Abstract

Objectives: The proposed work is to classify breast cancer with few attributes. Reducing the attributes reduces the time, so that the patient need not wait for result for a long time. For classification, the user friendly environment is created. The user can enter the details of the patient such as Clumpthickness, Uniformity in cell size etc., and the result is classified as benign or malignant. Statistical analysis: Variable selection is done by one of the variable reduction algorithm called Linear Discriminant Analysis (LDA). LDA is one of the statistical method. The dataset is passed to LDA function repeatedly and the combination of variables which gave the good accuracy is selected. The variables that are selected by using LDA are used in classifying breast cancer. Findings: This application is created to find whether the given record is benign or malignant tumor. In this proposed work, the dataset from UCI repository for breast cancer detection is used. There are many other works done for finding breast cancer risk, diagnosing breast cancer etc., and there may be at least ten variables used for classification which may be time consuming. But in this proposed work, only four are used and it gave the accuracy of up to 96%. Hence this may be the first step or idea for detecting breast
cancer with lesser variables, so that this may be helpful for the doctors. Improvements: The proposed work is done based on the UCI machine learning repository dataset, which was uploaded by Wisconsin Hospitals, Madrid. Some changes can be made in the coding and this methodology can also be implemented in other dataset also by reducing the attributes.

References

16. Lichman,M.,‘UCI Machine learning Repository’,
https://archive.ics.uci.edu/ml/datasets/BreastCancerWisconsin (Diagnostic),University of
California, Irvine,CA,2013.
17. Data preprocessing techniques for data
18. ‘Minitab-What is Mahalanobis distance’
19. ‘Naive Bayes classification algorithm’, http://software.ucv.rp/air/docs/naivebayes.pdf,
Feb 2015.

Index Terms

Computer Science Information Sciences

Keywords

Classification, Mahalanobis, Normalization, Fisher, data-preprocessing