

Accepted Manuscript

Wisdom of artificial crowds feature selection in untargeted metabolomics: An application to the development of a blood-based diagnostic test for thrombotic myocardial infarction

Patrick J. Trainor, Roman V. Yampolskiy, Andrew P. DeFilippis

PII: S1532-0464(18)30049-2
DOI: <https://doi.org/10.1016/j.jbi.2018.03.007>
Reference: YJBIN 2948

To appear in: *Journal of Biomedical Informatics*

Received Date: 5 August 2017
Revised Date: 15 December 2017
Accepted Date: 12 March 2018

Please cite this article as: Trainor, P.J., Yampolskiy, R.V., DeFilippis, A.P., Wisdom of artificial crowds feature selection in untargeted metabolomics: An application to the development of a blood-based diagnostic test for thrombotic myocardial infarction, *Journal of Biomedical Informatics* (2018), doi: <https://doi.org/10.1016/j.jbi.2018.03.007>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Wisdom of artificial crowds feature selection in untargeted metabolomics: An application to the development of a blood-based diagnostic test for thrombotic myocardial infarction

Patrick J. Trainor^{a*}, Roman V. Yampolskiy^{b§}, Andrew P. DeFilippis^{c§}

^aDepartment of Medicine, Division of Cardiovascular Medicine, University of Louisville

^bDepartment of Computer Science and Engineering, University of Louisville

^cDepartment of Medicine, Division of Cardiovascular Medicine, University of Louisville

§Equally contributing authors

*Corresponding Author

Keywords

Evolutionary Computation, Wisdom of Artificial Crowds, Metabolomics, Feature Selection, Classification, Diagnostic Test, Myocardial Infarction

Abstract

Introduction: Heart disease remains a leading cause of global mortality. While acute myocardial infarction (colloquially: heart attack), has multiple proximate causes, proximate etiology cannot be determined by a blood-based diagnostic test. We enrolled a suitable patient cohort and conducted a non-targeted quantification of plasma metabolites by mass spectrometry for developing a test that can differentiate between thrombotic MI, non-thrombotic MI, and stable disease. A significant challenge in developing such a diagnostic test is solving the NP-hard problem of feature selection for constructing an optimal statistical classifier.

Objective: We employed a Wisdom of Artificial Crowds (WoAC) strategy for solving the feature selection problem and evaluated the accuracy and parsimony of downstream classifiers in comparison with traditional feature selection techniques including the Lasso and selection using Random Forest variable importance criteria.

Materials and Methods: Artificial Crowd Wisdom was generated via aggregation of the best solutions from independent and diverse genetic algorithm populations that were initialized with bootstrapping and a random subspaces constraint.

Results / Conclusions: Strong evidence was observed that a statistical classifier utilizing WoAC feature selection can discriminate between human subjects presenting with thrombotic MI, non-thrombotic MI, and stable Coronary Artery Disease given abundances of selected plasma metabolites. Utilizing the abundances of twenty selected metabolites, a leave-one-out cross-validation estimated misclassification rate of 2.6% was observed. However, the WoAC feature selection strategy did not perform better than the Lasso over the current study.

Download English Version:

<https://daneshyari.com/en/article/6927470>

Download Persian Version:

<https://daneshyari.com/article/6927470>

[Daneshyari.com](https://daneshyari.com)