Editorial

George M. Spyrou*

Biomedical Informatics Unit,
Biomedical Research Foundation,
Academy of Athens,
4 Soranou Efessiou Str., Athens 115 27, Greece
E-mail: gspyrou@bioacademy.gr
*Corresponding author

Vasilis Promponas

Bioinformatics Research Laboratory,
Department of Biological Sciences,
New Campus, Office B161, University of Cyprus,
P.O. Box 20537, CY 1678, Nicosia, Cyprus
E-mail: vprobon@ucy.ac.cy

Evangelia Micheli-Tzanakou

Department of Biomedical Engineering,
Rutgers University,
599 Taylor Rd, Piscataway, NJ, 08854, USA
E-mail: etzanako@rci.rutgers.edu

Athanasios Vasilakos

National Technical University of Athens,
Heroon Polytechniou 9, 15780 Zografou, Greece
E-mail: vasilako@ath.forthnet.gr

Biographical notes: George M. Spyrou received his BSc in Physics from University of Athens, Greece. He holds a Masters of Science in Medical Physics and in Bioinformatics as well. During his PhD in Medical Physics he worked on algorithms and simulations applied on medical issues, especially on breast cancer imaging. Currently, he is working as a Senior Research Scientist in the Biomedical Informatics Unit of the Biomedical Research Foundation of the Academy of Athens (BRFAA) leading the Modeling and Computational Intelligence in Biomedical Informatics Group. Also, he has been assigned as the Head of the Department of Informatics and New Technologies in BRFAA.

Vasilis J. Promponas received his PhD from the Department of Biology at the University of Athens, Greece. He is currently a Lecturer in Bioinformatics and Head of the Bioinformatics Research Laboratory in the Department of Biological Sciences at the University of Cyprus. He is a Reviewer for *Bioinformatics*, *PloS One*, *Briefings in Bioinformatics, Proteins, BMC Genomics*, a member of the International Society for Computational Biology (ISCB) and serves as Interim President of the Hellenic Society for

Computational Biology and Bioinformatics (HSCBB). His research interests include protein sequence analysis, prediction of protein structure/function, comparative genomics, computational approaches towards genome and protein evolution.

Evangelia Micheli-Tzanakou is Professor II and Director of the Computational Intelligence Laboratories in the Department of Biomedical Engineering at Rutgers University. She established the first-ever experimental brain to computer interface. She is the recipient of many awards and has published over 300 scientific papers. An IEEE Fellow, she has served in many positions including the IEEE Board of Directors as Director of Division X (Systems and Controls), President of the Computational Intelligence Society, Chair of the IEEE Awards Board and more recently as the IEEE VP of Educational Activities.

A.V. Vasilakos is currently Visiting Professor at the Graduate Programme of the Department of Electrical and Computer Engineering, National Technical University of Athens (NTUA).He has authored or co-authored over 200 technical papers in major international journals and conferences. He is author or co-author of five books and 20 book chapters in the areas of communications. He served as General Chair, TPC Chair and Symposium Chair for many international conferences. He served or is serving as an Editor or/and Guest Editor for many technical journals. He is founding Editor-in-chief of the journals: *International Journal of Adaptive and Autonomous Communications Systems* (*IJAACS*, http://www.inderscience.com/ijaacs), and *International Journal of Arts and Technology* (*IJART*, http://www.inderscience.com/ijart).

In the last decade, there was an explosion in the amount of biological data, since detailed information concerning the DNA, RNA and protein biomolecules had been produced through modern and high throughput experimental systems. So, the scientific community had to decide between sinking or swimming as the tidal wave of data was approaching. The existing information concerns the biomolecular sequences (primary structure), basic characteristics of their 3D structure (secondary structure), their 3D structure, their function, the interactions between them, their relationships and their dependencies, their regulation and expression, their location, their thermodynamic characteristics, etc. As it is obvious, acquiring, storing and handling these data is not enough. Not even analysing them is enough. There is the need of thoroughly mine from them the hidden knowledge that will lead us to complexity reduction and understanding of the fundamental biological mechanisms. Computational intelligence methods have been applied to solve problems such as: finding the coding regions in a genome sequence, identification of various genome elements, sequence comparison, multiple alignments, discovery of functional domains, protein structure prediction, structural alignments and production of interaction networks, docking and drug design, whole genome comparisons, discovery of metabolic pathways, analysis of high throughput gene expression and proteomics data, biomedical literature text mining and natural language processing. Many times the aforementioned issues are interconnected trying to answer to complex questions such as: given the gene sequence that is responsible for a disease, is there a suitable drug that will stop the undesired action of the specified gene? A chain of bioinformatics actions strongly incorporating computational intelligence methods is necessary (but not always sufficient). Such a chain should include the translation of gene

sequence to the corresponding amino acid sequence, similarity finding with other sequences of known structure, model and predict the structure of the protein that is the product of the defective gene and predict – through docking procedures – the drug that would optimally bind to the modelled protein stopping thus the effects of the problematic gene.

