

PREDICTION OF HEART DISEASES AND CANCER IN DIABETIC PATIENTS USING DATA MINING TECHNIQUES

Savan M Mendapara¹, Deep R Kahar², Prof. Binita B. Acharya³

^{1,2}Dept. of Computer Engineering, Shree Swami Atmanand Saraswati Institute of Technology, Surat, Gujrat, INDIA

³Assistant Professor, Dept. of Computer Engineering, Shree Swami Atmanand Saraswati Institute of Technology, Surat, Gujrat, INDIA

Abstract - Data Mining Techniques are used in many areas like banking, healthcare, education because of extracting relevant information from the database. Data mining has become a main strategy in many industries to improve outputs and decrease costs. Now days in healthcare management this field will become very useful. Data mining techniques has become great potential for the healthcare industry to predict health deceases by using systematic data and analytics to identify inefficiencies and best practices that improve care and reduce costs. These techniques are fast in nature and take less time for the prediction system to improve the diabetic decease with more accuracy. In this research work, the risk level of diabetic heart disease patient & cancer disease patient can be predicted using two datasets such as Pima Indians Diabetes dataset and heart disease dataset. Based on the World Health Organization (WHO), diabetes is the one of the biggest health concerns so mining the diabetes data is an ambiguous task. Diabetic patients may also have to suffer from other diseases like heart disease, eye complications, kidney disease, nerve damage, foot problems, skin complications and dental diseases. However, the existing techniques faced the difficulty to identify inconsistent and redundant features.

Key Words: Analysis, Data Mining, Prediction, Heart Disease, Diabetic

1. INTRODUCTION

Data Mining is the process of extracting hidden knowledge from large volumes of raw data. The knowledge must be new, not obvious, and one must be able to use it. Data mining has been defined as “the nontrivial extraction of previously unknown, implicit and potentially useful information from data. It is “the science of extracting useful information from large databases”. It is one of the tasks in the process of knowledge discovery from the database. ^[1]

Data Mining is used to discover knowledge out of data and presenting it in a form that is easily understand to humans. It is a process to examine large amounts of data routinely collected. Data mining is most useful in an exploratory analysis because of nontrivial information in large volumes of data. It is a cooperative effort of humans and computers. Best results are achieved by balancing the knowledge of human experts in describing problems and goals with the search capabilities of computers. There are two primary goals of data mining tend to be prediction and description. Prediction involves some variables or fields in the data set to predict unknown or future values of other variables of interest. On the other hand Description focuses on finding patterns describing the data that can be interpreted by humans.

The Diabetic Disease Prediction plays an important role in data mining. There are different types of diseases predicted in data mining namely Hepatitis, Lung Cancer, Liver disorder, Breast cancer, Thyroid disease, Diabetes etc... This paper analyses the Heart disease & cancer disease predictions in Diabetic patient.

2. EXISTING SYSTEM

Clinical decisions are often made based on doctors’ intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. There are many ways that a medical misdiagnosis can present itself. Whether a doctor is at fault, or hospital staff, a misdiagnosis of a serious illness can have very extreme and harmful effects. The National Patient Safety Foundation cites that 42% of medical patients feel they have had experienced a medical error or missed diagnosis. Patient safety is sometimes negligently given the back seat for other concerns, such as the cost of medical tests, drugs, and operations. Medical Misdiagnoses are a serious risk to our healthcare profession. If they continue, then people will fear going to the hospital for treatment.

3. MEDICAL DATA MINING

Data mining has been used to uncover patterns from the large amount of stored information and then used to build predictive models. Medical field contains large amount of data that are needed to be processed. Data mining in medical field improves the quality of patient care and the prediction of healthcare patterns. Data mining tools helps us to discover unknown patterns, group the related items and decision making of healthcare oriented problems. Medical care is necessity; it gives patient and hope for a fruitful life. The collected data when published is used for social causes without harming the dignity of the patients. Early detection of disease can increase the survivability of patients [2]. Data mining techniques such as classification and prediction, clustering, association rule mining and various mining methods can be useful to apply on medical data. The collected data in the form of images can also be used to mine healthcare data. Nowadays, many image mining techniques improved the disease prediction and health care decision making task as easiest. This paper presents the application of data mining algorithms for the Heart disease & cancer disease predictions in Diabetic patient.

3.1 CANCER IN DIABETIC PATIENT

Diabetes and cancer represent two complex, diverse, chronic, and potentially fatal diseases. Cancer is the second leading cause of death, while diabetes is the seventh leading cause of death with the latter still likely underreported. There is a growing body of evidence published in recent years that suggest substantial increase in cancer incidence in diabetic patients. The worldwide prevalence of diabetes was estimated to rise from 171 million in 2000 to 366 million in 2030. About 26.9% of all people over 65 have diabetes and 60% have cancer. Overall, 8-18% of cancer patients have diabetes. In the context of epidemiology, the burden of both diseases, small association between diabetes and cancer will be clinically relevant and should translate into significant consequences for future health care solutions. This paper summarizes most of the epidemiological association studies between diabetes and cancer including studies relating to the general all-site increase of malignancies in diabetes and elevated organ-specific cancer rate in diabetes as comorbidity. Additionally, we have discussed the possible pathophysiological mechanisms that likely may be involved in promoting carcinogenesis in diabetes and the potential of different antidiabetic therapies to influence cancer incidence. [3]

3.2 HEARD DISEASE IN DIABETIC PATIENT

Various researches have been conducted to find the relationship between diabetes mellitus and heart diseases recently in diabetic patients. The diabetes is a group of metabolic disorders that affects life of human. A reinterpretation is required, so, the possibilities of heart diseases in diabetic patients are high [7,8] Most of the diabetic patients are affected by metabolic and hormonal disorders. Some of the risk factors of heart diseases and cancer among diabetic patients are lack of physical in activity, alcohol, tobacco smoking and obesity. The biological link mechanisms behind heart diseases and diabetes are hyperinsulinemia, hyperglycaemia and inflammation in the blood [8]. Very high blood sugar damages blood vessels can lead to blockage. People with diabetes have two to four times risk of developing heart disease. High blood sugar causes blocks in leg vessels can cause pain and also impair circulation. Some of the symptoms of heart failure are problems in breathing, swelling in the ankles, feet, legs, abdomen, and veins in neck. People with heart failure can live longer and more active lives if the condition is diagnosed earlier and if they follow their treatment plans regularly. People having diabetes can have a chance of developing heart disease and stroke. Several new techniques are used for diagnosing both heart diseases and diabetes. The machine learning techniques are one of the existing techniques which have a transparent diagnostic knowledge to diagnose diseases [9]

4. DATA MINING USES TWO STRATEGIES

Supervised and unsupervised learning. In supervised learning, a training set is used to learn model parameters whereas in unsupervised learning no training set is used (e.g., k-means clustering is unsupervised). Each data mining technique serves a different purpose depending on the modelling objective.

The two most common modelling objectives are classification and prediction. Classification models predict categorical labels (discrete, unordered) while prediction models predict continuous valued functions [4]. Decision Trees and Neural Networks use classification algorithms while Regression, Association Rules and Clustering use prediction algorithms.

5. DATA MINING TECHNIQUE & DATASET

5.1 NAIVE BAYES ALGORITHM

The Naïve Bayes Algorithm is a probabilistic algorithm that is sequential in nature. The Naive Bayes algorithm rely upon conditional probabilities which uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data. Bayes' Theorem finds the probability of an event that occurs when given the probability of another event already been occurred.

$$\text{Prob (B1 given A1)} = \text{Prob(A1and B1)}/\text{Prob(A1)}$$

Here, B1 represents the dependent event and A1 represents the prior event.

5.2 WEKA Tools

The WEKA ("Waikato Environment for Knowledge Analysis") tool is used for Data mining. Data mining finds valuable information hidden in large volumes of data. Weka is a collection of machine learning algorithms for data mining tasks, written in Java and it contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.^[5] The key features of Weka are it is open source and platform independent. It provides many different algorithms for data mining and machine learning^[6]. We have used Naïve bayes method to perform the mining and classification process.

Sr. No.	Attribute	Relabeled Values
1.	Age	Age
2	Obesity	Abes
3	Heart rate	heart
4	Chest Pain	chest
5.	Blood Pressure	Pres
6	Blood Sugar	Insu
7	Cholesterol	mass
8	Body Mass Index	mass
9	Triceps skin fold thickness	Skin
10	Number of times pregnant	Preg

Table -1: Dataset

6. CONCLUSION

This paper aimed to analyses the application of data mining in medical domain and some of the algorithms used to diseases. It is observed that results may vary for different disease diagnosis based on the tools and techniques used. Data mining provides good results in disease diagnosis when appropriate tools and techniques applied. Hence data mining is the promising field for healthcare predictions

REFERENCES

- [1] Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006
- [2] Dhanya P Varghese, Tintu P B, –A Survey on Health Data using Data Mining Techniques||, International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 07, e-ISSN: 2395-0056, p-ISSN: 2395-0072, Oct2015
- [3] Samy Lewiz Habib Diabetes and Risk of Cancer February 2013ISRN Oncology 2013(45):583786 DOI: 10.1155/2013/583786
- [4] Liao, S.-C., I.-N. Lee, "Appropriate medical data categorization for data mining classification techniques", MED. INFORM., 2002. Vol. 27, No. 1, pp. 59–67
- [5] Weka Data mining software <http://www.cs.waikato.ac.nz/ml/weka>
- [6] An Introduction to the WEKA Data mining system - <http://www.cs.ccsu.edu/~markov/weka-tutorial.pdf>
- [7] Shi Y, Eberhart R. A modified particle swarm optimizer. The Proceedings of the IEEE International Conference on Evolutionary Computation; 1998; Piscataway, NJ. p. 69–73
- [8] Bianchini F, Kaaks R, Vainio H. Overweight, Obesity and Cancer risk. Lancet Oncol. 2002 Sep; 3(9):565–74.

[8] Alessandra Saldanha de. Impact of Diabetes on Cardiovascular Disease: An Update. International Journal of Hypertension. 2013; 653789.

[9] Nagarajan S, Chandrasekaran RM. Design and implementation of expert clinical system for diagnosing diabetes using datamining techniques. IJST. 2015 Apr; 8:771-6.

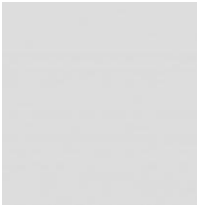
BIOGRAPHIES



Savan M. Mendapara
Bachelors of Engineering
Department of Computer Engg.
Shre Swami Atmanand Saraswati
Insti. of Tech. Surat Gujrat India



Deep R. Kahar
Bachelors of Engineering
Department of Computer Engg.
Shre Swami Atmanand Saraswati
Insti. of Tech. Surat Gujrat India



Prof. Binita B. Acharya
Assistant Professor
Department of Computer Engg.
Shre Swami Atmanand Saraswati
Insti. of Tech. Surat Gujrat India