

International Conference on **Big Data Analysis and Data Mining**

May 04-05, 2015 Kentucky, USA

Identification of gene biomarkers for cancer tumor classification using cross-validated area under the ROC curve

Waheed Babatunde Yahya^{1,2} and Alexander Hapfelmeier²

¹University of Ilorin, Nigeria

²Technical University of Munich, Germany

Microarray-based cancer classifications using gene expressions of biological samples 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 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 was applied on published lung cancer data set and it efficiently classified the two subtypes of lung cancer tumours; malignant pleural mesothelioma (MPM) and adenocarcinoma (ADCA) based on the expression profiles of 12,533 genes biomarkers that were measured on 181 mRNA samples. The feature selection method presented here efficiently selects 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.

Biography

Waheed Babatunde Yahya is the Head of Department of Statistics, University of Ilorin, Nigeria and his area of specialization is microarray analysis, multiple hypothesis testing, biostatistics etc.

wb_yahya@daad-alumni.de, wbyahya@unilorin.edu.ng

Notes: