An Improved gSVM-SCADL2 with Firefly Algorithm for Identification of Informative Genes and Pathways

Weng Howe Chan[a], Mohd Saberi Mohamad[a], Safaai Deris[b], Juan Manuel[c], Sigeru Omatu[d], Zuwairie Ibrahim[e]

[a]Artificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia,
[b]Faculty of Creative Technology & Heritage, Universiti Malaysia Kelantan, Locked Bag 01, 16300 Bachok, Kota Bharu, Kelantan, Malaysia
[c]Biomedical Research Institute of Salamanca/BISITE Research Group, University of Salamanca, Salamanca, Spain
[d]Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan
[e]Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

ABSTRACT

Incorporation of pathway knowledge into microarray analysis has been favoured by researchers owing to the improved biological interpretation of the analysis outcome. However, most of the pathway data are manually curated without specific biological context. Inclusion of non-informative genes in the analysis of context specific microarray data could lead to classifier with poor discriminative power. Thus, one of the main challenges is how to effectively identify informative genes from the pathway data. This paper proposes a firefly optimised penalised support vector machine with SCADL2 penalty function (SVM-SCADL2-FFA) in optimising tuning parameters for each pathway for efficient identification of informative genes and pathways. Experiments are done on lung cancer and gender data sets. Tenfold CV is used to evaluate the performance in terms of accuracy, specificity, sensitivity and F-score. The identified informative genes are validated through online databases. Our proposed method shows consistent improvements compared to previous works.