

# Analysis of Data Mining Techniques in Diagnosing Heart Disease

Nagaraj P<sup>a,1</sup>, Aakash M<sup>b</sup>, Arunkumar M<sup>b</sup>, Balanathanan K<sup>b</sup>, Dharanidharan A<sup>b</sup> and  
Rajkumar C<sup>b</sup>

<sup>a</sup> Assistant Professor, Dept of CSE, Kalasalingam Academy of Research and Education  
Krishnan Kovil, Srivilliputtur

<sup>b</sup> Dept of CSE, Kalasalingam Academy of Research and Education  
Krishnan Kovil, Srivilliputtur, India

**Abstract.** Data mining is a technique used for extracting the hidden database technology on a large database. Further, the medical data mining became extremely important research field due to its importance in the development of various applications in developing health care of every domain. While summarizing the deaths occurring over worldwide, the heart disease appears to be the leading cause. The identification of the possibility of heart disease in a person is a very complicated task for medical practitioners because it requires a few years of experience and intense medical tests to be conducted frequently. The main objective of this significant research work is to identify the best and suitable classification algorithm used for providing maximum accuracy when classification of normal and abnormal person is carried out. Thus the loss of lives can be controlled.

**Keywords.** Data Mining, Heart Disease, Outcomes, Dataset, Decision Tree.

## 1. Introduction



Figure 1. Introduction to data mining

Data mining is known for its ability to find the patterns in large data sets. Using different clinical reports, a large amount of data is generated from the field of medical science. Data mining is also used in healthcare to analyze the medical datasets and find

<sup>1</sup> Nagaraj P; Assistant Professor, Department of CSE, Kalasalingam Academy of Research and Education  
Krishnan Kovil, Srivilliputtur, India;  
E-mail: nagaraj.p@klu.ac.in

the hidden patterns and arrive at better solutions. These data sets and their results are organized, assembled and integrated with the hospital database systems.

One of the most chronic diseases in the world is the cardiovascular disease. According to a survey held by World Health Organization, cardiovascular diseases are the most dangerous and they cause millions of deaths in a year. Thus measures should be taken in order to reduce this. The diagnosis of the cardiovascular diseases is a very confusing and tough process. It should be diagnosed in a correct way.

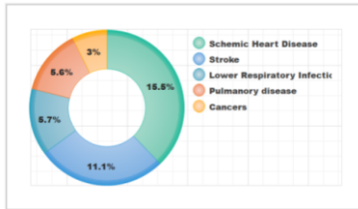


Figure 2. Showing the major heart diseases in the world

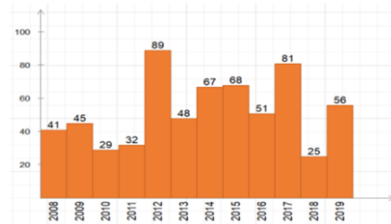


Figure 3. Showing the increasing level of heart diseases

## 2. Existing System

Currently, the deadliest cardiovascular disease is known to be as Acute Myocardial Infarction (AMI). Healthcare centers used to handle data with small and costly methods. But after the introduction of EHRs (Electronic Health Records), the process of data input became cheaper and easier and the care for an individual patient increased. The old concept “Big data” is now the key method to collect, organize and harness data. The primary modes in which the patients are discharged and taken care of are the outcome measures. The big data takes much time to process and exceeds the processing time and the processing capacity of the normal databases. The characteristics of big data can be described by the 6 Vs. The 6 Vs are:

- Volume
- Velocity
- Variety
- Value
- Variability and
- Veracity

By using these 6Vs data can be optimized and uploaded in the EHRs which can be analyzed and the predictions can be made for the particular patient. The data are also segregated by using the data mining process.

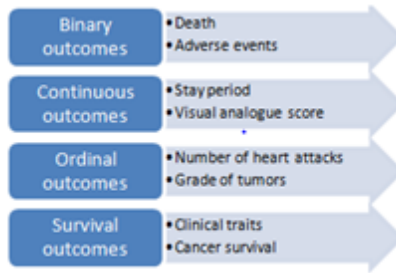


Figure 4. Showing some types of the outcomes

### 3. Data Mining Process

It is also known as the method of text analysis. It is used to give predictions in many areas of healthcare analytics like it can predict how much treatment should be given to the patients, to evaluate the impact of the treatment, etc.

