

Prediction System For Detection and Diagnosis of Coronary Diseases Using Data Mining Techniques: A Review

Jaspreet Kaur Sahiwal

Department of Computer Science & Engineering
Lovely Professional University
Phagwara, Punjab, India.
jaspreet.14752@lpu.co.in

ABSTRACT

The most spreading illness that generally affects individuals and is the cause of the majority of deaths in the nation is the coronary illness. By 2030, around 25 million individuals will lose their lives in light of heart diseases. Even though numerous analysts have recommended and proposed strategies for diagnosing the heart maladies from the considerable measure of coronary illness information, suitable methods are not available and are not appropriately mined. To get exactness and productivity in result, another methodology called improved k-mean calculation is proposed in this paper. The dataset utilized for the forecast is extracted and used from UCI machine learning archives. The exploration work depends on prediction analysis for heart disease recognition. The forecast investigation in data mining contains two stages in which the first stage is clustering, and the second stage is classification. The fundamental goal of this work is to derive a connection between traits of the dataset by utilizing improved k-mean grouping to acquire exactness level.

Keywords: *Data mining, Clustering, Classification, K-mean algorithm, Prediction Analysis*

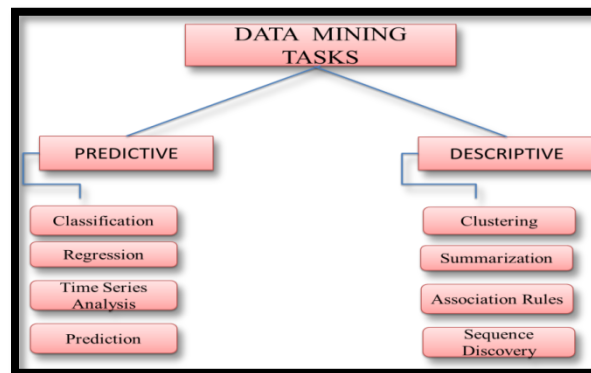
I. INTRODUCTION

Heart Diseases are expanding quickly in the present age. It has become the most common malady affecting individuals of all age groups indicating the risk level of individuals' life. The factors contributing to the increased risk of cardiovascular diseases are sedentary lifestyle and stress. All these factors result in a complex dataset. These days everybody is confronting an issue, for example, the administration, corporates, and industrial communities because of the ever-expanding number of databases. These databases should be managed and investigated. The databases contain crude dataset which comprises of conflicting and repetitive information. Data mining techniques can assist in uncovering the valuable data from dataset which apply to coronary illness forecast.

I. DATA MINING IN MEDICAL FIELD

In the healthcare sector , vast amount of information is generated every day, which is stored in the medical databases.

The health care system is no doubt very rich in aspects of storing data but at the same time, inferior in fetching knowledge. This database contains not only information of heart patients but also some irrelevant and inconsistent data. When the dataset becomes more massive and complex then algorithms for data mining(Decision tree, Naïve Bayes, Neural Network) are used. The task of data mining is generally divided into two types one is predictive and the other is descriptive. The first task utilizes the variables which can be used to predict the unknown values of the other variables of interest. The second task includes clustering, summarization, association rules, and sequence discovery.



II. Knowledge Data Discovery (KDD)

KDD is mainly used to express an issue that is made by digital information, which is an unacceptable fact for every individual: data overload [3]. KDD allows the entire process of extracting valuable information from the dataset, and data mining alludes only a particular step in the process of KDD. The data warehousing sets the Platform of KDD in two critical ways such as data cleaning and data access.

(1) Data cleaning: It is mainly used when the database contains all irrelevant, noise and errors, inconsistent data as well as missing data. To make this irrelevant data as relevant data, data cleaning is performed to obtain useful information.

(2) **Data access:** It is capable enough to access the data from the database, and well known methods should be made just for retrieving the data from the database and provides a path to access the data that are very hard to obtain (for example offline storing data).

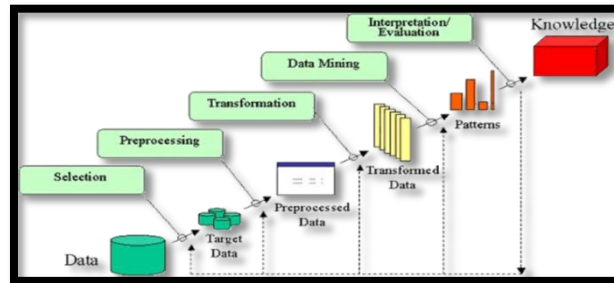
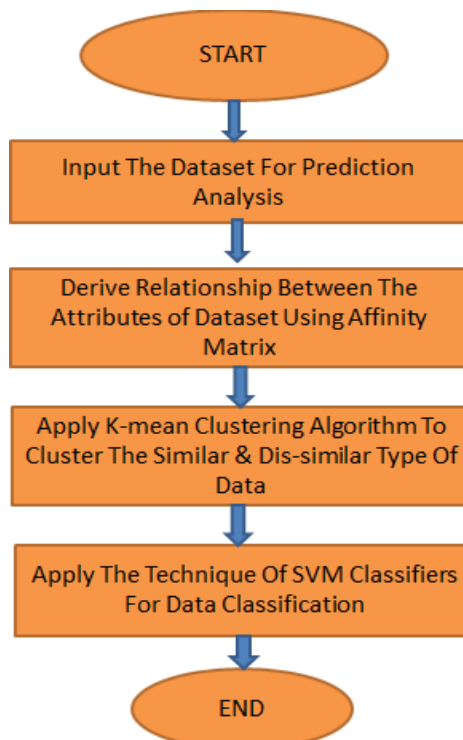


Figure 1: KDD PROCESS

III. WORK FLOW DESIGN



IV. RELATED WORK

A) Ranganatha.S et.al “Analysis For Heart Disease and Medical Data Mining Dataset Using Techniques Of Classification, (2013)”

The paper gives an insight into applying algorithms such as ID3 and Naive Bayes to the data of patients who are hospitalized for coronary diseases. The results obtained are in the human-understandable words and charts. Conclusion : Due to the immense measure of information which is produced ,algorithms are designed to give results once a query is submitted . The accomplishments of the created framework are:

- Less complex and minimal manual work.
- Storage of large volume of data
- Smooth work process.

B) Syed Umar Amin , et.al “ Data Mining Based on Genetic Neural Network In Prediction Of Heart Disease By Utilizing Risk Factors,(2013)”:

This paper represents a method for predicting coronary illness utilizing significant risk factors. An advanced model of Neural Networks and Genetic Algorithms is proposed, which may enhance the efficiency of the Artificial Neural Network. Conclusion: By utilizing hybrid procedures for data mining better clinical decision support systems can be designed for the detection of the diseases. We can fabricate an intelligent framework which could anticipate the ailment utilizing risk factors henceforth sparing cost and time to undergo medical tests and guaranteeing that the patient can screen his wellbeing all alone and plan preventive measures and treatment at an early stage of the ailment .

C) Theresa Princy et.al “Prediction Of Human Heart Disease System Using Techniques Of Data Mining,(2016)”:

This paper gives an overview of using various data mining methods such as Naïve Bayes, KNN, Decision Tree Algorithm, Neural Network for anticipating the hazard level of every individual depending on age, sexual orientation, circulatory strain, cholesterol, beat rate. The precision of the hazard level is high when utilizing number of characteristics progressively. End: Using KNN and ID3 calculation the hazard pace of coronary illness was recognized, and precision level likewise accommodated a distinctive number of properties. The future scope includes reducing the number of parameters and better precision by utilizing some different calculations.

