General Review

Stark Assessment of Lifestyle Based Human Disorders Using Data Mining Based Learning Techniques

M. Sharma [a,*], , G. Singh [b], R. Singh [b]

[a] *Dept. of Computer Sc. and App., DAV University, Jalandhar, India*
[b] *DCS, GNDU, Amritsar, Punjab, India*

Highlights

- Lifestyle based human disorders diagnosed have been done by using different data mining techniques.
- Common and live datasets for lifestyle based human disorders have been mentioned.
- Rate of accuracies achieved using different mining techniques is also highlighted.
- Effect of preprocessing, relationship between disease, datatype and mining approach is analyzed.
- A novel hybrid diagnosis model for lifestyle based human disorders is proposed.

Graphical abstract
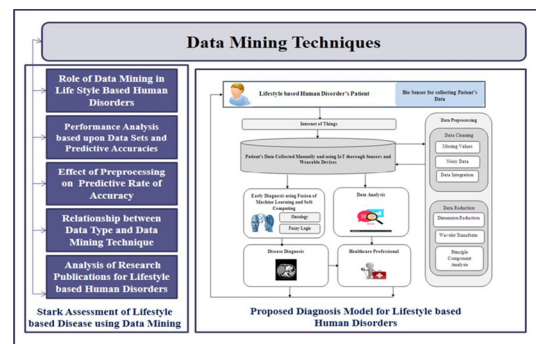
Abstract

Background: Medical informatics has observed an unrestrained growth in the database. Latest advancements in the field of medical sciences have wiped out lots of critical diseases. Nowadays, the medical industry is affluent in data sources. These data sources are of use only if these are effectively analyzed on time.

Methods: Data mining techniques are artificially intelligent and used to investigate known and unknown patterns available in the medical databases. Nowadays, data mining techniques are chronically used to mine abundant data sources of medical science. This paper explores the practice of diverse data mining techniques, the role of dataset used, effect of preprocessing, and the performances of different data mining techniques in diagnosis of different lifestyle based diseases. The venture of this paper is to fetch out stark assessments of different data mining techniques used in medical sciences.

Results: By far, surveillance discloses that significant effort has been made for mining the data allied to the Cardiology and Diabetes. As per Google Scholar, in last seven years, the percentage of articles published related to cardio, diabetes, digestive, dentistry and ophthalmology disease

---

* Corresponding author.
  E-mail address: manik10143@davuniversity.org (M. Sharma).

diagnosis using data mining are 42%, 26%, 18%, 10% and 4% respectively. So, a little attention has been paid to develop predictive model for the diseases viz. ophthalmology, dentistry and digestive disorders. In addition, the rate of usage of preprocessing in diagnosis of different disorders related to cardio, diabetes, digestive, dentistry and ophthalmology lies between 10.65%–17.75%, 8.48%–14.80%, 4.58–8.93%, 2.96%–7.73% and 5.83%–12.93% respectively.

Conclusion: An attention is obligatory to develop smart diagnostic system to aware and save human masses from wide critical spectrum of diseases related to ophthalmology, oral and digestive systems.

1. Introduction to data mining

In modern times, Data mining has acknowledged an immense deal of attention. Data mining is an automatic extraction of patterns or useful information. It is used to stumble on the consequence of the problem by probing the available facts and data. It helps in measuring the association, nature and degree of correlation in the different parameters of the dataset. Moreover, the dataset can also be classified into several classes. At core level, data mining is known as knowledge discovery database. One can also predict the future by carefully analyzing the past and current state of dataset.The commonly used data mining techniques are naïve bayes, neural network (NN), genetic algorithm (GA), support vector machine (SVM), decision tree induction (DTI) etc. The complete working of data mining techniques is based upon two sets of data i.e. Training and Testing. Data mining is a difficult process as one has to train the system regarding the characteristics or the features that have to be extracted. Data mining is one of the dominant research fields and is used in almost all streams viz. Agriculture, Computer Science, Mathematics, Chemistry, Finance, Economics, Medical Science, Zoology, Bio-informatics etc. The major applications of data mining are given below [1,2]:

- Agriculture: To study characteristics of soil, crop yield analysis, optimizing use of pesticide, diagnosis of plant diseases.
- Medical Science: Early diagnosis and prediction of various diseases.
- Computer Science: Image processing, Network Security, Computer Security.
- Business: Airline, Shopping Malls, Banks, Marketing.
- Natural and Life Sciences: Creation of new hypothesis, Stimulation of new facts, Analyzing Protein structure etc.
- Engineering; CAD, CAM, Fault detection in production lines.

Data mining [3] is a multistep approach. In the first phase, data is collected from heterogeneous sources and is converted into a homogeneous format. Different preprocessing and normalization procedures are used to reduce data inconsistency. In second phase, the data mining procedures are applied to excavate some meaningful information. Third phase analyzes the processed data and represent it in standardized format. Finally, the upshots of data mining progression are used in decision making process. In general, data mining is used to test hypothesis or to discover some new or hidden patterns. Traditionally, it was associated with hypothesis testing only. The idea was to first prepare a statement that has to be tested against particular set of data and condition.

There are certain anomalies [4] in mining the database. The most imperativecrisis in data mining is the lack of long term vision. Sometimes, data used in mining is incorrect or may not be updated. Moreover, some of the organization or departments within an organization do not want to reveal the nature and contents of their data. In addition, legal and privacy condition may further create a problem in data mining. One of the crucial factors of decision making is having the right information at the right time. Today, data collection is not of major issue. Rather, analysis of data is of concern. Nowadays, the survival of organization is based upon the effectiveness in generating information from their data. It is believed that the future and success of medical industry is heavily dependent upon the data mining process.

This paper briefly summarized the endeavor of different researchers for mining information for medical industry.

2. Related work

Suvarna Pawar and Smita Sikchi [5] have carried out a survey related to diabetes. The survey is focused on to unearth out intricacy with the existing diagnosis systems. Authors found that still it seems too difficult to diagnose diabetes with high precision. Authors tried to emphasize on diverse classification techniques that were used in diabetes diagnosis. Authors observed that for PIMA data set, SVM seems to be a good solution for early diagnosis of diabetes with 97% of accuracy.

Shubpreet Kaur, R.K. Bawa [6] stated that the use of data mining provides an efficient way to mine the required clinical information from capacious, raw and heterogeneous data. One may use the techniques viz. decision tree, naïve bayes, logistic regression and ANN to forecast different medical problems with low cost and high precision. In addition, author presented how data mining can be used to ascertain the interactions between health status and a disease.

Divya Tomar, Sonali Aggarwal [7] discovered the efficacy of an assortment of data mining techniques like association, classification, clustering and regression in health sphere. Authors briefly explained different data mining techniques with their intrinsic worth and demerits. Authors suggested that before applying the classification techniques, one should preprocess the