Data mining makes it possible for the healthcare industry to organize and use the data in an effective manner and to find any malfunctions in the storage units and the best solutions that can improve care and reduce costs. It helps the healthcare centers to get deep insights and makes it easier to analyze large.

It is also used to reveal new biomedical results and predictions that can be used in the clinical and administrative areas or huge datasets. In some cases, healthcare centers could increase their profit by predicting with data mining.

Due to its increased application in every field, it is now used for Knowledge Discovery in databases (KDD). Hence, by using the process of data mining we can discover knowledge and apply it to predict results out of data. This is used in diagnosing heart diseases. Avoid using illustrations taken from the Web. The resolution of images intended for viewing on a screen is not sufficient for the printed version of the book. If you are incorporating screen captures, keep in mind that the text may not be legible after reproduction.

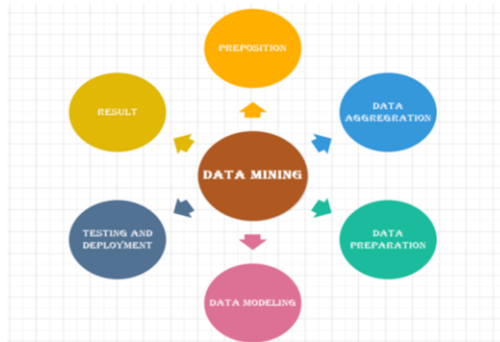


Figure 5. Showing the process of data analytics

### 4. Proposed System

It is believed that the healthcare will be benefitted by following data analytics. Many heart diseases are diagnosed and will be diagnosed in a faster way by using data analytics. The heart disease diagnosis consists of three steps. They are collection of

data, pre-processing the data and the classification of data.

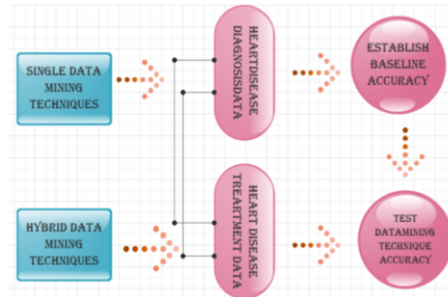


Figure 6. Showing the proposed system

Heart diseases can be diagnosed by creating a confusion matrix. A confusion matrix is a simple table that has various predictions and their corresponding different results and compares these values with the real-time values and helps us to arrive at solutions. For e.g.: In order to predict a heart disease, a confusion matrix table with two values A and B is created. Let us assume A to be the patients with disease and B to be the patients with no disease. So, the confusion matrix will be shown .

Table 1. confusion matrix

Patients	A	B
A	TP	FN
B	FP	TN

In this table, the meanings of the entries are:

- TP-data which are predicted as true when they are entered as true.
- TN-data which are predicted as true when they are entered as false
- FP-data which are predicted as false when they are entered as true.
- FN-data which are predicted as false when they are entered as false.

The data of the patients are collected from many records. They are collected with the patient details such as the patient’s age, height, weight, sex, cholesterol levels, chest pain type etc with some constant values. These data which are collected are converted into a structural database. The data is first checked for duplication and then the missing values are found out and filled and then the values are processed according to their corresponding domain values.

After processing the data with all these things, if the confusion matrix shows that there are only TP values and FN values, then we can get 100% accuracy in predicting the people having a heart disease.

For instance, consider the following data table,

**Table 2. Having 100% accurate predictions**

Patients	A	B
A	50-TP	0-FP
B	0-TN	598-FN

## 5. Data Set

Here is a data set that has the patient details and the corresponding domain values, only by using this data set the above actions are performed.