D) K.Srinivas et.al “Prediction And Analysis Of Coronary Heart Disease In Coal Mining Regions Using The Techniques Of Data Mining,(2010)”:

This study analyzes the Behavioral Risk Factor Surveillance System, study to test whether number of people who themselves report about cardiovascular diseases are higher in Singareni coal mining locales in Andhra Pradesh state, India, contrasted with different districts after control for different dangers. The data mining procedures, Decision Trees, Naïve Bayes and Neural Network are utilized for forecast of coronary illness. Conclusion:

- We consolidate various kinds of properties like individuals' financial condition, contamination, mental stress and medical history.
- The methods like Clustering, Time Series and Association Rules are utilized to ascertain the behavior of the patients
- Further, it can be used to study the condition of the patient's morbidity in the health care centers

E) R. Kavitha et.al “Framework Used In Data Mining For Heart Disease Classification Using Feature Selection And Feature Extraction Technique,(2016)”:

In this research paper,principal component analysis (PCA) is used to design a system which derives the characteristics and mathematical model. It would help in improving the proficiency, exactness and speed of the procedure .The designed system would be useful in data retrieval, image processing and pattern matching. Result: The efficiency of the framework is improved in comparison to the other scoring function like Euclidean distance and Pearson correlation coefficients. The future work suggested aims to improve the accuracy by addressing the outlier present in the class boundary.

F). Eman Abu Khousa et.al “Predictive Data Mining To Support Clinical Decisions: An Overview Of Heart Disease Prediction Systems,(2012)”:

This paper surveys modelsdeveloped out of data mining procedures to help clinical decisions in (Heart Disease) diagnosis and prediction. The models provide automatic pattern recognition and endeavors to reveal relationship among various parameters and symptoms of Heart Disease.

Every framework has its own characteristics and constraints depending upon the data it handles, accuracy, and simplicity of interpretation, reliability and generalization ability. Result: Each model derives a significant pattern for the detection, for example drawbacks and strengths, interpretation of simplicity, data accuracy and speculation ability and reliability.

G). Monika Gandhi et.al “Predictions in Heart Disease Using Techniques of Data Mining”:

This paper gives information regarding the different knowledge abstraction techniques by utilizing strategies for data mining which are being used for prediction of coronary illness. In this research, strategies for data mining like Naïve Bayes and Neural Networks will be applied on medical datasets. Result:

- Every data mining techniques have focal points just as inconveniences of characterization of data and extraction of knowledge
- Moreover, decision tree, Networks of Neural or Naive Bayes can be analyzed further, which can be utilized in future in healthcare industry.

H) HeonGyu Lee et.al “A Data Mining Approach For Coronary Heart Disease Prediction Using HRV Features And Carotid Arterial Wall Thickness,(2008)”:

The aim of this research is to design and later on propose a method valuable in building up the different features of heart rate variability (HRV) and carotid arterial wall thickness supportive in diagnosing cardiovascular malady. Result: HRV can be counted as a such as extended Bayesian classifiers, CMAR, SVM and MDA. From the yield of investigations, the techniques for classification, for example, SVM and CPAR surpasses the others.

I). Limia Abed et.al “Using Data Mining Technique To Diagnosis Heart Disease,(2012)”:

In this paper, an experiment was conducted using Naive Bayes to build a predictive model as a virtual analyze for coronary disease dependent on dataset containing parameters that were recently evaluated for individuals. It was trailed by standing out the results from the aftereffects of various strategies which used comparable data from UCI store. The outcome and the responses obtained from viable work met the expectations. The exactness of this model achieved 100 % ratio. This result was compared with other experiments that were performed on the same

data set. Result: A critical point in the zone of diagnosis field is that every disease has its own particular characteristics that guide in diagnosing, and are unique in relation to other ailment. Best outcomes were given in diagnosing variations from the norm of heart utilizing cardiovascular Single Proton Emission Computed Tomography (SPECT) pictures with naive bayes classifier .It may not perform similarly with another sickness that has different data space.

V. CONCLUSION

The aim of this research is to detect the heart diseases by using various data mining techniques by performing prediction analysis. These techniques help in digging meaningful information from large amount of data. In this study, the improved K-mean clustering technique is used for the given datasets in calculating the arithmetic mean of input dataset from the centroid point and then Euclidian distance is also calculated from the same centroid point. As the dataset becomes complex, k means algorithm can help in deriving a close relationship between the attributes of the dataset to predict the heart diseases accurately in health care organizations. In further study, these attributes of the dataset can be reduced in number and accuracy level can be increased by using some of the algorithms of data mining.

