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

1. Saravanakumar K. and Arthanariee A. M. Evaluate the multiple breast cancer factors and
calculate the risk by software tool breast cancer risk evaluator, Indian Journal of Science and
Technology, 3 (Apr 2015), 686-91.

2. Vaidhehi K and Subhashini T.S, Breast Tissue Characterization using combined K-NN
classifier, Indian Journal of Science and Technology, 8 (Jan 2015), 23-26.

Transforms, Indian Journal of Science and Technology, 8 (Oct 2015), 1-7

4. Singh S, Dr. Gupta, P.R. and Sharma M.K, Breast Cancer Detection and Classification
using Histopathological images, International Journal of Engineering Science and Technology, 3
(May 2011), 4-9.

5. Dumitru D. 2009 Prediction of recurrent events in breast cancer using the Naïve Bayesian
classification, Annals University of Craiova, Mathematics and Computer Science Series.


26-31.

8. Andrews J L and McNicholas P D, Variable selection for clustering and classification,

9. You H and Rumbe G, Comparative Study of Classification Techniques on Breast Cancer
FNA Biopsy Data, International Journal of Interactive Multimedia and Artificial Intelligence, 1
(Dec 2010), 6-13.

Naïve Bayes Machine Learning Classifier with KNN Missing Data Imputation, 3rd World
conference on innovation and Computer Sciences.

for Classification of Breast Cancer Data, Machine Learning and Applications, ICMLA ’08.
Seventh International Conference.

Classification for Diagnosing Breast Cancer, “Proceedings of the IASTED International
Conference of artificial intelligence and soft computing”.

13. Kitbumrungrat K, Comparison Logistic Regression and Discriminant Analysis in
classification groups for Breast Cancer, International Journal of Computer Science and Network


15. Nancy S G and Dr. Appavu alias Balamurugan S, A Comparative Study of Feature
Selection Methods for cancer classification using Gene Expression Dataset, Journal of
Computer Applications, 6 (Sep 2013), 78-84.


Index Terms

Computer Science			Information Sciences

Keywords

Classification, Mahalanobis, Normalization, Fisher, data-preprocessing