**Table 3. Datasets**

S.NO	Parameters	Parameter description	Values
1	Age	Age(X)	Continuous
2	Sex	Female or Male	1=Male 0=Female
3	Thespis	Resting blood pressure	1=typical type1
4	Cp	Chest pain type	2=typical type angina 3=No pain
5	Chol	Serum cholesterol	4=asymptomatic
6	Fbs	Fasting Blood Pressure	1>=120 mg/dL 0<=120mg/dL
7	Rest Cg	Resting Electrographic Result	0=Average 1=having ST-T wave abnormal
8	Thalach	High level Heart rate Achieved	1=having ST-T wave abnormal 2=Left Ventricular Hypertrophy
9	Old Peak	ST depression induced by using In order to rest	Continuous value
10	Exang	Use induced angina	0=no 1= yes
11	Ca	High number of Major Vessels used by Fluroscopy	0-3 values
12	Slope	Graph of the data By ST segment	1=High 2=Normal
13	Thal	Shortcoming	1=Yes 0=no
14	Obes	Obesity	1=Yes 0=no
15	Num	Diagnosning Heart Disease	0%<=50% 1%>=50%

We can also use the decision tree to classify the data. It is the easiest way to classify the data. The classification and regression classification types which are in the form of decision trees are much easier to understand and operations like finding the errors and removing them can be done. Easily.

By using this structure, we can classify the data and make further predictions. This is a basic method. This algorithm works by finding the information gain of the attributes and then dividing the data tree into branches.

## 6. Conclusion

In this paper the analysis of heart disease has been discussed by using data mining method with the help of different techniques which results in the absolute accuracy of people having heart disease. Data mining is used to convert large amount of raw health care data into information which helps to take better decision and to make better prediction. By using these method we can avoid late discovery of the disease by diagnosing the heart disease earlier and implement the required actions as soon as possible. Some experimental result reveal that the heart diseases predicted by the system are 100% accurate when neural networks are used. So that the people get aware about their body condition.

## References

- [1] Panda, B. S., Misra, A., &Gantayat, S. S. (2019). Methods And Concepts Of Data Mining Techniques To Impute Missing Data Information. [PDF] researchgate.net.
- [2] Laxmikanthlaxmikanth, P., & Bhramaramba, R. Data D
- [3] Techniques, Concepts And Applications For Diabetes Mellitus-A Decade Review From 2008 To 2018. [Pdf] Researchgate.Net.
- [4] Khanbabaei, M., Sobhani, F. M., Alborzi, M., &Radfar, R. (2018). Developing an integrated framework for using data mining techniques and ontology concepts for process improvement. *Journal of Systems and Software*, 137, 78-95. [PDF] researchgate.net.
- [5] Prasdika, P., & Sugiantoro, B. (2018). A Review Paper on Big Data and Data Mining Concepts and Techniques. *IJID (International Journal on Informatics for Development)*, 7(1), 33-35. [PDF] uin-suka.ac.id.
- [6] Shmueli, Patel, N. R., &LichtendahlJr, K. C. (2017). Data mining for business analytics: concepts, techniques, and applications in R. John Wiley & Sons. [PDF] ozone.ru.
- [7] Jayamalini, K., &Ponnaivaikko, M. (2017, February). Research on web data mining concepts, techniques and applications. In 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET) (pp. 1-5). IEEE. [PDF] e-tarjome.com
- [8] Demigha, S. (2016, March). A case-based reasoning tool for breast cancer knowledge management with data mining concepts and techniques. In *Medical Imaging 2016: PACS and Imaging Informatics: Next Generation and Innovations* (Vol. 9789, p. 97890I). International Society for Optics and Photonics.
- [9] Pradhan, P. R., &Kulkarni, R. B. (2016, March). Secure e-learning using data mining techniques and concepts. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 1498-1501). IEEE.
- [10] Kester, Q. A. (2016, July). Using Formal Concepts Analysis Techniques in Mining Data from Criminal Databases and Profiling Events Based on Factors to Understand Criminal Environments. In *International Conference on Computational Science and Its Applications* (pp. 480-496). Springer, Cham.
- [11] Galit, S., Nitin, R. P., & Peter, C. B. (2016). Data Mining for Business Intelligence: Concepts, Techniques and Applications in Microsoft Excel® with XLMiner®.
- [12] V.D.Ambeth Kumar, G.Gokul, S.Malathi, K.Vengatesan, D.Elangovan, B.Chitra, "Implementation Of The Pulse Rhythmic Rate For The Efficient Diagonising Of The Heart Beat", ", *Healthcare Technology Letters (IET)* 2019 Apr 17;6(2):48-52.