets of clinical risk factors are considered. In this study, theproposed system that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. Here we evaluate many factors Hereditary and genetics factors, Stress, Body Mass Index, Increased cholesterol level, High carbohydrate diet, Nutritional deficiency, Nature of Exercises, Tension and worries, High blood pressure, Insulin deficiency, Insulin resistance. Then we evaluate and compare this system using suitable rules and Map Reduce algorithm. The performance of the system is evaluated in terms of different parameter like rules used, classification accuracy, and classification error. By considering all these parameters, the system can predict diabetics in a great accuracy. This paper mainly discuss on the non-clinical parameter Exercise and its influence on diabetics.

Keywords*: Diabetes Mellitus; Big Data, Hadoop/Map Reduce, Clinical factors, Non-clinical factors*

I. Introduction

Diabetes mellitus (DM), commonly known asdiabetes, in which the body's ability to produce or respond to the hormone insulin is impaired, resulting in abnormal metabolism of carbohydrates and elevated levels of glucose in the blood.
Type 1 diabetes (T1D), also known as juvenile diabetes, is a form of diabetes in which very little or no insulin is produced by the pancreas. Before treatment this results in high blood sugar levels in the body. Type 2 diabetes, formerly known as adult-onset diabetes, is a form of diabetes that is characterized by high blood sugar, insulin resistance, and relative lack of insulin. Type 2 diabetes primarily occurs as a result of obesity and lack of exercise. Some people are more genetically at risk than others. Type 2 diabetes makes up about 90% of cases of diabetes, with the other 10% due primarily to type 1 diabetes and gestational diabetes. Type 2 diabetes is largely preventable by staying a normal weight, exercising regularly, and eating properly. Treatment involves exercise and dietary changes. Onset of type 2 diabetes can be delayed or prevented through proper nutrition and regular exercise
Type 2 DMbegins with insulin resistance, a condition in which cells fail to respond to insulin properly. As the disease progresses a lack of insulin may also develop. The most common cause is excessive body weight and insufficient exercise.Gestational diabetes is the third main form, and occurs when pregnant women without a previous history of diabetes develop high blood sugar levels.Prevention and treatment involve maintaining a healthy diet, regular physical exercise, a normal body weight, and avoiding use of tobacco.Control of blood pressure and maintaining proper foot care are important for people with the disease.

As per the official WHO data, India is in top of the countries with the highest number of diabetics; China, America, Indonesia, Japan, Pakistan, Russia, Brazil, Italy and Bangladesh. In the year 2030 it may reach 79.4 million. It is marked that, persons with excessive weight are at higher risk of falling the diseases such as diabetes. So, doing healthy and regular physical activity is theimportant thing to beat the lifestyle diseases including diabetes.

Big data: The termbig datameans the large volume of data – both structured, semi-structured and unstructured. But amount of data is not important.The important thing is what organizations do with the data . Big data can be analysis can be lead to better

decisions and predictions system. Now days more fields involve Big Data problems, ranging from global economy to society administration and from scientific researches to national security.Apart that Big Data is used in the domain of: health care, and public sector administration. For the theatrical definition of the Big Data [1], there are various different explanations from 3Vs to 4Vs. Doug Laney used volume, velocity and variety, known as 3Vs , to characterize the concept of Big Data. The term volume is the size of the data set, velocity represents the speed of data in and out, and variety explains the range of data types and sources. Some, people expands its definition by adding another V according to their special requirements. The fourth 'V' can be value, variability, or virtual. In simple words, Big Data is a collection of very huge data sets with a great mixtureof typesso that it becomes difficult to process by traditional data processing platforms. In our study we can utilize the 4V definition. We have to collect data from different verticals of a person such as, medical history, family history of diabetics, value of body mass index, nature of blood pressure, nature of food habit. All these data together forms our experimental data.

Hadoopis an open-source distributed data processing platform from Apache[2]. Hadoop has the potential to process extremely large amounts of health data by allocating partitioned data sets to numerous servers like clusters, each of which solves different parts of the larger problem and then integrates them for the final result. Hadoop uses two main components to do its job: Map/Reduce and Hadoop Distributed File System.