VI. REFERENCES

- [1] Andritsos, P., Tsaparas, P., Miller, R.J. and Sevcik, K.C., “Limbo: Scalable Clustering of Categorical Data”, Extending Database Technology (EDBT), 2004.
- [2] Han, J. and Kamber, M., “Data Mining Concepts and Techniques”, Morgan Kaufmann, 2001.
- [3] Anil, K. Jain, and Richard, C. Dubes., “Algorithms for Clustering Data”, Prentice-Hall International, 1988.
- [4] AnjanGoswami, Ruoming Jin, and GaganAgrawal, “Fast And Exact K-Means Clustering”, International Conference on Data Mining (ICDM), 2004.
- [5] Huang, Z., “Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values”, Data Mining and Knowledge Discovery, 1998.

- [6] Hartigan, J. A. and Wong, M.A., “A k-means clustering algorithm”, Applied Statistics, pp. 100–108, 1979.
- [7] Subramanian, D.K., Narasimha Murthy, M., Vijaya, P.A., “Leaders-Subleaders: An Efficient hierarchical clustering algorithm for large data sets”, ELSEVIER, PP. 505–513, 2004.
- [8] Danyang Cao, Bingru Yang, “An improved k-medoids clustering algorithm”, Computer and Automation Engineering (ICCAE), 2010
- [9] Chen, H.L., Chuang, K.T and Chen, M.S., “Labeling Un clustered Categorical Data into Clusters Based on the Important Attribute Values”, IEEE International Conference on Data Mining (ICDM), 2005.
- [10] J.Zhu, H.Wang, M.Zhu, B.K.Tsou, and M.Ma,,” Aspect-Based Opinion Polling from Customer Reviews,” T. Affective Computing2(1):pp. 37- 49, 2011.
- [11] M.Karamibekr,A.A.Ghorbani,”Verb Oriented Sentiment Classification,” Processed of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT), Vol (1): pp.327-331, 2012.
- [12] Ranganatha S., Pooja Raj H.R., Anusha C., Vinay S.K.,” MEDICAL DATA MINING AND ANALYSIS FOR HEART DISEASE DATASET USING CLASSIFICATION TECHNIQUES”, 2013, IEEE, 32987u-45096-45
- [13] Syed Umar Amin, KavitaAgarwal, Dr. Rizwan Beg,” Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors”, 2013 IEEE Conference on Information and Communication Technologies (ICT) 978-1-4673-5758
- [14] Theresa Princy, R, J. Thomas,” Human Heart Disease Prediction System using Data Mining Techniques”, 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT], 978-1-5090-1277-0
- [15] K.Srinivas, Dr.G.RaghavendraRao, Dr. A.Govardhan,” Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques”, 2010, IEEE, 978-1-4244-6005-2
- [16] R.Kavitha, E.Kannan,” An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining”, 2016, IEEE, 29753-48506-65-7

- [17] EmanAbuKhousa, Piers Campbell,” Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems”, 2012 International Conference on Innovations in Information Technology (IIT), 978-1-4673-1101-4
- [18] Monika Gandhi, Dr. Shailendra Narayan Singh,” Predictions in Heart Disease Using Techniques of Data Mining”, 2015 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management (ABLAZE)
- [19] HeonGyu Lee, Ki Yong Noh, and Keun Ho Ryu,” A Data Mining Approach for Coronary Heart Disease Prediction using HRV Features and Carotid Arterial Wall Thickness”, 2008 International Conference on BioMedical Engineering andInformatics, 978-0-7695-3118-2
- [20] Lamia AbedNoorMuhammed,” Using Data Mining technique to diagnosis heart disease”, 2012, IEEE, 3545-56445-4
- [21] SellappanPalaniappan, RafiahAwang,” Intelligent Heart Disease Prediction System Using Data Mining Techniques”, 2008, IEEE, 978-1-4244-1968-5
- [22] Mai Shouman, Tim Turner, Rob Stocker,” USING DATA MINING TECHNIQUES IN HEART DISEASE DIAGNOSIS AND TREATMENT”, 2012, IEEE, 978-1-4673-0484-9
- [23] Sivagowry .S, Dr. Durairaj. M and Persia.A,” An Empirical Study on applying Data Mining Techniques for the Analysis and Prediction of Heart Disease”, 2013, IEEE, 23543-56-465