
Original Article

**Diagnosis of Breast Cancer Subtypes using the Selection of Effective Genes from Microarray Data**

Tabatabaei A¹, Derhami V¹*, Sheikhpour R², Pajoohan M-R¹

¹ Computer Engineering Department, Faculty of Engineering, Yazd University, Yazd, Iran
² Department of Computer Engineering, Faculty of Engineering, Ardakan University, Ardakan, Iran

**Abstract**

**Introduction:** Early diagnosis of breast cancer and the identification of effective genes are important issues in the treatment and survival of the patients. Gene expression data obtained using DNA microarray in combination with machine learning algorithms can provide new and intelligent methods for diagnosis of breast cancer.

**Methods:** Data on the expression of 9216 genes from 84 patients across 5 different types of cancer was obtained using microarray technology. In this study, we proposed a feature selection method based on the correlation between abnormal expression of genes and cancer for diagnosis of breast cancer. Then, we used K-nearest neighbor (KNN), support vector machine (SVM), and naive Bayesian (NB) classifiers to evaluate the performance of the proposed method in the selection of relevant genes.

**Results:** The proposed feature selection method coupled with the KNN classifier predicted all types of cancer with 100% accuracy and using 38 of the 9216 genes. The proposed method could also identify the genes associated with each class. Moreover, the proposed feature selection method coupled with NB and SVM classifiers achieved accuracy rates of 90% and 96.67% using 17 and 22 genes, respectively.

**Conclusion:** The results of this study demonstrated that the proposed feature selection method has better performance compared with other methods. The proposed method is able to distinguish the genes involved in each cancer class and detect overexpression or underexpression of selected genes, which can be used by physicians and researchers in the field of health care.

**Keywords:** Breast Cancer, Feature Selection, Microarray Data, Classification