Data Mining Technique to Improve the Prediction of Diabetes

Ruwaisa Saba & Ravi. P
1M.Tech Student, 2Assistant Professor,
1,2 Vidyavardhaka College of Engineering, Mysuru, Karnataka.

Abstract: From the last decade in India diabetes has reached epidemic proportions. It has been estimated that over 75 million people in India would become diabetic by 2025. In health concern business data mining plays an important role in early prediction of disease. However the existing healthcare infrastructure is inadequate to meet the demands of this exploding population. Providing a web based patient support system would presently help to improve the quality of life among patients and to reduce the treatment cost from a very large web based medical database extracting useful information and providing scientific decision making is critical and difficult this problem can be dealt by applying data mining techniques for heterogeneous and very large databases. This explores a new way of monitoring health, especially for diabetes and trying to minimize the problems faced by diabetic patients.

Keywords: Data mining, clustering, k-means, diabetes, visualization, doctor, patient.

1. Introduction

Diabetes is growing very rapidly in India; diabetes is a disease where for a body to work properly sugar should be converted into energy.

With diabetes, a hormone called insulin, which is essential for the conversions of the glucose (sugar) into energy, is no longer produced in sufficient amounts by the body or the insulin produced is not working properly.

Generally diseases will be of several types because of which prediction becomes difficult and we need a laboratory to predict the disease or we need to compare the values for that particular disease. In existing system, Prediction module and diagnosis module are two of the main features in Web-Based Medical Diagnosis and Prediction. Prediction module utilizes neural networks technique to predict patient illness or conditions based on the previous similar cases has been used which is a difficult task. For each disease of a patient, where we need to create a new node structure which makes the decision complicated and prediction difficult.

Diabetes is a disease. For our bodies to work properly we need to convert sugar into energy. With diabetes, a hormone called insulin, which is essential for the conversions of the glucose (sugar) into energy, is no longer produced in sufficient amounts by the body or the insulin produced is not working properly.

There are two main types of diabetes.

- Type 1 or insulin–dependent diabetes (Juvenile Diabetes) which usually affect children, teenagers and young adults and requires daily injections of insulin.
- Type 2 or non-insulin dependent Diabetes (Adult-onset Diabetes), which usually affects people over the age of 45 years and is treated by healthy eating and regular exercise. Tablets and insulin injections are sometimes necessary.
2. Existing System

An expert system that aims to provide the doctors with limited practical experience and the diabetic patient more background for suitable diagnosis of diabetic micro vascular complications. In modern society, more and more people get diabetes. This disease affects almost every organ in the body like heart, eyes, kidney, skin, nerves, blood vasculum and foot etc. They leave the very serious consequences for the patient. Early diagnosis of complications of diabetes can get tremendous value to patients and society.[3] Hence an expert system for diabetes micro vascular complication diagnosis can be used to assist for the doctors with limited clinical experience in situations without consultation directly to a specialist. In constructing an expert system, the knowledge representation plays a very important key.

This expert system presents the design and implementation of an intelligent medical expert system for diabetes diet that intended to be used in Sudan. [4] The development of the proposed expert system went through a number of stages such problem and need identification, requirements analysis, knowledge acquisition, formalization, design and implementation. Visual prolog was used for designing the graphical user interface and the implementation of the system.

This existing system presents a fuzzy expert system Framework that combines case-based and rule-based reasoning effectively to produce a usable tool for Type 2 Diabetes Mellitus (T2DM) management. The major targets are on combined therapies (i.e., lifestyle and pharmacologic), and the recognition of management data dynamics (trends) during reasoning. [5,8] The Knowledge base (KB) is constructed using fuzzified input values which are subsequently de-fuzzified after reasoning, to produce crisp outputs to patients in the form of low-risk advice. The extended framework features a combined reasoning approach for simplified output in the form of decision support for clinicians. With seven operational input variables and two additional pre-set variables for testing, the results of the proposed work will be compared with other methods using similarity to expert’s decision as metrics.

In this expert system we are proposing an efficient decision support system for Diabetes Disease, apart from the traditional simple support vector machine. [10] We are proposing an efficient two level approach for classifying data. In initial phase we extract optimal feature set from the training data by analyzing the optimality in the dataset, and then new dataset is formed as optimal training dataset. Then classification mechanism on the optimal feature set is applied.

3. Proposed System

After I surveyed the literature, I found a lot of researches have been conducted in the area of diagnoses and determining the treatment of diabetes. Later I also decided to introduce a system which will help for diabetic patient, named as diabetic medical system using data mining. This system will give complete information about diabetics. The patient can get general tips, information, and suggestion from doctor, prescription from the doctor.

In our system doctor plays an important role, He will add new patient and new suggestion into the system. Has the permission to view the history of the patients. And provide suggestions to the patients.

In data mining technique, at first we choose a medical data set and insert the data into the machine. The Clustering algorithm groups a set of objects. Segregation of data into different clusters according to the disease. Prediction algorithm is used to give the required output for the disease.

In this paper a Learned algorithmic approach to improve the prediction of diabetes has been designed following a hierarchical approach. Diabetic Medical System Using Data Mining is based on several elements.

- A Patient, which request for the suggestion and get suggestion for the doctor.
- A Doctor, which adds new suggestion and provide suggestion for patient who request for service.
- A General Tips, where the patient request for the general tips about the diabetic disease for his knowledge.
- A Pre Processing, which is going to processes its input data to produce output that is used as input to another program.
- A Clustering, which is used to cluster the data set (Patient details). Here we use K-Mean algorithm for clustering.

Advantages

1. High reliability.
2. Easily understandable.
3. Efficient use of data mining techniques.
4. Less expensive.
Overall functions of Diabetic Medical System is

General Tips
- Gives complete information about diabetics.

Data Mining
- Preprocessing.
- Clustering.
- Visualization.
- View the patient history.
- Gives suggestion for the patient.
- Logout

The overview of the software can be depicted by the following figure:

![Figure-1: Overview of the software](image1)

Patient requests for general tips from the doctor it goes through authentication and Data mining steps are carried out that is pre-processing, clustering and visualization and reaches terminal.

Software activities

1. General tips activity
   In this general tips module, the patient will request for general tips about diabetics, which include several information about the disease. He will select what all information is needed. This plays an important role for patient who has doubt about the disease.

   Login activity

![Figure-2: Login screen for patients](image2)

![Figure-3: General tips option for diet](image3)

![Figure-4: General tips about symptoms](image4)

![Figure-5: General tips about food taking](image5)
2. Data mining activity

In the data mining module, the preprocessing of the data is done. Here data refers to the patient disease details followed by clustering. The clustering process will segregate the data into clusters by using k-mean clustering algorithm. Later, the visualization part will help in easy classification of patient through clustering k-mean algorithm.

4. Conclusion

The future for medicine will be better. The use of computer and communication tools can change the medical practice into a better implementation. Advancement in technology will form a platform for developing a better design of telemedicine application. Here, it is being proposed to focus on some feature selection and move over to classification of medical data which is based on the feature selection algorithm.

This will enable the clinician to access the system and provide the consultation as expert does regardless of the location. Patient’s record or patient’s database could be installed at the main server. The electronic record could be accessed by health-care providers and the data could be stored and updated frequently. By using this method, the system knowledge will always be updated.

K-means is better by many factors like first, it gives better results when compared with other algorithms, by increasing factor. Secondly, K-means takes lesser time to cluster the data. Thirdly, K-means is considered to be a hard clustering and in hard clustering, after some iteration most of the centers are converged to their final positions and the majority of data points have only few candidates to be selected as their closest centers.

5. Future Enhancement

Though this is a well-modulated system, it has been limited to certain restrictions. By understanding trends in technology, it is possible to make accurate prediction about what will happen in future. Future work will consist of:

- Extending it for all human diseases.
- To give direct link to the online medicine delivery website, such that patients can order it from there place itself.
- To validate the whole platform at a large scale.

4. References

[1] Thirumal P. C. and Nagarajan N. “Utilization Of Data Mining Techniques For Diagnosis Of Diabetes Mellitus - A Case Study”, Department of IT, Coimbatore Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India, Jan 2015.