• Map/Reduce: Hadoop's implementation of Map/Reduce is based on programming models to process large data or datasets by dividing them into small blocks of tasks. Map/Reduce uses distributed algorithms, on a group of computers in a cluster, to process large datasets. It consists of two functions: The Map ( ) function which resides on the master node and then divides the input data or task into smaller subtasks, which it then distributes to worker nodes that process the smaller tasks and pass the answers back to the master node. The subtasks are run in parallel on multiple computers. The Reduce ( ) function collects the results of all the subtasks and combines them to produce an aggregated final result — which it returns as the answer to the original big query.

Hadoop Distributed File System (HDFS): HDFS replicates the data blocks that reside on other computers in the data center (to ensure reliability) and manages the transfer of data to the various parts of the distributed system. The HDFS system then acts to distribute data across a network or migrate it as necessary.

II. Clinical parameters

The proposed system which use Big Data , that can efficiently find out the rules to predict the risk level of patients based on the given parameter about their health. Here we evaluate many factors. The main clinical factors are 1:number of times pregnant, 2:concentration of plasma glucose rate, 3:blood pressure(mm Hg), 4:triceps skin fold thickness(mm), 5:serum insulin amount(mu U/ml), 6:body mass index, 7:diabetes pedigree , 8:age in years. There are number of predictions are implemented based on the clinical parameters and its values.

III. Influence of Physical Activity/ Exercise and Diabetes

The routine of physical activity / exercise[4] has major role in blood glucose level management in our body. Recommendations and precautions are different depending on individual characteristics and health nature. Physical activity includes all movement that increase energy use,. Exercise increase blood glucose control in type 2 diabetes. The regular exercise contributes to weight loss and improves our health. The exercise habit prevents or delay type 2 diabetes development.

Aerobic exercise

Aerobic training increases mitochondrial density, insulin sensitivity, oxidative enzymes, compliance and reactivity of blood vessels, lung function, immune function, and cardiac output [5]. In individuals with type 2 diabetes, regular training reducesA1C, triglycerides, blood pressure, and insulin resistance [6]. The effect of resistance exercise on glycemic control in type 1 diabetes is unclear [7]. When resistance and aerobic exercise are undertaken in one exercise session, performing resistance exercise first results in less hypoglycaemia than when aerobic exercise is performed first [8]. Resistance training benefits for individuals with type 2 diabetes include improvements in glycemic control, insulin resistance, fat mass, blood pressure, strength, and lean body mass [9].

Strength training

Strength training or resistance training has not become popular as aerobic exercise. There is now a clear evidence of the merits of strength training. When combined with aerobic exercise, some of the benefits are additive, and some others are unique to strength training. This cannot be achieved through aerobic activity alone. Many of these benefits are useful for the management of chronic diseases such as diabetes.Aerobic exercise long is recommended in the management of type 2 diabetes, in large part because of its ability to improve insulin sensitivity and glucose tolerance [10]. Several studies have suggested that strength training improves insulin sensitivity and glucose tolerance when compared to aerobic training [11].

Combining both types of activity appears to take advantage of differing mechanisms of action, enhancing insulin sensitivity and glucose disposal further than either activity could achieve alone[12]. This is exciting news for patients with type 2 diabetes who have already affected diet and aerobic exercise as their means of managing the disease through lifestyle efforts.

Balance exercises

Balance control, muscle strength and joint mobility are also important fall risk factors that may be influenced by exercise. Fear of falls is a cognitive behavioural component, which was recently shown to be related to the gait velocity of diabetic patients [13]. The balance exercise program improved balance and trunk proprioception. These results suggested that a balance exercise is suitable for individuals with diabetic neuropathy.

IV. Proposed Work

Our proposed system which use Big Data, that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. Here we evaluate many factors , 1:number of times pregnant, 2:concentration of plasma glucose rate, 3:blood pressure(mm Hg), 4:triceps skin fold thickness(mm), 5:serum insulin amount(mu U/ml), 6:body mass index, 7:diabetes pedigree , 8:age in years, 9: Nature of Exercises.

Then we evaluate and compare this system using C4.5 rules and Map Reduce algorithm.

In this study, we propose to use Hadoop as the software for Big Data Analysis.Hadoop is an open source distributed processing framework that manages data processing and storage for big data applications running in clustered systems. . It includes various main components, including a MapReduce set of functions and a Hadoop distributed file system (HDFS).

Data Collection

The raw diabetic big data or data set is given as input to the system. The unstructured voluminous input data can be obtained from various Electronic Health Record (EHR) / Patient Health Record (PHR), Clinical systems and external sources (individuals, laboratories, pharmacies, insurance companies etc.), in various formats and residing at various locations.

The proposed system evaluate many factors Hereditary and genetics factors, Stress, Body Mass Index, Increased cholesterol level, High carbohydrate diet, Nutritional deficiency, Nature of Exercises, intake of oil and sugar Overeating, Tension and worries, High blood pressure, Insulin deficiency, Insulin resistance. All these are treated as parameters. We Must use some classification and correlation algorithms to convert these parameters into related values. We the performance of the system is evaluated in terms of techniques used like rules used, classification accuracy, and classification error. By considering all these parameters, the system can predict diabetics in a great accuracy.

V. Conclusion

Big Data Analysis in Hadoop's implementation provides systematic way for achieving better outcomes like availability and affordability of healthcare service to all population. Non-Communicable Diseases like diabetes, is one of a major health hazard in India. According to the official WHO data, India tops the list of countries with the highest number of diabetics; China, America, Indonesia, Japan, Pakistan, Russia, Brazil, Italy and Bangladesh follow.. The goal of this research is to predict diabetics. The design of this system of diabetic treatment may give enhanced data and analytics yield the greatest results in healthcare. Treatment can be offered when it is identified in advance.

.

References

[1] C.L Philip Chen, Chun-Yang Zhang, Data intensive application, challenges, techniques and technologies: A survey on Big Data, Information Sciences 275 (2014) 314–347

[2] Big Data (CoversHadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization) Black Book, Authored By DT Editorial Services. Pages 83 – 114

[3] Predictive Methodology for Diabetic Data Analysis in Big Data. Dr Saravanakumar N M, Associate Professor, Dept of CSE, Bannari Amman Insitute of Technology,Sathyamangalam. EswariT ,2,4Assistant Professor, Dept of IT, Sri Krishna College of Engineering&Techechnology,Coimbatore. Sampath P, Associate Professor, Dept of CSE, Bannari Amman Institute of Technology, Sathymangalam. Lavanya S, Assistant Professor, Dept of IT, Sri Krishna College of Engineering &Techechnology,Coimbatore. Procedia Computer Science 50 (2015) 203 – 208

[4] Physical Activity / Exercise and Diabetes: A position Statement of the American Diabetes Association: Diabetes Care 2016; 39:2065-2079 | DOI: 10.2337/dc16-1728

[5] Garber CE, Blissmer B, DeschenesMR,et al.; American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011;43: 1334–1359

[6] Snowling NJ, Hopkins WG. Effects of different modes of exercise training on glucose control and risk factors for complications in type 2 diabetic patients: a meta-analysis. Diabetes Care 2006;29:2518–2527

[7] Tonoli C, Heyman E, Roelands B, et al. Effects of different types of acute and chronic (training) exercise on glycaemic control in type 1 diabetes mellitus: a meta-analysis. Sports Med 2012;42:1059–1080

[8] Yardley JE, Kenny GP, Perkins BA, et al. Effects of performing resistance exercise before versus after aerobic exercise on glycemia in type 1 diabetes. Diabetes Care 2012;35:669–675

[9] Gordon BA, Benson AC, Bird SR, Fraser SF. Resistance training improves metabolic health in type 2 diabetes: a systematic review. Diabetes Res ClinPract 2009;83:157–175

[10] American Diabetes Association*: Diabetes and exercise (Position Statement).* Diabetes Care25*(Suppl. 1):*S64–*S68*,2002
[11] Ryan AS, *Pratley RE, Goldberg AP, Elahi D: Resistance training increases insulin action in postmenopausal women*. J Gerontol51A *:* M199–*M205*,1996
[12] Wallace MB, *Mills BD, Browning CL: Effects of cross-training on markers of insulin resistance / hyperinsulinemia*. Med Sci Sports Exerc29 *:* 1170–*1175*,1997
[13] Erikson J, *Tuominen J, Valle T, Sundberg S, Sovijarvi A, Lindholem H, Tuomilehto J, Koivisto V: Aerobic endurance exercise or circuit-type resistance training for individuals with impaired glucose tolerance?* HormMetab Res30:37 – *41*, 1998.