

An improved parallelized mRMR for gene subset selection in cancer classification

ABSTRACT

DNA microarray technique has become a more attractive tool for cancer classification in the scientific and industrial fields. Based on the previous researchers, the conventional approach for cancer classification is primarily based on morphological appearance of the tumor. The limitations of this approach are bias in identify the tumors by expert and faced the difficulty in differentiate the cancer subtypes due to most cancers being highly related to the specific biological insight. Thus, this study propose an improved parallelized Minimum Redundancy Maximum Relevance (mRMR), which is a particularly fast feature selection method for finding a set of both relevant and complementary features. The mRMR can identify genes more relevance to biological context that leads to richer biological interpretations. The proposed method is expected to achieve accurate classification performance using small number of predictive genes when tested using two datasets from Cancer Genome Project and compared to previous methods.

Keyword: Feature selection; Cancer classification; mRMR filter method; Parallelized mRMR; Random forest classifier