Abstract ID: *ASRS2018 – 21*

A NOVEL FILTER-WRAPPER BASED FEATURE SELECTION APPROACH TO ENHANCE THE ACCURACY OF CANCER CLASSIFICATION

M.M. Mohamed Mufassirin[1*] and Roshan G. Ragel[2]

[1]Department of Mathematical Sciences, Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sammanthurai, Sri Lanka
[2]Department of Computer Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya, Sri Lanka
[*]mufassirin666@gmail.com

The improvement in DNA microarray technology is an important area of interest among many researchers and medical scholars to investigate the expression levels of enormous number of genes in a DNA simultaneously. It has been shown that the use of this technology is beneficial for cancer data classification. However, the DNA microarray data usually contains thousands of irrelevant and redundant gene information, which need to be eliminated to increase the classification accuracy. Usually, little mutated genes are responsible for cancer susceptibility. The objective of this study is to effectively select the relevant mutated gene information from cancer data to enhance the accuracy of cancer classification. Thus, in order to select the relevant gene information, a novel feature selection technique based on a filter-wrapper approach is proposed in this study. Wrapper approach chooses all possible subsets of features to evaluate useful features and provides the most informative subset which will increase the accuracy of the classifiers. On the other hand, filter methods extract features from the data without any learning involved. However, compared to filters, the computation demand of wrappers are high and therefore consume a massive amount of time when applied to microarray data. Hence, in the proposed work, the wrapper is applied after the filter approach with the intention of reducing the computational complexity of wrappers. The datasets were initially employed using a filter called Gain Ratio Filter to remove redundant and irrelevant genes from dataset with the Ranker search method, and then the resultant gene subsets were evaluated using a wrapper called Wrapper Subset Evaluator with the best first forward selection searching strategy using WEKA machine learning workbench. The selected gene subset by wrapper was then used to classify the cancer microarray using machine learning classifiers namely, Decision Tree (J48), Naïve Bayes, Sequential Minimal Optimization (SMO), Deep Learning and Bayes Net. The proposed approach was tested on five benchmark cancer microarray datasets. The accuracy of 89.69%, 95.16% and 97.04% were obtained for Breast, Colon and Lung cancer datasets respectively while Leukaemia and Ovarian cancer datasets scored 100%. As per the findings of this study, the proposed method is more efficient compared to the existing classification models

.

Keywords *- DNA Microarray, Machine Learning, Feature Selection, Classification*

**Corresponding Author*