Abstract

Heart Disease Dataset (HDD) contains high dimensions which poses challenges to research community in terms of complexity and efficient analysis. Heart disease is also called as cardiovascular disease (CVD). Feature selection will be made to reduce the irrelevant and redundant number of attributes. Fast diagnosis of the heart disease can be done using a knowledge driven approach. A comparison was made for medically important features to that of computerized subset of features, to bring out much simpler set of features used for the diagnosis. It focuses on the experts' judgement for medical driven feature selection process termed as MFS, and the performance of various classifiers on Cleveland dataset for the computerized feature selection termed as CFS and also a combination of both to enhance the prediction accuracy. Further, this paper categorizes the MFS, CFS and the combination of both into discrete and continuous sets of attributes. Our work has proved that the discrete features do not contribute much to the classification as do the continuous ones, in its accuracy, speed and performance.
References

15. David W. Aha & Dennis Kibler. "Instance-based prediction of heart-disease presence with the Cleveland database."


Index Terms

Computer Science

Biomedical
Keywords

Medical Feature Selection, Computerized feature selection, SMO