PREDICTION OF HEART DISEASE

Author: Anuradha Srinivasa Raghavan, and Vincy Joseph.

Abstract:

The heart is important organ of human body part. Life is completely dependent on efficient working of the heart. What if a heart undergoes a disorder, cardiovascular diseases are the most challenging disease for reducing patient count. According to survey conducted by WHO, about 17 million people die around the globe due to cardiovascular diseases i.,e 29.20% among all caused death, mostly in developing countries. Thus there is a need of getting rid of the this complicated task CVD using advanced data mining techniques, in order to discover knowledge of Heart disease prediction. In this paper, we propose an efficient hybrid algorithmic approach for heart disease prediction. This paper serves efficient prediction technique to determine and extract the unknown knowledge of heart disease using hybrid combination of K-means clustering algorithm and artificial neural network. In our proposed model we considered 14 attribute out of 74 attributes of UCI Heart Disease Data Set [19]. This technique uses medical terms such as age, weight, gender, blood pressure and cholesterol rate etc for prediction. To perform grouping of various attributes it uses k-means algorithm and for predicting it uses Back propagation technique in neural networks. The main objective of this paper is to develop a prototype for predicting heart diseases with higher accuracy rate.

Introduction:

At the age above 30, the heart attack or CVD is a common problem can be seen in all human beings. Along with changing lifestyle there are many such factors such as smoking, alcohol, cholesterol level, obesity, high blood pressure, diabetes etc. which are responsible factors for the risk of having heart problems. However, resent studies says that, with the introduction of artificial intelligence and medical sciences, we can actually help in preventing any such kind of diseases. Data mining plays a vital role in healthcare domain. Data Mining and Machine learning comes up as an emerging field of high importance for providing prognosis and a deeper understanding of medical data. In an old survey. The World Health Organization (WHO) has evaluated that 17 million deaths occur in world, every year due to the Heartdiseases. Prediction by using data mining techniques gives us accurate result of Heart Diseases The prediction can solve complicated queries for detecting heart disease and thus assist medical practitioners to make smart clinical decisions. Researchers are suggesting that applying data mining techniques in identifying effective treatments for patients can improve practitioner performance. Researchers have been investigating and applying different data mining techniques in the diagnosis of heart disease to identify which data mining technique can provide more reliable accuracy. Different data mining techniques have been used to help health care departments in the diagnosis of heart disease . Those most frequently used focus on classification: Naïve Bayes, decision tree, and neural network. In such one of the systems, has used Back-Propagation in neural network which is stated as the best prediction algorithm. The system shows a

non-linear relationship between the data and the target output. The characteristics of BP algorithm are that it is adaptive and tolerant towards the noisy data or other outliers present in the medical data. In our proposed system, we are proposing a hybrid approach to predict or diagnose heart disorders using UCI heart disease dataset [11]. by combining K-means and ANN algorithm. The main goal is to obtain high accuracy rate of prediction. Flow of the paper is given as; after the introduction section, a proper literature survey is made in section II. Section III specifies the proposed system architecture and flow chart of the implementation steps. Next section specifies about the steps required for K means and ANN algorithm. In section IV experiment result is summarized.

Methodlogy:

In this section, Data mining techniques used for decision making in heart disease are analysed. Ankita Dhewan and Meghana Sharma proposed a methodology of hybridizing two data mining techniques like Artificial Neural Network and Genetic Algorithm which was implemented to achieve high accuracy with least error [1]. Limitations: The very big disadvantages of GA are unguided mutations. The mutation operator in GA functions like adding a randomly generated number to a parameter of an individual of the population [10]. This is the only reason of a very slow convergence of genetic algorithm. The time consumed for optimization is much high. M.Akhil jabbar, B.L Deekshatulua proposed algorithm into two parts i.e first part deals with evaluating attributes using genetic search and second part deals with building classifier and measuring accuracy of classifier. In this paper it compares the accuracy of datasets with and without GA. Results shows that accuracy is increases by 5% when this two are combined. Limitations: Accuracy is very low with K-nearest neighbour and genetic algorithm takes much more time for optimization [7]. Rovina Dbritto, Aniruddha has given three data mining techniques viz. Naïve Bayes, Support Vector Machine, K-nearest neighbour and Logistic regression. Results shows that Naïve Bayes gives more accuracy compared to other classifier even. Limitations: The disadvantage is that the Naive Bayes classifier makes a very strong assumption on the shape of your data distribution, i.e. any two features are independent given the output class. Due to this, the result can be very bad. Dependencies among attributes cannot be modelled using Bayesian classifier [2]. Humar Kahramanli, Novruz Allahverdi, used a hybrid neural network that includes artificial neural network and fuzzy neural network. A datasets of 303 samples were taken from patients with heart disease which give 87.4% accuracy on attributes of UCI repository[5] [12]. Limitations: When fuzzy system is combined with neural network, fuzzy systems need to be tuned which is very time consuming and error-prone. Sudha, Sarath Kumar proposed two algorithms i.e KNN and K-means [15]. Their accuracy was measured which shows that KNN achieve 100% accuracy for different cluster with nearest value while K-means achieves 100% accuracy when value of K number of cluster have is very high. Limitations: Computation cost is very high as we have to calculate the distance of each query instance to all training samples [8] Mai Shouman, Tim Turner used a single data mining techniques on different datasets which shows that results can't be compared because of use of different datasets. When single and hybrid data mining techniques on Cleveland datasets in heart disease diagnosis results shows that hybrid techniques shows better results than single techniques. The hybrid technique used was Neural Network ensemble [3]. Limitations: Ensemble training is several times slower than traditional neural network. When solving some rare problems, the ensemble error is greater than error of a traditional neural network. M. Veera Krishna and S. Prem Kumar proposed different data mining techniques like KNN, K-means, Apriori, PLS-DA. Comparison is done on Performance measures like computational time, positive precision values, negative precision values etc. Results

shows that PLS-DA outperforms based on all performance measures. Limitations: PLS-DA is a complex algorithm which is very difficult to use

Conclusion:

As heart disease patients are increasing every year, huge amount of medical data is available. Researchers are applying data mining techniques on this data to diagnosis heart disease. It is analysed that artificial neural network algorithm is best for classification of knowledge data from large amount of medical data. Population is growing in exponential way. Death rate due to cardiovascular diseases is also increasing. The only solution to control this is to predict the heart disease and medicate it before it gone worse. Our hybrid approach gives higher accuracy rate of 97% of disease detection than earlier proposed method.