Das et al. (2017) Can J Biotech, Volume 1, Special Issue, Page 35
https://doi.org/10.24870/cjb.2017-a22

Canadian Journal of Biotechnology

ISSN 2560-8304

Poster Presentation **OPEN ACCESS**

Category: Bioinformatics

sigFeature: an R-package for significant feature selection using SVM-RFE & t-statistic

Pijush Das[1], Susanta Roychoudhury[1] and Sucheta Tripathy[1*]

[1]Computational Genomics lab, Structural Biology and Bioinformatics Division, CSIR- Indian Institute of Chemical Biology; Kolkata 700032, INDIA

Presenting author: topijush@gmail.com

Abstract

Depending on the sub-site of the primary tumour, up to thirty percent of the patients with clinical and radiological node negative HNSCC may have occult metastases. Therefore, currently, up to seventy percent patients with node negative neck disease receive unnecessary therapy to ensure a minority who are truly at risk [1]. The treatment of HNSCC involves surgery, radiotherapy or multimodality therapy like surgery together with adjuvant radiotherapy or chemo radiotherapy. HNSCC is typically considered as a homogeneous tumour group, i.e., histopathologically identical, but they are often genetically disparate and exhibit variable biological behaviour and response to treatment between and within anatomical sub-sites [2]. Currently, treatment decisions for patients with HNSCC are still based on clinical, radiological and pathologic parameters. No molecular markers are used for treatment decision, except in ongoing research protocols. To identify those patients who are truly at risk, a novel feature selection method has been introduced based on expressional genomic data in this study. In data mining, feature selection is an extremely dynamic field of research for classification in the field of machine learning technology. The aim of feature selection is to select a small subset of a feature from a larger pool, rendering not only a good performance of classification but also biologically meaningful insights. Filter methods e.g. the support vector machine recursive feature elimination (SVM-RFE) is recognised as one of the most effective methods. The RFE-SVM algorithm is a greedy method that only hopes to find the best possible combination for classification without considering the differentially significant feature between the classes. To overcome this limitation of SVM-RFE, our proposed approach which is based on RFE-SVM and t-statistic is to find out differentially significant features along with the good performance of classification. The experimental results which we obtained after analysing six publicly available micro array datasets are very promising and show the contribution in feature selection in machine learning technology. The main conclusion is that the selected features are differentially significant between the classes and able to produce good classification accuracy which will help further downstream analysis for strengthening the biological aspect.

References

[1] Vaish, R., Gupta, S. and D'Cruz, A.K. (2015) Elective versus therapeutic neck dissection in node-negative oral cancer. *N Engl J Med* 373: 2477. https://doi.org/10.1056/NEJMc1511351

[2] Rodrigo, J.P., Ferlito, A., Suarez, C., Shaha, A.R., Silver, C.E. et al. (2005) New molecular diagnostic methods in head and neck cancer. *Head & Neck* 27: 995-1003. https://doi.org/10.1002/hed.20257

Citation: Das, P., Roychoudhury, S. and Tripathy, S. sigFeature: an R-package for significant feature selection using SVM-RFE & t-statistic [Abstract]. In: Abstracts of the NGBT conference; Oct 02-04, 2017; Bhubaneswar, Odisha, India: Can J biotech, Volume 1, Special Issue, Page 35. https://doi.org/10.24870/cjb.2017-a22