

Comparison of partial least squares and random forests for evaluating relationship between phenolics and bioactivities of *Neptunia oleracea*

ABSTRACT

Background: *Neptunia oleracea* is a plant consumed as a vegetable and which has been used as a folk remedy for several diseases. Herein, two regression models (partial least squares, PLS; and random forest, RF) in a metabolomics approach were compared and applied to the evaluation of the relationship between phenolics and bioactivities of *N. oleracea*. In addition, the effects of different extraction conditions on the phenolic constituents were assessed by pattern recognition analysis. **Results:** Comparison of the PLS and RF showed that RF exhibited poorer generalization and hence poorer predictive performance. Both the regression coefficient of PLS and the variable importance of RF revealed that quercetin and kaempferol derivatives, caffeic acid and vitexin-2-O-rhamnoside were significant towards the tested bioactivities. Furthermore, principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) results showed that sonication and absolute ethanol are the preferable extraction method and ethanol ratio, respectively, to produce *N. oleracea* extracts with high phenolic levels and therefore high DPPH scavenging and α -glucosidase inhibitory activities. **Conclusion:** Both PLS and RF are useful regression models in metabolomics studies. This work provides insight into the performance of different multivariate data analysis tools and the effects of different extraction conditions on the extraction of desired phenolics from plants.

Keyword: *Neptunia oleracea*; Metabolomics; Partial least squares; Random forest; Phenolics; Extraction conditions