On the other hand, there are also questions concerning the optimum method of computational intelligence used and the answers (if any) are always related to the type of biological question and much more to the type of prior information and data that are going to be used as well. There is a plethora of pure or hybrid computational intelligence methodologies such as: expert systems, artificial neural networks, fuzzy logic and systems, feature selection methods, dimension reduction, pattern classification and recognition, possibility theory, Bayes network and hidden Markov models, genetic and hybrid evolutionary algorithms, support vector machines (SVMs), relative vector machine, data mining and knowledge discoveries, quantum-inspired evolutionary algorithms, fusion of systems, swarm intelligence and hybrid computational intelligence, affinity propagation and game theory. It is very challenging for the researchers in the bioinformatics field not only to select the best methodology that fits his/her problem and data, but to synthesise the aforementioned methodologies in order to have them working in complementary and thus more effective way.

This special issue of the *International Journal of Computational Intelligence in Bioinformatics and Systems Biology (IJCIBSB)* deals with the topic 'classify the classifiers: investigating the optimum classification technique per case in bioinformatics'. The contributing papers are of high quality and multidisciplinary.

In the paper entitled 'Risk haplotype pattern discovery for gene mapping by recursive partitioning method based on weighted classification trees', Trang and Minh present a combinatorial approach based on recursive weighted longest prefix trees (RWLPT) for mining a massive genetic marker data. They demonstrate that the case-control association test based on haplotype patterns reduces degrees of freedom and number of tests while increasing the power of disease-haplotype association analysis. Their experimental results through the real published SNP dataset show that their approach tends to be more effective and powerful than the sliding window approach, classification and regression trees and single marker analysis.

In the paper entitled 'Table of periodic properties of human immunodeficiency virus inhibitors', Torrens and Castellano propose classification algorithms based on information entropy studying the feasibility of mixing a given human immunodeficiency virus type 1 (HIV 1) inhibitor with dissimilar ones, in a complex drug. Their analysis is in agreement with principal component analysis and compares well with other classifications.

In the paper entitled 'A minimum classification error framework suitable for multicriteria gene selection: discovery of differentially methylated genes in small B-cell lymphomas' , Popescu et al. present a framework for the identification of differentially methylated genes in several types of small B-cell lymphomas (SBCLs) using DNA differential methylation microarrays. The proposed approach uses the patient classification error to tune various steps of the algorithm such as the data normalisation, the gene ranking and the threshold for gene selection. Some of the identified genes are known to be involved in critical pathways such as apoptosis and proliferation, while others function as tumour suppressor genes or oncogenes.

In the paper entitled 'Evaluation of machine learning techniques for prostate cancer diagnosis and Gleason grading', Alexandratou et al. used 16 well-established, supervised machine learning algorithms and compared them based on their performance on prostate cancer diagnosis and Gleason grading. More specific, the classification problems addressed were: tumour vs. non-tumour, low vs. high grade and the four class problem of diagnosis and grading (normal, Gleason 3, Gleason 4 and Gleason 5). For the best performing algorithm in each case, the accuracy obtained for diagnosis (tumour-non-tumour), for low-high grade discrimination and for accomplishing both diagnosis and Gleason grading were calculated. Logistic regression and sequential minimal optimisation (SMO) for training an SVM were among the four top scoring algorithms in each classification problem.

In the paper entitled 'Automatic brain MRI segmentation scheme based on feature weighting factors selection on fuzzy c-means clustering algorithms with Gaussian smoothing', Xiao et al. introduce a new clustering method and apply it to brain magnetic resonance imaging (MRI) lateral ventricular compartments segmentation. The method uses Gaussian smoothing to enable fuzzy c-mean (FCM) to create both a more homogeneous clustering result and reduce effect caused by noise. Compared with the standard FCM with or without Gaussian smoothing, they found that the proposed scheme provides a better clustering performance for brain MRI lateral ventricular compartments segmentation.

In the paper entitled 'A comparative study of multiclassification methods for protein fold recognition', Valavanis et al. comparatively assess five different classification techniques, namely multilayer perceptron and probabilistic neural networks, nearest neighbour classifiers, multiclass SVMs and classification trees for fold recognition on a reference set of proteins that are organised in folds and are described by n-dimensional vectors of sequence-derived features. They evaluate all classifiers in terms of total accuracy, mutual information coefficient, sensitivity and specificity measurements using a ten-fold cross-validation method. A polynomial SVM and a multilayer perceptron of one hidden layer performed better and achieved satisfactory multiclass classification accuracies given the complexity of the problem and the reported similar classification performances of other researchers.