

3rd International conference on

ARTIFICIAL INTELLIGENCE & ROBOTICS

June 28-29, 2017 San Diego, USA

Identification of Gene Biomarkers For Cancer Tumour Classification using Cross-validated Area Under The ROC Curve**W. B. Yahya**

Department of Statistics, University of Ilorin, Nigeria

Microarray-based cancer classifications using gene expression signatures has been embraced as a viable alternative to clinical identification and diagnosis of cancer tumours. However, the efficiency of the various gene-based classifiers depends largely on the goodness of the crop of genes selected and employed for tumour prediction. Thus, one of the common challenges in microarray studies is how to select the crop of genes subset that would be highly predictive of the tissue samples and make biological sense. In this study, an efficient primary gene selection (filtering) method that employs the area under the receiver operating characteristic (ROC) curves for feature selection is presented for binary response microarray data. Gene candidates were selected based on their individual univariate predictive strength of the two tumour subgroups as measured by their respective estimated areas under the ROC curves over a 10-fold cross-validation. Results of the hierarchical clustering with complete linkage search and principal component analysis employed on the selected gene signatures showed a good discrimination of the two biological groups based on the expression levels of the selected gene biomarkers via Monte Carlo experiments. The method, when applied on published lung cancer data set, efficiently classified the two subtypes of lung cancer tumours; *malignant pleural mesothelioma* (MPM) and *adenocarcinoma* (ADCA) based on the expression profiles of few selected features from the entire 12,533 genes biomarkers that were measured on 181 mRNA samples. It can be concluded that the new feature selection method proposed here is quite efficient at selecting informative gene inputs that can be further employed by any standard machine learning methods for proper classification of mRNA samples into their respective tumour subgroups in any binary response microarray data problem.

dr.yah2009@gmail.com

Notes: