TITLE : Heart disease Analyses And Prediction

S.NO	NAME	MAIL ID	CONTACT NUMBER
1	K.L.N.SAI	99210041570@klu.ac.in	6300732764
2	R.AKASH	99210041413@klu.ac.in	7673986206
3	T.YUVA KISHORE	99210042119@klu.ac.in	8143936297
4	P.YASWANTH GUPTA	99210041442@klu.ac.in	7569099553

Department of computer science and engineering , Kalasalingam academy of research and education

DESCRIPTION

Heart related diseases are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, havebeen using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases. This paper presents a survey of various models based on such algorithms and techniques and analyze their performance. Models based on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Decision Trees (DT), Random Forest (RF) and ensemble models are found very popular among the researchers..

1. Introduction

Heart is an important organ of the human body. It pumps blood to every part of our anatomy. If it fails to function correctly, then the brain and various other organs will stop working, and within few minutes, the person will die. Change in lifestyle, work related stress and bad food habits contribute to the increase in rate of several heart related diseases.

Medical organisations, all around the world, collect data on various health related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend, can be easily explored using various machine learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately.

2. Dimensionality Reduction

Dimensionality Reduction involves selecting a mathematical representation such that one can relate the majority of, but not all, the variance within the given data, thereby including only most significant information. The data considered for a task or a problem, may consists of a lot of attributesor dimensions, but not all of these attributes may equally influence the output. A large number of attributes, or features, may affect the computational complexity and may even lead to overfitting which leads to poor results. A. Feature Extraction

In this, a new set of features is derived from the original feature set.Feature extraction involves a transformation of the features. This transformation is often not reversible as few, or maybe many, useful information is lost in the process.In [3]and[4]Principal Component Analysis (PCA)is used for feature extraction. Principal Component Analysis is a popularly used linear transformation algorithm. In the feature space, it finds the directions that maximize variance and finds directions that are mutually orthogonal. It is a global algorithm that gives the best reconstruction. B. Feature Selection

In this, a subset of original feature set is selected. In [5],key features are selected by CFS(Correlation based Feature Selection) Subset Evaluation combined with Best First Search method to reduce dimensionality. In [6]chi-square statistics test is used to select the most significant features.

3. Algorithms and Techniques Used

K – Nearest Neighbour

In 1951, Hodges et al. introduced a non-parametric technique for pattern classification which is popularly known the K-Nearest Neighbour rule[13]. K-Nearest Neighbour technique is one of the most elementary but very effective classification techniques. It makes no assumptions about the data and is generally be used for classification tasks when there is very less or no prior knowledge about the data distribution. This algorithm involves finding the k nearest data points in the training set to the data point for which a target value is unavailable and assigning the average value of the found data points to it. In [10] KNN gives an accuracy of 83.16% when the value of k is equal to 9 while using 10-cross validation technique. In [14]

KNN with Ant Colony Optimization performs better than other techniques with an accuracy of 70.26% and the error rates is 0.526.Ridhi Saini et al. have obtained a efficiency of 87.5% [15], which is very good.

4. Acknowledgment

We sincerely thank the staff of SRM Institute of Science and Technology, that have provided their immense support and guidance throughout the project.

5. Conclusion

Based on the above review, it can be concluded that there is ahuge scope for machine learning algorithms in predicting cardiovascular diseases or heart related diseases. Each of the above-mentioned algorithms have performed extremely well in some cases but poorly in some other cases. Alternating decision trees when used with PCA, have performed extremely well but decision trees have performed very poorly in some other cases which could be due to overfitting. Random Forest and Ensemble models have performed very well because they solve the problem of overfitting by employing multiple algorithms (multiple Decision Trees in case of Random Forest). Models based on Naïve Bayes classifier were computationally very fast and have alsoperformed well.SVM performed extremely well for most of the cases. Systems based on machine learning algorithms andtechniques have been very accurate in predicting the heart related

Reference

[1] Ramadoss and Shah B et al.“A. Responding to the threat of chronic diseases in India”. Lancet. 2005; 366:1744–1749. doi: 10.1016/S0140-6736(05)67343-6.

[2] Global Atlas on Cardiovascular Disease Prevention and Control. Geneva, Switzerland: World Health Organization, 2011

[3] Dhomse Kanchan B and Mahale Kishor M. et al. “Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis”, 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication.

[4] R.Kavitha and E.Kannan et al. “An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining “, 201