

Accepted Manuscript

Gene selection in autism – comparative study

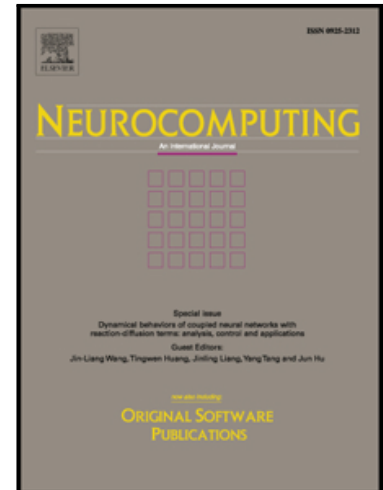
Tomasz Latkowski , Stanislaw Osowski

PII: S0925-2312(17)30234-5
DOI: [10.1016/j.neucom.2016.08.123](https://doi.org/10.1016/j.neucom.2016.08.123)
Reference: NEUCOM 18030

To appear in: *Neurocomputing*

Received date: 24 February 2016
Revised date: 21 June 2016
Accepted date: 16 August 2016

Please cite this article as: Tomasz Latkowski , Stanislaw Osowski , Gene selection in autism – comparative study, *Neurocomputing* (2017), doi: [10.1016/j.neucom.2016.08.123](https://doi.org/10.1016/j.neucom.2016.08.123)



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.

GENE SELECTION IN AUTISM – COMAPRATIVE STUDY

TOMASZ LATKOWSKI

Faculty of Electronics, Military University of Technology,
01-476 Warsaw, Kaliskiego 2, Poland
tlatkowski@wat.edu.pl

STANISLAW OSOWSKI¹

Faculty of Electrical Engineering, Warsaw University of Technology
00-661 Warsaw, Koszykowa 75, Poland
Faculty of Electronics, Military University of Technology,
01-476 Warsaw, Kaliskiego 2, Poland
sto@iem.pw.edu.pl

Abstract

The paper investigates application of several methods of feature selection to identification of the most important genes in autism disorder. The study is based on the expression microarray of genes. The applied methods analyze the importance of genes on the basis of different principles of selection. The most important step is to fuse the results of these selections into common set of genes, which are the best associated with autism. These genes may be treated as the biomarkers of this disorder and used in early prediction of autism. The paper proposes and compares three different methods of such fusion: purity of the clusterization space, application of genetic algorithm and random forest in the role of integrator. The numerical experiments are concerned with the identification of the most important biomarkers and their application in autism recognition. They show the applied fusion strategy of many independent selection methods leads to the significant improvement of the autism recognition rate.

Keywords

Autism; Gene expression microarrays; Feature selection; Clustering; Genetic algorithm; Random forest

1. Introduction

Autism disorder belongs to the pervasive neurodevelopmental disorders, affecting a broad spectrum of human functions [1,2]. The important problem is early recognition of this disorder, enabling the proper treatment of the autistic individuals. Nowadays, microarray gene expression data are studied to find the genes or sequences of genes which are the best associated with autism and might be treated as biomarkers. The difficulty in identifying these genes are many outliers, high variance of data and bad conditioning of the problem [1,3], manifested by the small number of available observations (usually measured in hundreds) in comparison to very huge number of genes (dozens of thousands).

These complexities raise the challenge of how to identify the genes, that are the most informative for this disorder and that can be used to distinguish the class of autistic from the other individuals. Many methods developed in feature selection have been used in solving the

¹ Corresponding author: Tel.: +22234 7235, fax: +22234 5642.

E-mail address: sto@iem.pw.edu.pl (S. Osowski)

URL: <http://www.iem.pw.edu.pl/~sto>.

Download English Version:

<https://daneshyari.com/en/article/4947308>

Download Persian Version:

<https://daneshyari.com/article/4947308>

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