Evaluation of Ensemble Classifier (EC) Machine Learning Methods for Introduction of Breast Cancer Genomic Biomarkers

L. Mirsadeghi [1,*], K. Kavousi [2], R. Hajihosseini [1], A. Banaei-Moghaddam [2]

[1] *Payame Noor University, Tehran, Iran*
[2] *Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

* *Corresponding author: L. Mirsadeghi, 1 Payame Noor University, Tehran, Iran.*
E-mail: l.mirsadeghi@gmail.com

Abstract

Introduction: The prognosis and diagnosis of cancers are two of crucial issues in Precision Oncology. Researchers are investigating the possibility of using ensemble machine learning tools to help detect cancers including breast cancer. This study reviews ensemble classifier methods and their applications in breast cancer detection and compares their results. At first, we surveyed 14 ensemble classifier machine learning methods in related to breast cancer prognosis and diagnosis. Then, we selected five of these approaches that prioritize genes and introduce breast cancer genomic biomarkers. They are include of Multiple RFE (Recursive Feature Elimination), DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples), HyDRA (Hybrid Distance-score Rank Aggregation), GenEnsemble for 4 base classifiers (NBS, IB3, SVM and C4.5), and the collective approach for 4 individual methods (correlation, color palette, color proportion and SVM).

Conclusions: This study can help to researchers in using ensemble methods in field of molecular biology and breast cancer diagnosis to choose most suitable ensemble method according to their problem. Today, there are many approaches for this problem and other cancers, and a key question is "which proposed ensemble method is the best?" The results show there is not best approach for all classification problems, and the best solution depends on kind of problem, the structure of the available data and prior knowledge about related algorithm [7]. Finally, using machine learning is less costly than biomolecular testing. However, we suggest in order to provide better evaluation of accuracy of the results of this approaches, biomolecular techniques such as qRT-PCR and the next generation sequencing (NGS) are applied. In the light of this research, we hope to help for introducing of novel driver genes as genomic biomarkers for precision oncology to increase patients' life expectancy.